UNIVERSITAT DE BARCELONA

Functional interplay between ER and MAF transcription factors in breast cancer bone metastasis

Alicia Llorente Lope

Universitat de Barcelona
Facultat de Biologia
Programa de doctorat en Biomedicina

Institut de Recerca Biomèdica de Barcelona

Functional interplay between ER and MAF transcription factors in breast cancer bone metastasis

Memòria presentada per Alicia Llorente Lope per optar al grau de doctora per la Universitat de Barcelona

Barcelona, 2020

Doctoranda:
Alicia Llorente Lope

Director:
Dr. Roger Gomis Cabré

Tutor:
Dr. Joan J. Guinovart Cirera

A mis padres.

CONTENTS

SUMMARY

Metastatic breast cancer (BCa) is the leading cause of death in BCa patients. Current therapies for BCa aim at cell proliferation and, therefore, latent disseminated cells are able to escape treatment. For this reason, there is an urgent need to understand the molecular mechanisms by which BCa cells invade distant tissues to develop new therapeutic opportunities.

It was recently discovered that the v-maf avian musculoaponeurotic fibrosarcoma oncogene homolog (MAF) transcription factor drives specifically bone colonization in estrogen receptor (ER)-positive BCa patients. MAF was shown to transcriptionally regulate genes that support metastatic functions, thereby emerging as an attractive target for the prevention of BCa metastasis to the bone. However, the nuclear localization of MAF, its lack of enzymatic activity and its intrinsically disordered structure made this transcription factor a challenging therapeutic target and brought to the forefront the potential of MAF-transcriptionally controlled proteins as well as MAF-interacting partners to become pharmacological targets for the prevention of bone metastasis.

Membrane proteins are involved in direct interactions between cancer cells and the host tissue and can be easily targeted. For this reason, in the present thesis, we applied a One-Bead-One-Compound (OBOC) combinatorial peptide library screening to identify MAF-transcriptionally controlled proteins located at the plasma membrane as well as new synthetic peptide binders to target these proteins. We optimized a novel methodology to isolate high-affinity ligands from OBOC libraries, designed and synthesized different libraries and identified peptide ligands that specifically recognize metastatic BCa cells. However, further experiments are still required to unravel the identity of the membrane proteins recognized by peptide binder hits.

Additionally, we performed a proximity-dependent biotin identification (BioID) screen to identify MAF-interacting partners that may cooperate with this transcription factor to promote bone colonization. BioID proximity labeling captured the MAF interactome in BCa cells, mainly comprised of chromatin-remodeling proteins that may influence histone modification to generate transcriptionally active or repressive chromatin structures for gene expression regulation. Moreover, ER, the main driver of ER+ BCa, emerged as a MAF interactor.

MAF-positive tumors treated with bisphosphonates, a major therapeutic option for the treatment of bone metastasis, were associated with increased adverse outcomes in pre-

menopausal patients, thereby supporting a role for estrogen (E2), the main ER ligand, in the modification of MAF-positive tumors behavior. Given this observation, we considered the MAF-ER interaction of major interest and explored the molecular mechanisms underlying the cross-talk of both transcription factors.

Gene expression analyses revealed a MAF-E2 signature, which suggested that MAF modulates the E2 response either by triggering a switch of target genes or by potentiating the expression of specific E2-responsive genes. Interrogation of the chromatin landscape in BCa cells led to the observation that MAF expression was associated with increased chromatin accessibility. Moreover, assessment of ER binding to chromatin showed an increase in ER recruitment to DNA regulatory regions depending on MAF presence. Thus, our results demonstrate that MAF recruits chromatin-remodeling proteins to generate accessible chromatin regions, which result in an increase of ER recruitment to chromatin potentially to induce the expression of genes that confer metastatic properties.

LIST OF ABBREVIATIONS

Ac	Acetylation
ADH	Atypical ductal hyperplasia
ADP	Adenosine diphosphate
AITL	Angioimmunoblastic T-cell lymphoma
AJCC	American Joint Committee on Cancer
ALH	Atypical lobular hyperplasia
ANGPTL4	Angiopoietin-like 4
AP-1	Activator protein 1
ARID1A	AT-Rich Interaction Domain 1A
ARK5	AMPK-related protein kinase 5
ATAC-seq	Assay of transposase accessible chromatin sequencing
ATF	Activating transcription factor
ATP	Adenosine triphosphate
BALP	Bone alkaline phosphatase
BBB	Blood-brain barrier
BCa	Breast Cancer
BCS	Breast-conserving surgery
BFDR	Bayesian false discovery rate
BMP	Bone morphogenetic protein
BMU	Basic multicellular units
BRCA1/2	Breast cancer susceptibility gene ½
BRD4	Bromodomain-containing protein 4

BSP	Bone sialoprotein
BTG1	B-cell translocation gene 1
bZIP	Basic leucine zipper
CAPG	Macrophage-capping protein (CAPG)
CaSR	Calcium-sensing receptor
CBP	cAMP-responsive element-binding protein (CBP)
CCL2	C-C motif chemokine ligand 2
CD13/APN	Aminopeptidase N
CDH11	Cadherin 11
CDK4/6/9	Cyclin-dependent kinase 4/6/9
CHD	Chromodomain-helicase DNA-binding
ChIP-seq	Chromatin immunoprecipitation sequencing
CISH	Chromogenic *in situ* hybridization
CNA	Copy number aberration
Co-IP	Co-immunoprecipitation
COPAS	Complex Object Parametric Analyzer and Sorter
COX 2	Cyclooxygenase 2
CpG	Cytosine-phosphodiester-guanine
CRE	cAMP-responsive element
CREB	Cyclic AMP-responsive element-binding protein
CREBBP	CRE binding protein
CSF1R	Colony stimulating factor 1 receptor
CTC	Circulating tumor cells
CTX	C-terminal telopeptide of type I collagen

CXCL1/12	C-X-C Motif Chemokine Ligand 1/12
CXCR2/4	C-X-C Motif Chemokine Receptor 2/4
DAPK	Death-associated protein kinase
DBT	Digital breast tomosynthesis
DCIS	Ductal carcinoma *in situ*
DKK1	Dickkopf 1
DNMT	DNA methyltransferases
DOCK4	Dedicator of cytokinesis protein 4
DRIP	Vitamin D receptor interacting protein
DTC	Disseminated tumor cell
E2	Estrogen
ECM	Extracellular matrix
EGF	Epidermal growth factor
EGFR	Epidermal growth factor receptor
EHR	Extended homology region
EMP3	Epithelial membrane protein 3
EMT	Epithelial-to-mesenchymal transition
ER	Estrogen receptor
ERE	Estrogen response element
ERK	Extracellular signal-regulated kinase
ET-1	Endothelin 1
FA	Formic acid
FACS	Fluorescence-activated cell sorting
FDA	US Food and Drug Administration

FGF	Fibroblast growth factor
FISH	Fluorescence *in situ* hybridization
Fmoc	9-fluorenylmethoxycarbonyl
FOS	Proto-oncogene c-Fos
FOXA1	Forkhead Box A1
FSH	Follicle-stimulating hormone
GATA3	GATA binding protein 3
GCO	Global Cancer Observatory
GH	Growth hormone
GIPC1	PDZ domain-containing protein GIPC1 (GIPC1)
GLI2	GLI family zinc finger 2
GnRH	Gonadotropin-releasing hormone
GO	Gene ontology
GR	Glucocorticoid receptor
GSK3	Glycogen synthase kinase 3
GSTP1	Glutathione S-transferase pi
HA	Hemagglutinin
HAT	Histone acetyltransferases
HDAC	Histone deacetylase
HEK 293T	Human embryonic kidney 293-T
HER2	Human epidermal growth factor receptor 2
HGF	Hepatocyte growth factor
HIF	Hypoxia-inducible factor
HIF2A	Hypoxia-inducible transcription factor 2A

HMBA	Hydroxymethylbenzoic acid
HPLC	High-performance liquid chromatography
HR	Hormone receptor
HSC	Hematopoietic stem cell
HSP	Heat shock protein
I-κB	Inhibitor of kappa B
ICTP	Cross-linked C-terminal telopeptide of type I collagen
ID1	Inhibitor of DNA binding 1
IF	Immunofluorescence
IGF-1	Insulin-like growth factor 1
IHC	Immunohistochemistry
IL-11	Interleukin-11
ISWI	Imitation switch
JAG1	Jagged 1
JUN	Jun proto-oncogene
KYNU	Kynureninase
L1CAM	L1 cell adhesion molecule
LCIS	Lobular carcinoma *in situ*
LH	Luteinizing hormone
LIFR	Leukemia inhibitory factor receptor
LPA	Lysophosphatidic acid
LSD1	Lysine specific demethylase 1
M-CSF	Macrophage colony-stimulating factor
m/z	Mass/charge ratio

MAF	v-maf avian musculoaponeurotic fibrosarcoma oncogene homolog
MAF L	MAF long isoform
MAF S	MAF short isoform
MALDI-TOF	Matrix-assisted laser desorption ionization time-of-flight
MAPK	Mitogen-activated protein kinase
MARE	MAF-recognition element
MaSC	Mammary stem cell
Me	Methylation
MET	Mesenchymal-to-epithelial transition
METABRIC	Molecular taxonomy of breast cancer international consortium
miRNA	microRNA
MM	Multiple myeloma
MMPs	Matrix metalloproteinases
MRI	Magnetic resonance imaging
MS	Mass spectrometry
MSCs	Mesenchymal stem cells
MSK1	Mitogen-and stress-activated kinase 1
mTOR	Mammalian target of rapamycin
MUC1	Mucin 1
NaCl	Sodium chloride
NADPH	Nicotinamide adenine dinucleotide phosphate
NCOA1/2/3	Nuclear-receptor coactivator 1/2/3
NCOR1/2	Nuclear-receptor corepressor 1/2
NFκB	Nuclear factor kappa-light-chain-enhancer of activated B cells

NFAT	Nuclear factor of activated T-cells
NK	Natural killer
NST	Invasive carcinoma of no special type
NTX	N-terminal telopeptide of type I collagen
NurD	Nucleosome remodeling and deacetylase
O/N	Overnight
OBOC	One-bead-one-compound
OBTC	On-bead-two-color
OPG	Osteoprotegerin
OPN	Osteopontin
OS	Overall survival
Osx	Osterix
P1NP	Procollagen type 1 N-terminal propeptide
PAK4	p21-activated kinase 4
PARP	Poly (ADP-ribose) polymerase
PBS	Phosphate-buffered saline
PCAF	p300/CBP-associated factor
PD-1	Programmed death 1
PD-L1	Programmed death-ligand 1
PDA	Photodiode array
PDGF	Platelet-derived growth factor
PEI	Polyethyleneimine
PFS	Progression-free survival
PGI_2	Prostacyclin

PI3K	Phosphoinositide 3-kinase
PMMA	Polymethyl methacrylate
POSTN	Periostin
PR	Progesterone receptor
PRC1/2	Polycomb repressive complex 1/2
PROTAC	Proteolysis targeting chimera
PTH	Parathyroid hormone
PTHrP	Parathyroid hormone related protein
RANK	Receptor activator of nuclear factor-κB
RANKL	Receptor activator of nuclear factor-κB ligand
RARRES3	Retinoic acid receptor responder protein 3
REST	Repressor Element-1 Silencing Transcription factor
RFP	Red fluorescent protein
RNA-seq	RNA sequencing
ROS	Reactive oxygen species
RT	Room temperature
RT-qPCR	Reverse transcription quantitative polymerase chain reaction
Runx2	Runt-related transcription factor 2
SERD	Selective estrogen receptor degrader
SERM	Selective estrogen receptor modulator
Sir2	Silent information regulator 2
Sirt1/2	Sirtuin 1/2
SOX	SRY-related HMG box
SPPS	Solid-phase peptide synthesis

SRE	Skeletal-related events
STAT	Signal transducer and activator of transcription
TAU	Microtubule-associated protein tau
TBK1	TANK-binding kinase 1
TBP	TATA binding protein
TCGA	The cancer genome atlas
TDLU	Terminal ductal lobular unit
TEB	Terminal end bud
TGF-β	Transforming growth factor β
TIC	Total ion chromatogram
TLR	Toll-like receptor
TN	Triple negative
TNC	Tenascin C
TNM	Tumor-Node-Metastasis
TRAP	Thyroid hormone receptor-associated protein
TRE	12-O-tetradecanoyl phorbol 13-acetate-responsive element
TrxG	Trithorax group
TSS	Transcription start site
uPA	Urokinase
UPLC-MS/MS	Ultra performance liquid chromatography-tandem mass spectrometry
VCAM1	Vascular cell adhesion molecule 1
VE	Vascular endotelial
VEGF	Vascular endotelial growth factor
VEGFR1	Vascular endothelial growth factor receptor 1

WHO	World Health Organization
Y2H	Yeast-two-hybrid
ZEB1	Zinc finger E-box binding homeobox 1
ZOL	Zoledronic acid

Name of Amino Acid	Three Letter Code	One Letter Code
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic Acid	Asp	D
Cysteine	Cys	C
Glutamic Acid	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

Introduction

1. BREAST CANCER

Breast cancer (BCa) is the most frequent type of cancer in women worldwide. Men can also suffer from BCa, although they make up less than 1% of all breast cancer diagnoses and it is considered a rare disease in this population[1]. The latest data from the Global Cancer Observatory (GCO) estimated 2,088,849 new BCa cases in 2018, accounting for 25% of all new cancer cases in women, and 626,679 BCa deaths, accounting for 15% of all cancer deaths and making BCa the leading cause of cancer death in females[2].

BCa usually gives no symptoms when the tumor is small, and screening is therefore indispensable for early detection[3]. BCa screening methods include mammography, digital breast tomosynthesis (DBT), magnetic resonance imaging (MRI) and breast ultrasound, and the choice of method depends on the characteristics of the patient. When the tumor grows, the main physical sign is a painless lump in the breast. However, BCa can sometimes spread to axillary lymph nodes and cause swelling even before the primary breast tumor can be detected. When cancer is suspected, the clinical praxis consists of either a needle or a surgical biopsy, depending on the size and location of the mass, for microscopic analysis.

BCa mortality has declined since 1990 due to improvements in both screening methods and treatment. Nowadays, the 5-year relative survival rate for women diagnosed with BCa is 90%. Nevertheless, when metastases are present at the time of diagnosis, this percentage is dramatically reduced to 27%[3].

Metastatic BCa remains an incurable disease. In this regard, current therapeutic goals for this condition seek to prolong life and palliate symptoms. Therefore, to improve therapeutic approaches to target cell dissemination and prevent metastatic BCa, there is an urgent need to understand the mechanisms by which BCa cells invade distant tissues.

1.1 The mammary gland: physiology and pathology

Mammary glands, a synapomorphy that defines mammals, are specialized structures that serve to produce milk to nourish offspring. It is fundamental to understand the structure and function of the mammary gland to comprehend normal development as well as the malignant transformation that occurs during BCa tumorigenesis.

1.1.1 Structure of the mammary gland

The mammary gland is a branching structure composed of ducts and alveoli (Figure 1A). Mammary gland development and differentiation is completed during puberty and adulthood, and the structure of the gland changes during the female reproductive cycle in response to ovarian and pituitary hormones. During these stages, cells proliferate, differentiate or undergo apoptosis, resulting in a marked remodeling of the tissue architecture[4].

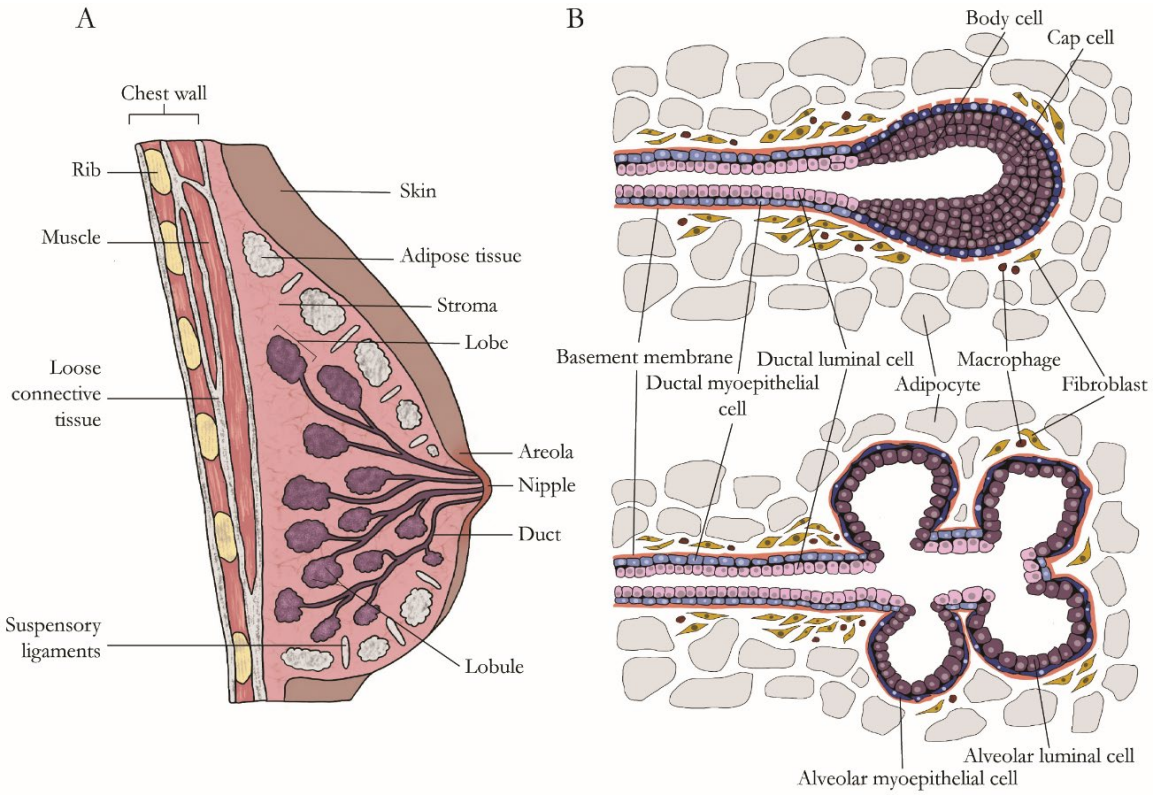


Figure 1. Anatomy and structure of the female breast. (A) Schematic representation depicting the anatomy of the human breast. The mammary gland is formed by 17-30 individual lobes, each of them connected to the nipple through a branching duct system embedded in the mammary fat pad. **(B)** Structure of a terminal end bud (TEB) during duct extension (top) and midpregnancy (bottom). TEBs are formed by a layer of cap cells surrounding multiple layers of body cells. Cap cells give rise to myoepithelial cells and body cells either generate luminal epithelial cells or undergo apoptosis for lumen generation. Differentiated myoepithelial cells and luminal epithelial cells line the ducts and, during pregnancy, a basket-like layer of myoepithelial cells surrounds milk-producing alveolar cells, forming alveoli for milk storage. Adapted from Ali and Coombes (2002), Sternlicht (2005) and Sopel (2010).

Two compartments compose the mammary gland: the epithelium, formed by ducts and alveoli; and the stroma, called the mammary fat pad. A variety of cell types with diverse functions and characteristics form these compartments. The vast network of ducts and alveoli is composed of epithelial cells and is embedded in the stroma, formed mainly by adipocytes, fibroblasts, hematopoietic cells, blood vessels and neurons. Milk is stored in alveoli, which cluster into lobules, which in turn group to form a single lobe[4,5].

Epithelial cells are organized in a bi-layered structure of ducts and alveoli with a central lumen that drains the milk to the nipple. This branching system is formed by an outer layer of basal myoepithelial cells, which contract to facilitate the delivery of milk, surrounding an inner layer of luminal cells divided into ductal luminal cells, which line in the ducts, and alveolar luminal cells, which produce and secrete milk during lactation[4,5].

During puberty, ducts start to grow and invade the mammary fat pad. The tips of the growing ducts become specialized structures named terminal end buds (TEBs), which guide duct extension through the adipose tissue (Figure 1B). TEBs are composed of multiple inner layers of body cells, which will give rise to the luminal cell compartment, and an outer layer of cap cells, which will eventually become myoepithelial cells. TEBs will originate new branches and alveolar buds that will gather around a terminal duct to form a lobule[5]. Mammary stem cells (MaSCs) are located at TEBs and have the capacity to give rise to mammary epithelial progenitor cells, which are crucial for alveolar renewal in each pregnancy. Experiments have proven that MaSCs are capable of regenerating an entire mammary gland[6,7]. In humans, the functional part of the mammary gland is called terminal ductal lobular unit (TDLU), which is formed by the combination of a TEB and alveolar buds[8].

Adipocytes comprise an extensive part of the mammary fat pad in the non-lactating gland. During pregnancy and lactation, adipocytes use their lipid reservoir to supply fat for the energy-consuming process of milk production[9]. Adipocytes also regulate the growth and function of the mammary epithelium by secreting paracrine and endocrine molecules such as prolactin, estrogen (E2), insulin-like growth factor I (IGF-I), transforming growth factor-β (TGF-β), hepatocyte growth factor (HGF) and leptin[10–15]. Furthermore, adipocytes also stimulate angiogenesis by secreting vascular endothelial growth factor (VEGF)[16].

Fibroblasts are usually located close to the basal side of the epithelium and can interact with it through growth factors and proteases. Fibroblasts can support epithelial cell survival and

synthesize extracellular matrix (ECM) components such as collagens, proteoglycans, fibronectin and matrix metalloproteinases[17].

Endothelial cells form a vast vascular and lymphatic network in the mammary fat pad. The generation of these vessels is driven by signals from epithelial cells and adipocytes[17].

Immune cells such as macrophages and eosinophils are recruited for the invasion of the epithelium into the mammary fat pad[18]. Macrophages also participate in processes such as epithelial cell death and adipocyte repopulation during mammary remodeling[19].

1.1.2 Mammary gland development

During embryonic development, the initial structure of the breast is established. Prior to puberty, mammary gland growth is independent of hormone signaling and there are no major differences between males and females. In females, the onset of puberty causes TEBs to become specialized structures at the tip of the ducts to drive their branching and elongation until the end of the mammary fat pad. Lobule formation starts, but development and differentiation of the mammary gland is not completed. After puberty, the mammary gland undergoes growth and involution during menstrual cycles until pregnancy, when it is largely remodeled[17,20].

During pregnancy and lactation, alveolar and ductal cells start to proliferate to form the lobuloalveolar buds that will produce and secrete milk. At this point, the mammary epithelium occupies most of the mammary fat pad. Alveolar cells secrete milk into the lumen and myoepithelial cells contract to facilitate the movement of the milk through the ducts. After weaning, the accumulation of milk inhibits further milk synthesis, lobuloalveolar buds are remodeled in an involution process, and cells undergo apoptosis, leaving a mammary gland similar to that present before pregnancy. In further pregnancies, growth and involution stages are repeated[17,20].

1.1.3 Hormones that regulate mammary gland development

Mammary gland development is hormone-independent up to puberty, but thereafter it is rigorously controlled by ovarian and pituitary hormones.

At puberty, neurons of the hypothalamus secrete gonadotropin-releasing hormone (GnRH), which stimulates gonadotropic cells in the anterior pituitary gland to produce follicle-stimulating hormone (FSH) and luteinizing hormone (LH) (Figure 2). Likewise, FSH and LH stimulate the synthesis of E2 and progesterone by the ovaries in oscillating amounts during each menstrual cycle. In response to E2 and progesterone release, the duct system of the mammary gland starts expanding and branching into the mammary fat pad[20,21].

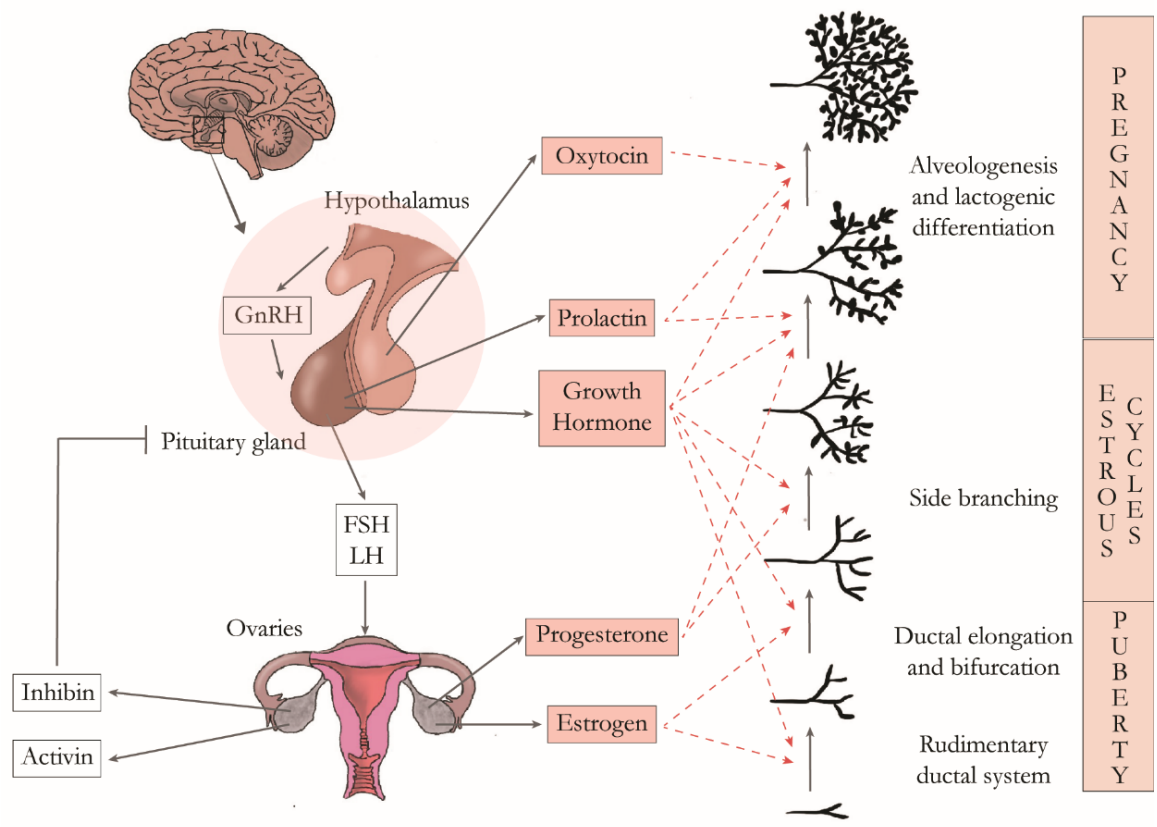


Figure 2. The hypothalamus-pituitary-ovarian axis in mammary gland development. Gonadotropin-releasing hormone (GnRH) produced by the hypothalamus is sensed by the anterior pituitary gland and stimulates the secretion of the follicle-stimulating hormone (FSH) and luteinizing hormone (LH) gonadotropins which, in turn, trigger the production of estrogen and progesterone by the ovaries. These hormones, together with ovarian inhibins and activins and the pituitary gland-secreted growth hormone, prolactin and oxytocin, have an important impact on different stages of mammary gland development. Dotted arrows depict hormonal regulation. Adapted from Brisken and O'Malley (2010).

During pregnancy, progesterone and prolactin are released and alveoli are formed. E2 stimulates prolactin synthesis and induces the expression of the progesterone receptor (PR).

In turn, prolactin and progesterone trigger estrogen receptor (ER) expression to enhance ductal growth. Progesterone interferes with prolactin signaling, thus preventing lactation. During the first days after childbirth, high progesterone levels lead to the delivery of colostrum, the first fluid produced by the mammary glands. Days later, a reduction in circulating progesterone allows prolactin signaling and consequently further alveologenesis and milk production. Suckling of the baby causes the release of oxytocin, which triggers alveolar cell proliferation and stimulates the contraction of the myoepithelial cells that surround the alveoli, facilitating milk ejection[21]. When suckling is stopped, the loss of prolactin signaling leads to massive luminal cell death and results in the involution of the mammary gland to its original adult state[4,22].

The pituitary gland also produces growth hormone (GH), essential for mammary gland development during both puberty and pregnancy. GH stimulates the mammary stroma to produce IGF-1, which is recognized by epithelial cells and triggers the formation of TEBs and extension of the ducts into the mammary fat pad[23,24]. Activins and inhibins are members of the TGF-β family and are also important regulators of the menstrual cycle and mammary gland development. These proteins are produced in the ovaries and they have opposite effects; activins enhance FSH synthesis and secretion, while inhibins block these processes. Activin and inhibin signaling is required for duct elongation and epithelial cell differentiation during puberty and pregnancy[25].

In summary, an extensive hormonal network orchestrates mammary gland development and remodeling by acting on several organs and regulating hormone synthesis and secretion.

1.1.4 Mammary tumorigenesis

Sensing of E2 by epithelial cells of the mammary gland triggers cell division, which is required for proper mammary development. Some malignant cells in the breast retain this feature and are dependent on hormone signaling for proliferation. For this reason, the risk of developing BCa is highly influenced by E2[26].

1.1.4.1 Estrogen in breast carcinogenesis

E2 is the most relevant hormone in BCa progression. E2 is synthesized mainly in the ovaries. However, it is also synthesized in other organs, such as the adipose tissue, bone, vasculature and brain, albeit at much lower amounts. E2 is essential for female sexual and reproductive development but also for the maintenance of bone homeostasis. The arrival of menopause results in cessation of ovarian E2 production, and E2 synthesis in the adipose tissue is therefore important for preserving bone density in postmenopausal women[21,27].

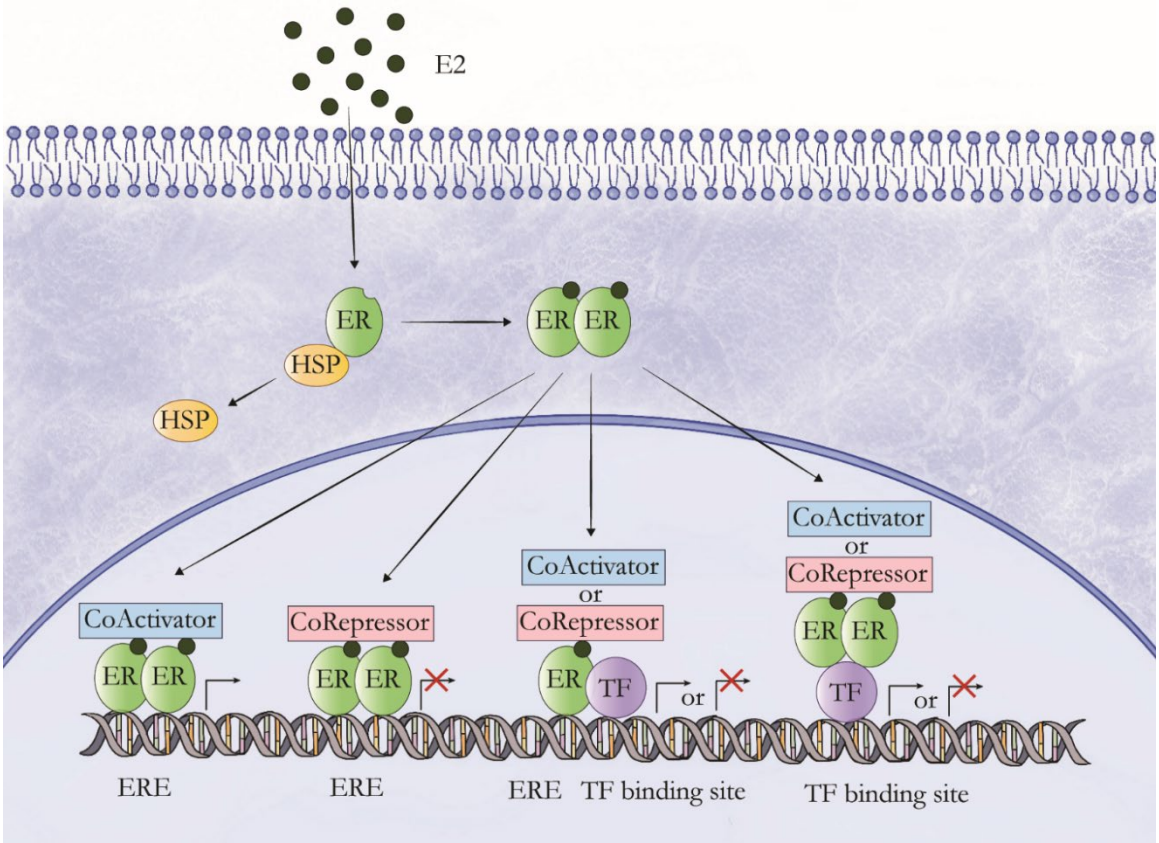


Figure 3. Mechanism of estrogen receptor (ER) action. In the absence of estrogen (E2), the ER is located in the cytoplasm bound to heat shock protein (HSP) complexes. Upon E2 binding, the ER dissociates from HSP complexes, homodimerizes and translocates to the nucleus, where it binds to specific DNA sequences known as estrogen response elements (EREs) and recruits coactivators or corepressors to activate or repress target gene transcription. Alternatively, the ER can bind to other DNA-binding transcription factors (TF) either by binding to EREs near their response element or by tethering to the partner TF. Adapted from Treviño and Weigel (2014).

E2 binds to two nuclear receptors, ERα and ERβ, which are encoded by *ESR1* and *ESR2*, respectively. Epithelial cells of the mammary gland express ERα, which is indispensable for ductal elongation and alveolar development[28]. Lower levels of ERβ are also expressed in the mammary gland. However, although evidence supports a role of ERβ in the organization, adhesion and differentiation of epithelial cells, this receptor is not required for ductal growth[29]. Because ERβ has a minor effect on mammary gland development and on BCa progression, ERα will be herein referred to as ER.

ER dimerizes upon E2 binding and translocates from the cytoplasm to the nucleus. In the nucleus, the dimer binds to small palindromic DNA motifs called estrogen response elements (EREs), located at the promoter of ER target genes, and functions as a transcription factor[30] (Figure 3). The transcriptional activity of the ER can be modulated by several coactivators and corepressors that facilitate recruitment of the general transcription machinery or activate or repress transcription through chromatin remodeling. Some ER coactivators include nuclear-receptor coactivator 1 (NCOA1), NCOA2 and NCOA3, the switch/sucrose non-fermentable (SWI/SNF) and the thyroid hormone receptor–associated protein (TRAP)/ vitamin D receptor-interacting protein (DRIP) chromatin remodeling complexes and histone-acetyltransferases (HATs) such as CREB-binding protein (CBP) and p300/CBP-associated factor (PCAF). HATs promote histone hyperacetylation, which correlates with chromatin opening and actively transcribed genes. Conversely, some ER corepressors include nuclear-receptor corepressor 1 (NCOR1) and NCOR2, which can associate with histone deacetylase complexes (HDACs) to repress gene expression[4].

Epidemiological studies show that many BCa risk factors are related to exposure of the breast tissue to hormones throughout life. For instance, early menarche, late menopause and the use of hormonal contraceptives or menopausal hormone therapy, both based on a combination of E2 and a synthetic hormone with similar effects to those of progesterone called progestin, are associated with an increased risk of developing BCa[31–33]. Moreover, evidence demonstrates that elevated endogenous E2 blood and urinary levels correlate with increased risk of BCa and that their decrease by either hormonal therapy or ovarian suppression in BCa patients results in an important reduction of recurrence and death[31]. Collectively, these data support a causal role for E2 in BCa.

E2 stimulates the proliferation of mammary cells, resulting in increased DNA replication and increased probability of DNA damage and mutations, thereby facilitating the appearance of

malignant cells. Furthermore, E2 metabolism generates E2 metabolites with mutagenic potential that might contribute to BCa progression[34–36].

Most breast tumors are characterized by ER overexpression and rely on E2 signaling for proliferation and survival, and they are remarkably sensitive to hormonal therapy. E2 production in the adipose tissue is extremely important for BCa progression, especially in postmenopausal women. The concentration of E2 in BCa tumors of postmenopausal patients is significantly higher than that found in plasma due to high expression of aromatase, an essential enzyme in E2 synthesis, in epithelial cells[37].

1.1.4.2 Histology of breast cancer progression

Abnormal proliferation of epithelial cells in the TDLUs of the breast gives rise to atypical lobular hyperplasia (ALH) or atypical ductal hyperplasia (ADH), depending on whether the epithelial cells are from a lobe or a duct, respectively. ALH can progress to lobular carcinoma *in situ* (LCIS), which is considered a benign breast condition associated with increased risk of developing BCa but without the potential to progress to invasive cancer[38]. The term *in situ* carcinoma refers to a hyperproliferation of cells confined to cell layer of origin. Alternatively, ADH can progress to a ductal carcinoma *in situ* (DCIS), considered a preinvasive lesion and defined by the proliferation of ductal epithelial cells with all the morphologic features of malignancy but without basement membrane penetration (Figure 4). DCIS can progress to invasive ductal carcinoma if the abnormal cells break through the basement membrane of the ducts and penetrate the surrounding breast tissue. At this point BCa is considered to be in a local stage, where cells have broken the myoepithelial cell layer and penetrated the basement membrane but have not left the breast. BCa cells can then spread to surrounding tissue or nearby lymph nodes, known as regional stage, and finally reach distant organs[39].

Most of the characteristic molecular changes of invasive BCa are already present in DCIS, yet the lesion does not have a malignant phenotype. Final changes in the expression of genes involved in adhesion and invasion at the same time as alterations in the surrounding stroma, such as angiogenesis, fibroblast proliferation and lymphocyte infiltration, are associated with the evolution from preinvasive DCIS to invasive carcinoma[40]. Not all DCIS progress to invasive BCa, but approximately 50% of recurrence after local therapy for DCIS are invasive[41].

Although BCa progression has historically been defined as the aforementioned simple pathway of histologically different stages, nowadays it is perceived as a complex series of stochastic events that lead to invasive BCa through divergent pathways, with nuances in histological characteristics[42]. The World Health Organization (WHO) currently classifies precancerous lesions and invasive BCa into distinct histological subtypes with different prognostic implications based on cell morphology, growth and architecture patterns to serve as a reference for pathologists, clinicians and researchers[43].

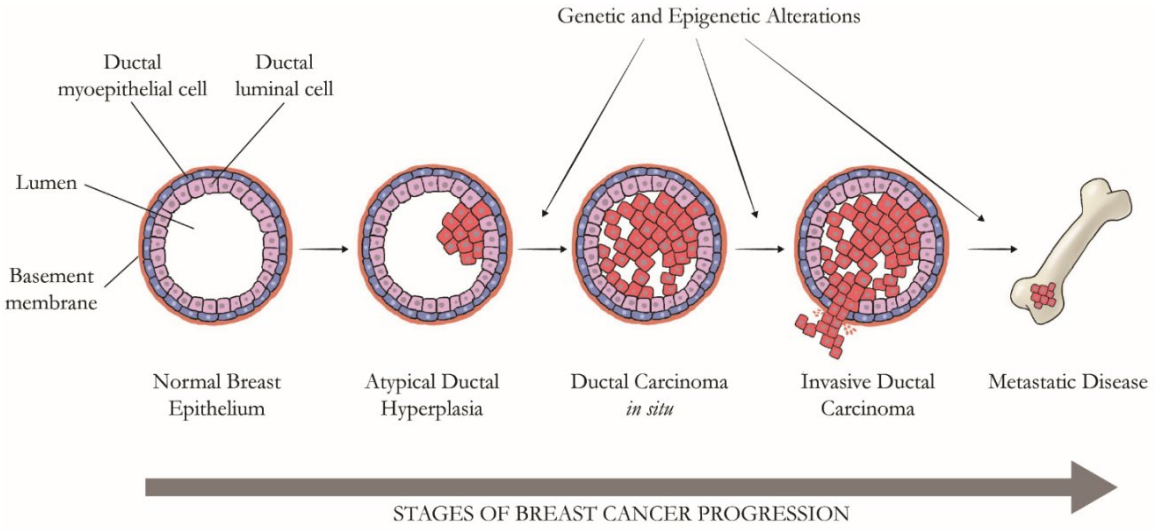


Figure 4. **Schematic outline of breast cancer progression.** Breast cancer development goes through defined histological stages and is driven by the accumulation of genetic, epigenetic and microenvironmental alterations. Atypical ductal hyperplasia (ADH) is characterized by abnormal ductal epithelial cell proliferation and can progress to ductal carcinoma *in situ* (DCIS), defined by a complete filling of the mammary duct with cancer cells. Once cancer cells break the basement membrane, the lesion is classified as invasive ductal carcinoma and at this stage cancer cells might be able to colonize distant tissues. Adapted from Kothari *et al.* (2018).

This histological classification allows categorization of the heterogeneity found in BCa based on growth patterns and architectural features. However, the lack of a molecular component limits the ability to predict response to specific therapies. Hence, to help clinicians select the best therapeutic option, the histological assessment of BCa tumors still needs to be complemented with the evaluation of certain biological markers.

1.2 Clinical classification of breast cancer

BCa is no longer conceived as a single disease, but it is considered a group of diseases that differ in biological and pathological features, clinical presentation, response to treatment and outcomes, which requires different therapeutic approaches. In clinical management, breast tumors are currently classified on the basis of the status of three biological markers: ER, PR and Human epidermal growth factor receptor 2 (HER2/ERBB2).

The hormonal status of the tumor is assessed by immunohistochemistry (IHC). Ttumors are considered hormone receptor-positive (HR+) if the presence of ER, PR or both represents at least 1% of the total nuclei in the sample[44]. When tumors are positive for HRs, their proliferation rate can be further assessed by the presence of the Ki67 marker[45].

HER2 expression is also determined by IHC and positivity is defined by a HER2 score. A value of 0 indicates no staining or incomplete or faint membrane staining in 10% or less of tumor cells, 1+ incomplete or faint membrane staining in more than 10% of tumor cells, 2+ weak to moderate complete membrane staining in more than 10% of tumor cells and 3+ complete and intense circumferential membrane staining in more than 10% of tumor cells. A HER2 score of 0 or +1 is considered negative and only a score of 3 directly denotes positivity. When the assigned score is 2 the amplification ratio must be assessed by fluorescence *in situ* hybridization (FISH) or by chromogenic *in situ* hybridization (CISH) in order to confirm HER2 positivity[46].

The ER, PR and HER2 biological markers are reliable, inexpensive and effective for categorizing BCa patients into 3 groups with differences in prognosis, survival and response to therapy[47]:

Hormone receptor positive (HR+) tumors include the luminal A and luminal B subtypes. These tumors are characterized by ER and PR expression and lack of HER2 expression. HR+ tumors are the most common BCa subtype and they show a remarkable response to hormonal therapy[47]. The luminal A and B subtypes differ in their proliferation rate; luminal B tumors are highly proliferative compared to luminal A tumors and have a poorer outcome. The high proliferation rate associated with luminal B tumors is defined by the presence of the proliferation marker Ki67 in at least 20% of the tumor cells[45.] Luminal A tumors have the most favorable prognosis among all BCa subtypes.

HER2-overexpressing tumors are characterized by expression of the transmembrane tyrosine kinase HER2. This protein is a member of the epidermal growth factor receptor (EGFR) family, which, after activation upon dimerization and transphosphorylation, triggers numerous signaling pathways involved in cell proliferation, survival, differentiation, angiogenesis, invasion and metastasis. HER2+ tumors are aggressive and have poor prognosis, although remarkable advances in targeted therapies for HER2 in recent years have considerably improved the outcome of these patients[48]

Triple negative (TN) tumors have no expression of ER, PR or HER2. These tumors are extremely aggressive and have the worst prognosis with high risk of relapse and short progression-free survival (PFS) and overall survival (OS). There is a lack of significant advances in therapeutic treatment for TN BCa[49].

Table 1. TNM classification of breast tumors. Adapted from Giuliano *et al.* (2018)[50].

Primary tumor size		Lymph node involvement		Distant Metastasis	
TX	Tumor cannot be assessed	**NX**	Lymph nodes cannot be assessed	**MX**	Metastasis cannot be assessed
		N0	No lymph nodes affected		
T0	No evidence of primary tumor	**N1**	Micrometastases in axillary lymph nodes	**M0**	No distant metastasis
			1-3 axillary lymph nodes affected		
Tis	Carcinoma *in situ*		Internal mammary lymph nodes affected but only found on sentinel lymph node biopsy	**M1**	Distant metastasis present
T1	<2 cm				
T2	2-5 cm				
		N2	4-9 lymph nodes affected		
T3	>5 cm		Internal mammary lymph nodes affected		
T4	Any size but growing into chest wall or skin	**N3**	>10 axillary lymph nodes affected		
			Infraclavicular lymph node affected		
			Internal mammary lymph nodes + ≥1 axillary lymph node affected		
			≥4 axillary lymph nodes + internal mammary lymph nodes affected found on sentinel lymph node biopsy		
			Supraclavicular lymph nodes affected		

To fine-tune clinical management, the American Joint Committee on Cancer (AJCC) has established a pathological prognostic staging for BCa on the basis of to the status of the three biological markers ER, PR and HER2, in combination with information about the histological type and grade of the tumor, size, lymph node involvement and distant metastases. These new

measures add information of the biology of the tumor and allow a refinement of prognosis and better selection of therapy.

Tumor grade, which ranges from 1 to 3, refers to the differentiation state of the cells compared with normal tissue. Grade 1 tumor cells are well differentiated and low-proliferating, grade 2 denotes moderately differentiated cells with faster proliferation rates and grade 3 indicates poorly differentiated and highly proliferative cells[38,50].

Tumor size, lymph node involvement and metastasis are evaluated using the Tumor-Node-Metastasis (TNM) system (Table 1). This system is based on three parameters that measure the size of the primary tumor (T), the number and location of affected lymph nodes (N) and the presence of metastasis (M)[38,50]. Higher numbers associated with these parameters are interpreted as a more advanced stage of disease.

With all this information, BCa is classified in different stages that describe the extent of the cancer within the body at the time of diagnosis, as well as patient prognosis (Table 2). BCa stages range from 0 to IV, with stage 0 describing non-invasive cancers (*in situ*) and stage IV invasive cancers that have metastasized. The higher the stage of the cancer, the poorer the prognosis[50].

Table 2. Example of the 8th edition of the AJCC pathological prognostic staging with biomarkers for breast cancer. Adapted from Giuliano *et al.* (2018)[50].

When TNM is…	And Grade is…	And HER2 Status is…	and ER Status is…	And PR Status is…	Then the Pathological Prognostic Stage Group is…
T1, N0, M0	G1	Negative	Positive	Positive	IA
T3, N0, M0	G3	Positive	Negative	Negative	IIB
T2, N2, M0	G2	Negative	Negative	Negative	IIIB
Any T, Any N, M1	Any	Any	Any	Any	IV

1.3 Molecular classification of breast cancer

The complexity of BCa is not fully captured by the aforementioned parameters used in clinics to divide patients into groups with different prognosis and treatment decisions. Gene expression profiling has revealed distinct gene expression patterns among breast tumors and

has made it possible to establish five intrinsic molecular BCa subtypes and a normal breast-like group with remarkable differences in incidence, survival and response to therapy[51–53]. DNA microarray experiments suggest that tumor subtypes might originate from different cell types. Molecular BCa subtypes should be contemplated as independent diseases that require different treatment approaches.

Luminal A is the most common BCa subtype, accounting for 50-60% of all BCas[54]. Luminal A tumors are characterized by high expression levels of ER and luminal markers such as cytokeratins 8 and 18, Forkhead Box A1 (FOXA1) and GATA binding protein 3 (GATA3) and low expression of proliferation-related genes. This group represents the least aggressive subtype and is associated with the most favorable prognosis[52,53].

Luminal B tumors represent 15-20% of all BCas[54]. Luminal B tumors also express ER and luminal markers but differ from luminal A tumors in that they show an increased expression of proliferation-related genes, which translates into a more aggressive clinical behavior and poorer prognosis[52,53].

HER2-enriched tumors comprise 15-20% of all BCas[54]. This subgroup is characterized by overexpression or amplification of the tyrosine kinase *HER2/ERBB2* gene or high expression levels of HER2-related genes. HER2-enriched tumors have high expression of proliferation-related genes, intermediate expression of luminal-related genes and low expression of basal-related genes. Moreover, these tumors have the highest number of mutations across the genome, 72% of them showing mutations in tumor protein p53 (*TP53*). HER2-enriched tumors are aggressive and associated with poor prognosis[52,53].

Basal-like tumors account for 8-37% of all BCas[54]. The basal-like subtype is characterized by high expression of proliferation-related genes and basal myoepithelial markers such as cytokeratins 5, 14, 17 and laminin and very low or no expression of HER2-related genes and luminal genes such as ER or PR. Of note, mutations in Breast cancer susceptibility gene 1 (*BRCA1*) mutations are associated with the basal-like tumor subtype and 80% of these tumors present *TP53* mutations. Basal-like tumors are very aggressive and are related to poor clinical outcome[52,53].

Claudin-low tumors account for 7-14% of all BCas[55,56]. This subtype was defined later than the previous subtypes. Claudin-low tumors are characterized by low expression of claudin 3, 4 and 7, tight junction proteins and E-cadherin, all of them involved in cell-cell adhesion

processes. These tumors also present low expression of luminal differentiation markers and features of mesenchymal and mammary stem cells, which results in poor prognosis[57,58].

Normal breast-like tumors comprise 5-10% of all BCas[54]. These tumors share characteristics with fibroadenomas and normal breast samples and they are enriched for genes usually expressed in cells from the adipose tissue. Normal breast-like tumors show low expression of luminal epithelial genes and high expression of basal epithelial genes, although they are negative for cytokeratin 5 and EGFR. This subgroup presents intermediate expression of proliferation-related genes and has the most favorable prognosis after luminal A tumors. Normal breast-like cancers are poorly characterized, and it is believed that these tumors could be a technical artifact derived from high contamination with normal tissue during analysis[52,53,59].

Understanding the molecular profile of BCa tumors provides a more refined stratification and facilitates the selection of the most appropriate therapeutic treatment. For instance, patients with HR+ HER2- tumors but with a HER2-enriched profile can benefit from lapatinib, a drug that inhibits HER2 signaling, in combination with hormonal therapy[60]. Molecular profiling is extremely relevant, as these patients would receive only hormonal treatment if classified solely on the basis of classical clinical parameters.

During BCa progression, cancer cells undergo many genetic alterations that can translate into a variation of the intrinsic molecular subtype between the primary tumor and the metastatic lesion. A recent study showed that more than 50% of luminal A primary tumors switched to a more aggressive luminal B or HER2-enriched subtype in the metastatic lesion. Likewise, 30% of luminal B primary tumors switched to a HER2-enriched subtype at the metastatic site. Metastatic tumors are enriched for proliferation- and migration-related genes and have a decreased expression of luminal-related genes. Thus, evaluation of the intrinsic molecular subtype of the primary but also the metastatic lesion is crucial to predict patient outcome[61].

Although molecular profiling has brought about a significant advance understanding BCa heterogeneity and paved the way to the selection of the most effective therapy, it is still not routinely implemented in the clinical setting. Some prognostic molecular platforms such as MammaPrint, Veridex 76-gene, MapQuant Dx/simplified, Oncotype Dx and Endopredict have been developed to analyze the expression of specific genes and identify patients at high risk of distant recurrence in order to avoid aggressive therapeutic options in those with a favorable prognosis[62]. Moreover, other molecular platforms such as BluePrint and PAM50

allow efficient classification of BCa into specific molecular subtypes and serve as an effective replacement for full microarray analysis, which would be too expensive for regular clinical practice[63,64].

1.4 Breast cancer treatment

Many considerations regarding the stage and biological characteristics of the tumor, the patient's age and menopausal status, and the risks and benefits associated with each therapeutic option are made by clinicians for BCa treatment[65] (Table 3).

Table 3. Prevalence, prognosis and typical therapeutic options for the three clinical subtypes of breast cancer. Adapted from Waks and Winer (2019)[65].

	Hormone Receptor (HR) +/ HER2-	HER2+ / HR+ or HR-	Triple-Negative
Percentage of breast cancer cases	70%	15-20%	15%
Prognosis			
Stage I (5 year breast cancer specific survival)	≥99	≥94	≥85
Metastatic (median overall survival)	4-5 years	5 years	10-13 months
Typical systemic therapies for non-metastatic disease			
Endocrine therapy	Yes	Yes (if HR+)	No
HER2-targeted therapy	No	Yes	No
Chemotherapy	Yes (when hormonal therapy resistance develops)	Yes	Yes

Local therapy comprises approaches that target the breast and the nearby tissues and includes surgery and radiation therapy. On the other hand, **systemic therapy** consists of treatments that affect the entire body and includes chemotherapy, targeted therapy, hormonal therapy, and immunotherapy. Local therapy is the most common approach used for localized tumors while systemic therapy is the strategy of choice for patients with metastatic disease. **Neoadjuvant therapy** refers to a systemic therapy administered to a patient in order to reduce

the tumor size prior to surgical intervention, thereby facilitating tumor removal, whereas **adjuvant therapy** refers to a treatment to reduce the risk of recurrence after the main therapeutic intervention[65].

1.4.1 Surgery

Surgery implies the removal of a tumor and it is the first strategy of choice to treat BCa.

Breast-conserving surgery (BCS) is also known as lumpectomy, quadrantectomy or partial mastectomy, depending on the area of the breast that is removed. In BCS, only the part of the breast that contains the tumor and the surrounding normal tissue are removed[66].

Mastectomy refers to the removal of the entire breast and sometimes neighboring tissues. Women who undergo mastectomy can have breast reconstruction with saline or silicone implants or tissue from another part of the body[66].

Both BCS and mastectomy are commonly followed by removal and analysis of axillary lymph nodes to determine whether the disease has spread and whether further treatment is required.

1.4.2 Radiation therapy

Radiation therapy is generally used after surgery to eliminate tumor cells that remain in the breast or nearby tissues, thereby reducing the risk of recurrence[67]. Radiation therapy can be administered as external beam radiation, where the radiation comes from a machine and is placed on the affected area, or as internal radiation therapy, where radioactive pellets are administered through catheters[68].

1.4.3 Chemotherapy

Chemotherapy targets proliferating cells and the benefit of this therapy is dependent on factors such as tumor size, lymph node affection, and HR and HER2 status. This systemic therapy is usually given in cycles over 3-6 months and is effective in patients with aggressive tumors such as TN and HER2+[69]. Patients with metastatic BCa are treated with chemotherapy. Although metastasis is generally considered incurable, this treatment delays disease progression and can

result in symptom palliation. For patients with early BCa, chemotherapy is administered after breast surgery or lymph node resection. However, if the tumor is large, chemotherapy may be used in the neoadjuvant setting[70]. Several chemotherapy drugs have been developed for BCa treatment and can be used as single cytotoxic agents or in combination. Additionally, chemotherapy drugs can be alternated during BCa progression in order to overcome resistance to therapy[71].

Alkylating agents include cyclophosphamide and platinum analogs such as cisplatin, carboplatin and oxaliplatin. These compounds act by alkylating the DNA, cross-linking DNA strands and disrupting DNA replication. They are effective in BCa treatment but their use is limited due to severe side effects such as gonadal and bladder toxicity, nephrotoxicity, myelosuppression and neurotoxicity[72,73].

Anthracyclines include doxorubicin and epirubicin. Derived from the antibiotic rhodomycin B, these drugs also disrupt DNA replication through a variety of mechanisms, including DNA intercalation, inhibition of topoisomerase-II function, and generation of reactive oxygen species (ROS). These compounds are very effective cytotoxic agents, but their administration is limited due to high cardiotoxicity[74].

Taxanes include paclitaxel and docetaxel and act by inhibiting microtubule depolymerization, leading to mitotic arrest, chromosome missegregation and abnormal multipolar spindles[75,76]. Docetaxel has superior efficacy to paclitaxel but generates more toxicity. The major toxicity for both drugs is neutropenia although there are other non-hematologic complications such as neurotoxicity, fluid retention and fatigue[77,78].

Antimetabolites include the pyrimidine antagonist 5-fluorouracil, the purine antagonist 6-mercaptopurine, and the folic acid antagonist methotrexate, among others. These drugs either mimic the natural purines and pyrimidines required for DNA synthesis or block the production of major nucleotide metabolites to disrupt DNA synthesis. The major side effects associated with antimetabolites are bone marrow, kidney and gastrointestinal tract toxicity[79,80].

Topoisomerase-I inhibitors include the camptothecin analogs irinotecan, etirinotecan and topotecan and act by blocking the activity of the topoisomerase-I enzyme, required to relax

DNA supercoiling generated by transcription, replication and chromatin remodeling. Their main adverse effects are neutropenia, diarrhea, nausea and vomiting[81].

1.4.4 Hormonal therapy

Hormonal therapy is the standard choice for HR+ BCa and it works by blocking E2 signaling, thereby impairing tumor growth. There are different options for hormonal therapy and the strategy of choice is guided by menopausal status. Premenopausal women with a high risk of recurrence are often treated with hormonal therapy combined with ovarian suppression[82].

Although hormonal therapy is extremely effective, most HR+ patients develop resistance to this treatment. This resistance can be de novo, when patients do not respond to the first line of treatment, or acquired, when resistance appears after an initial response. Hormonal therapies can be alternated or given in combination to minimize or overcome resistance[65].

Selective estrogen receptor modulators (SERMs) include tamoxifen and raloxifene. Tamoxifen competes with E2 for binding to its receptor, inducing a conformational change that impairs the function of the receptor. Tamoxifen is effective in pre- and post-menopausal patients and has been the hormonal therapy of choice for ER+ BCa for the last 30 years. After 5 years of adjuvant tamoxifen treatment the risk of BCa recurrence and death is greatly decreased[83]. However, a clinicopathologic tool has recently been developed to estimate risk of late recurrence in order to identify patients that might benefit from an extension of tamoxifen treatment[84]. Tamoxifen has a protective effect on bone mineral density but is associated with adverse effects such as hot flashes and increased risk of thromboembolic events and endometrial cancer[65]. Raloxifen is approved only for postmenopausal women and is mostly used to prevent osteoporosis as it has more potent antiresorptive capacity compared to tamoxifen[85].

Aromatase inhibitors include letrozole, anastrozole and exemestane and are effective only in postmenopausal women. All three drugs are equally effective and act by blocking aromatase, a key enzyme in estrogen synthesis, thus lowering serum E2 concentration. Adverse effects associated with the use of aromatase inhibitors are hot flashes and arthralgias or myalgias. These inhibitors are more effective than tamoxifen, although treatment decisions are guided by menopausal status and adverse effects[86].

Gonadotropin-releasing hormone analogs include leuprolide acetate and goserelin and are used to suppress ovarian function in premenopausal women, inducing a menopause-like condition. GnRH analogs downregulate pituitary GnRH receptors and suppress LH and FSH release, thereby blocking E2 production in the ovaries. An alternative way to induce menopause is through oophorectomy, although this approach has largely been replaced by GnRH analog therapy as both approaches are equally effective[87]. After inducing menopause, patients are treated with tamoxifen or aromatase inhibitors, the latter showing higher rates of benefit. The main adverse effects of GnRH analogs are those of menopause, including hot flashes and loss of bone mineral density[88].

Selective estrogen receptor degraders (SERDs) exert their mechanism of action by binding to the ER and inducing misfolding, which ultimately leads to proteasome-dependent protein degradation. SERDs include fulvestrant, the only clinically approved drug of this class. Fulvestrant is a potent and highly specific ER degrader with no estrogen agonistic effects, therefore its side effects are more favorable compared to tamoxifen. Fulvestrant is recommended for postmenopausal women and associated with increased long-term survival, particularly in patients that receive previous adjuvant hormonal therapy[89]. Other SERD molecules have been successfully developed and some of them, such as AZD9496, GDC-0810 and LSZ102, are currently being tested in clinical trials as new therapies for metastatic BCa[90–92].

1.4.5 Targeted therapy

Targeted therapies consist of blocking the action of a specific protein that contributes to cancer cell growth and survival.

HER2 inhibitors are given to patients with HER2 overexpression and are generally administered in combination with chemotherapy. The development of HER2-targeted therapy is considered one of the greatest advances in BCa treatment. In this regard, multiple drugs are used to treat BCa by blocking HER2 signaling and can be used in combination depending on acquired resistance to therapy. Trastuzumab is a monoclonal anti-HER2 antibody that targets the extracellular juxtamembrane domain of the HER2 receptor, thus inhibiting its function. This drug was the first approved as a HER2 inhibitor after clinical trials demonstrated that it significantly improved PFS and OS for women with HER2+ BCa[93,94]. Pertuzumab is a

monoclonal antibody that targets the HER2 dimerization domain, preventing heterodimerization and subsequent HER2 signaling. When combined with trastuzumab and chemotherapy, this drug significantly improves PFS in HER2+ BCa patients[95]. Given its higher cost and toxicity, pertuzumab is currently recommended only for high-risk patients[65]. Lapatinib is a reversible and selective inhibitor of the intracellular tyrosine kinase domains of EGFR and HER2 receptors[96]. Finally, neratinib, an oral small tyrosine kinase inhibitor that irreversibly binds to the intracellular signaling domain of multiple HER2 family members, has been approved for patients with advanced BCa that have received two or more prior anti-HER2 treatment regimens[97].

Cyclin-dependent kinase 4/6 (CDK4/6) inhibitors abemaciclib, palbociclib and ribociclib. These drugs cause cell cycle arrest by targeting CDK4 and CDK6, two enzymes with key roles in cell division. When used in combination with hormonal therapy, CDK4/6 inhibitors lead to a significant increase of PFS in HR+ HER2- metastatic BCa and their main possible side effects include myelosuppression and liver toxicity[98–100].

Poly (ADP-ribose) polymerase (PARP) inhibitors block PARP enzymes, which are involved in cellular processes such as transcription, replication, recombination, and DNA repair. Many tumors rely on PARP-mediated DNA repair to survive and are sensitive to its inhibition[65]. PARP inhibitors include olaparib and talazoparib and are used to treat HER2- metastatic BCa in people with *BRCA1/2* mutations who have received previous chemotherapy or hormonal therapy when the tumor is HR+[101,102].

Phosphoinositide 3-kinase (PI3K) inhibitors target PI3K, which regulates important functions such as cell growth, proliferation, differentiation, motility and survival and is commonly upregulated in cancer cells. Alpelisib is used in combination with hormonal therapy to treat HR+ HER2- metastatic breast tumors that carry a *PIK3CA* gene mutation. Adverse effects of Alpelisib include hyperglycemia, fatigue and diarrhea[103].

Mammalian target of rapamycin (mTOR) inhibitors are directed to the mTOR kinase, which is upregulated in many types of cancer, including BCa. mTOR integrates multiple intracellular and extracellular signals in order to regulate cell metabolism, growth, proliferation and survival. Everolimus is the main mTOR inhibitor and is used in combination with

hormonal therapy to treat advanced HR+ HER2- metastatic BCa in postmenopausal women. Possible side effects of everolimus include mouth ulcers, infections, fatigue and diarrhea[104,105].

1.4.6 Immunotherapy:

Immunotherapy is an emerging approach for BCa treatment. It seeks to stimulate the patient's immune system to recognize and eliminate cancer cells. At the moment, immunotherapy in BCa is not as promising as in other types of cancer. However, some breast tumors benefit from this therapy[65]. Immunotherapy drugs include atezolizumab, a checkpoint inhibitor that targets programmed death-ligand 1 (PD-L1) and blocks activation of the T-cell receptor programmed death 1 (PD-1), which leads to suppression of immune evasion and enhanced T-cell-mediated immune response. In combination with chemotherapy, atezolizumab shows prolonged PFS among patients with advanced PD-L1+ TN BCa. The adverse effects of this therapy include diabetes, thyroid problems, lung inflammation, liver problems and colitis[106].

2. METASTASIS

Metastasis is the main cause of death in cancer patients and is considered a hallmark of cancer[107]. To colonize distant organs, cancer cells need to acquire several specialized functions to overcome many different obstacles (Figure 5). Cancer cells must cross endothelial barriers to leave the primary tumor and enter into the vasculature, survive the mechanical forces of the bloodstream, arrest in capillaries at distant sites, pass through endothelial barriers again to leave the circulation, adapt to supporting niches to colonize the host tissue, survive in a latent state and eventually acquire the ability to proliferate and generate a secondary tumor. Additionally, during the metastatic process, cancer cells must evade the immune system in order to survive[108,109]. This steeplechase that cancer cells undergo to generate distant metastases is known as the "metastatic cascade".

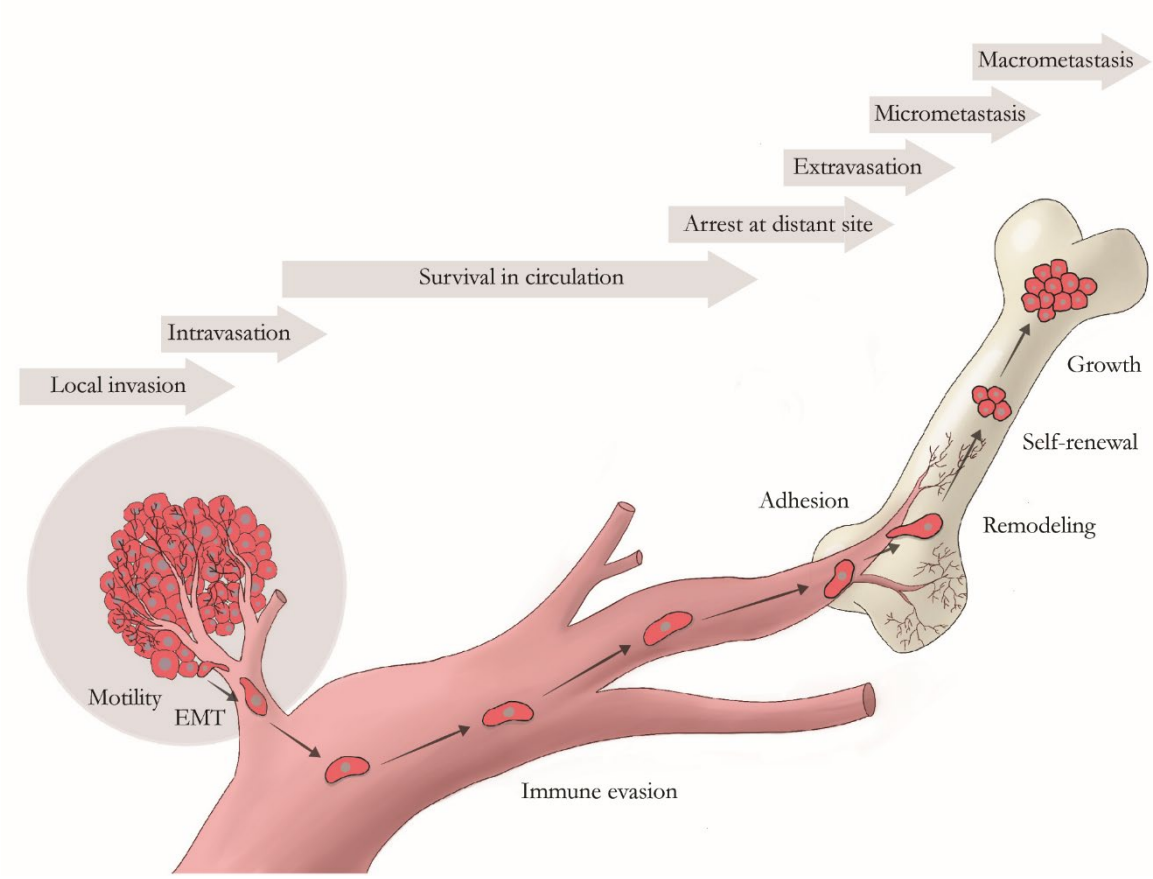


Figure 5. The metastatic cascade. Tumor cells from a primary tumor must acquire specific cellular functions (written in black) in order to progress through the sequential steps of the metastatic cascade (illustrated as grey arrows) and colonize distant organs. Adapted from Gomis and Gawrzak (2017).

All these obstacles make metastasis a highly inefficient process in which most of the cells that leave the primary tumor die. The number of circulating tumor cells (CTCs) found in the blood of cancer patients goes far beyond the number of metastatic lesions that these patients develop[110]. Moreover, tumor cells can remain in the bone marrow of cancer patients for years but only half of these patients will develop overt metastases[111]. However, despite the inefficiency of the metastatic process, thousands of cancer cells might have already seeded distant organs once the primary tumor has been diagnosed, thereby highlighting the importance of early cancer detection.

Of note, current therapies for BCa focus on cell proliferation and latent disseminated cancer cells are able to escape treatment. Thus, to design therapies that target cell dissemination and prevent the development of metastasis, it is important to understand the molecular mechanisms that enable distant tissue colonization.

2.1 The metastatic cascade

The metastatic cascade includes the following steps: local invasion, intravasation, survival in circulation, arrest, extravasation, adaptation to supportive niches, tumor dormancy and overt colonization.

2.1.1 Local invasion

The process by which cancer cells surpass the primary tumor boundaries and penetrate the surrounding stroma is known as local invasion. Cancer cells must have special characteristics to achieve invasion and migration through the stroma. For instance, they can secrete matrix metalloproteinases (MMPs) to degrade and remodel the ECM as a way to facilitate their movement across the stroma[112]. Migration can also occur independently of proteolytic ECM remodeling through cytoskeletal rearrangements that confer cancer cells a highly motile amoeboid phenotype[113,114]. Malignant cells can migrate as single cells, along collagen fibers or nerve fibers or collectively as a cohesive group with properties such as collective polarization and force generation[115–118].

The tumor microenvironment plays a relevant role in the invasion process. Cancer cells can associate with macrophages and use them as a source of proteases to support migration and invasion[119]. Furthermore, fibroblasts can adhere to groups of cancer cells and drag them across the ECM through physical forces, resulting in cooperative tumor invasion[120]. The formation of new blood vessels, a process named angiogenesis, is essential for the growth of cancer cells but also for allowing them to reach the circulation and spread to distant tissues. During tumor progression, an angiogenic switch is activated by factors that bind to vascular endothelial cell receptors to orchestrate the growth of new blood vessels. These factors, which include VEGF and members of the fibroblast growth factor (FGF) family, are produced by cancer cells and also by inflammatory immune cells of the tumor microenvironment[107].

2.1.2 Intravasation

Intravasation refers to the entry of cancer cells into a blood or lymphatic vessel. To this end, cancer cells can undergo an epithelial-to-mesenchymal transition (EMT), a physiological process that is essential in embryonic development where cells lose epithelial features such as cell-cell adhesion and cell polarity and acquire migratory, invasive and stem cell properties, characteristic of mesenchymal cells. Cancer cells take advantage of this plastic program to penetrate the vasculature, resist stress and migrate to distant organs[121].

Cancer cells that undergo EMT lose expression of E-cadherin, the core protein of adherens junctions. This leads to disruption of cell-cell adhesion, loss of cell polarity, disconnection with the basement membrane and reorganization of the actin stress fibers of the cytoskeleton, thereby confering cancer cells a spindle-like mesenchymal morphology. Moreover, the expression of the mesenchymal markers N-cadherin, vimentin and fibronectin is activated. Cancer cells can trigger EMT through various signaling pathways, including TGF-β, Wnt, Notch and mitogenic growth factors. However, this transition can also be induced by the tumor microenvironment[122].

Despite EMT importance in metastatic dissemination, recent studies suggest that it might not be indispensable for the establishment of metastasis, as dissemination also occurs with retention of epithelial cell identity such as E-cadherin expression[123–125]. It has been proposed that cancer cells trigger a partial EMT, retaining some epithelial features, or that only those

cells located at the leading edge of clusters of migrating cells display mesenchymal features associated with EMT activation[122].

2.1.3 Survival in circulation

Upon entry into the bloodstream, CTCs are highly exposed to mechanical forces, the immune system and oxidative stress. CTCs associate with platelets to protect themselves from the mechanical forces of the bloodstream, and rely on platelet-derived signals for efficient metastasis[126]. Moreover, platelets adhered to the surface of CTCs prevent their recognition by the immune system[127]. Natural killer (NK) cells and macrophages are critical tumor suppressors; hence, CTCs need to evade their action when traveling through the bloodstream. For instance, it has been proven that down-regulation of toll-like receptor (TLR) 2 and TLR4, responsible for NK cell activation, in monocytes of breast and colorectal cancer patients correlates with increased number of CTCs[128]. Other mechanisms of CTC-induced NK cell inhibition include the production of inhibitory cytokines, the interaction and inhibition of killer cell immunoglobulin receptors (KIRs) on the NK cell surface, and increased platelet activation[129]. The loss of a matrix support and an increased oxygen tension in the circulation lead to high ROS in CTCs, and to cope with oxidative stress, these cells undergo metabolic changes[130]. For instance, melanoma cells rely on nicotinamide adenine dinucleotide phosphate (NADPH)-generating enzymes in the folate pathway to reduce oxidative stress and survive in the circulation[131]. Furthermore, in CTCs from breast, prostate and lung cancers, increased ROS trigger β-hemoglobin expression to suppress intracellular oxidative stress, thus increasing cell survival[132].

Interestingly, clusters of CTCs derived from the primary tumor and not from intravascular aggregation have been found in the blood of cancer patients and it has been demonstrated that these clusters have a 23-50 fold increased metastatic capacity[133].

2.1.4 Arrest

CTCs arrest on the wall of a blood vessel is caused mainly by the structure of the capillaries. When capillaries become thinner, CTCs are not able to move freely in the bloodstream and they arrest before entering a new tissue. A study based on *in vivo* imaging showed that major

CTC extravasation occurs in small capillaries and at branch points[134]. Additionally, CTCs associated with platelets can create microemboli, which facilitate arrest and adherence to the endothelium through recognition of E-selectin expressed on activated endothelial cells[135].

2.1.5 Extravasation

Extravasation is defined as the exit of cancer cells from the vasculature to invade distant tissues. Depending on the structure of the capillaries, extravasation can be challenging. Capillaries in the bone marrow and liver are fenestrated, with a discontinuous basal lamina that facilitates CTCs entry into the parenchyma. In contrast, lung capillaries are rich in tight junctions and have a basement membrane, and the brain has the blood-brain barrier (BBB), thereby hindering penetration[136].

Arrested microemboli formed by CTCs and platelets might start to grow in the lumen of blood vessels and eventually break the vascular wall and reach a distant tissue[137]. Alternatively, arrested CTCs must release extracellular cues, use pairs of ligand-receptor molecules or associate with other circulating cells in order to extravasate.

CTCs can release extracellular vesicles to deliver microRNAs (miRNAs) that target tight junction proteins in the BBB, thereby promoting extravasation to the brain tissue[138–140]. VEGF expression by CTCs increases permeabilization in the lung endothelium by disrupting the vascular endothelial (VE)-cadherin-β-catenin complex at adherens junctions and its blockade suppresses CTC extravasation *in vivo*[141]. Moreover, the release of vascular cell adhesion molecule 1 (VCAM1) in melanoma cells and angiopoietin-like 4 (ANGPTL4) in BCa cells disrupts the capillary endothelium and promotes lung colonization[142,143].

Multiple adhesion receptors expressed on the surface of CTCs bind to endothelial cells for extravasation. The transmembrane glycoprotein Mucin 1 (MUC1) on CTCs binds to endothelial intercellular adhesion molecule 1 (ICAM1) and mediates cell migration through the vasculature[144]. Furthermore, the cell surface glycoprotein CD44 on CTCs binds to E-selectin expressed by endothelial cells to favor migration across the endothelium[145,146].

Arrested CTCs and endothelial cells themselves can release cytokines and chemokines to facilitate extravasation, but they also further interact with other circulating cells such as platelets, monocytes, neutrophils and NK cells, which secrete factors to increase endothelial

barrier permeability[147]. Platelets secrete adenine nucleotides that activate endothelial purinergic receptors by adenosine triphosphate (ATP) to increase vascular permeability and also release TGF-β to trigger EMT in CTCs, which results in a more invasive phenotype suited for extravasation and subsequent movement through the distant tissue[148]. Moreover, CTCs recruit macrophages through the release of the C-C motif chemokine ligand 2 (CCL2), which in turn secretes cues to increase vascular permeability[149].

2.1.5 Adaptation to supportive niches

Cancer cells that managed to invade distant organs, called disseminated tumor cells (DTCs), must adapt to the new harsh environment in order to generate a micrometastasis. Supportive niches are prone to harboring DTCs by providing specific signals that sustain tumor cell survival and proliferation. These niches, which include stem-cell niches, perivascular niches, pre-metastatic niches and *ad hoc* niches, are used by malignant cells as a shelter.

Stem-cell niches support normal adult stem cells by providing factors to sustain self-renewal and differentiation and to maintain a balance between proliferation and quiescence. DTCs can hijack these specialized niches and benefit from their signals to ensure their own survival. For instance, it has been shown that prostate cancer cells compete with hematopoietic stem cells in the bone marrow for occupancy of hematopoietic stem-cell niches[150].

Perivascular niches surround blood vessels and provide supportive signals for angiogenesis. Cancer cells remain close to endothelial cells after extravasation and benefit from the attachment, oxygen, nutrients and paracrine factors that the activated endothelium supplies. For example, it has been demonstrated that metastatic breast and lung cells express L1 cell adhesion molecule (L1CAM) to remain close to brain capillaries for metastatic outgrowth and also high levels of serpins to prevent L1CAM inactivation by the reactive brain stroma[151].

Pre-metastatic niches are created by signals from the primary tumor that influence the microenvironment of distant organs and make them more suitable for cancer cells before their arrival. These signals include tumor-derived inflammatory cytokines, exosomes and ECM-remodeling enzymes.

Inflammatory cytokines recruit immune cells to prepare a distant tissue for metastatic cells. Colorectal cancer cells secrete VEGF in order to stimulate tumor-associated macrophages to

produce C-X-C Motif Chemokine Ligand 1 (CXCL1). An increase in CXCL1 recruits liver-infiltrating C-X-C Motif Chemokine Receptor 2 (CXCR2)-expressing immune cells, which will form a pre-metastatic niche in the liver[152]. Moreover, C-X-C Motif Chemokine Receptor 4 (CXCR4) expression in colorectal cancer cells controls the expression of interleukin 10 (IL10) and CXCL1, which favors liver metastasis formation[153].

Exosomes are small membrane vesicles released by fusion of multivesicular bodies with the plasma membrane that are loaded with proteins, lipids DNA, mRNA or miRNA to mediate local and systemic cell communication. Melanoma-derived exosomes induce vascular leakiness and inflammation at pre-metastatic bone marrow sites by delivery of the MET receptor tyrosine kinase. Moreover, pancreatic-derived exosomes loaded with macrophage migration inhibitory factor (MIF) promote pre-metastatic niche formation in the liver through stimulation of TGFβ secretion, which triggers fibronectin production by hepatic stellate cells and subsequent macrophage recruitment[154,155].

Ad hoc niches can be established by DTCs by producing elements of stem-cell niches themselves. Metastatic BCa cells induce stromal expression of periostin (POSTN) to increase Wnt signaling and initiate lung colonization[156]. Furthermore, BCa cells that infiltrate the lungs can express tenascin C (TNC) to enhance the expression of stemness genes and therefore their own metastasis-initiating capacity[157].

2.1.6 Tumor latency

Tumor latency, also known as tumor dormancy, refers to the capacity of tumor cells to remain in a latent state at distant tissues without giving rise to an overt metastasis and without manifestation of clinical symptoms. This concept is supported by the observation that patients can develop metastases years after the removal of the primary tumor and by the detection of DTCs in the bone marrow of non-metastatic patients. Dormant cells can go undetected for long periods of time and generate resistance to therapies[158].

Two models of tumor latency have been postulated: solitary cell dormancy and tumor mass dormancy. In the former, a single DTC enters a quiescence state by arrest at the G0/G1 phase of the cell cycle. This arrest can be induced by factors present in the host tissue microenvironment but can also be self-imposed by the tumor cell[158]. Once proliferation has

stopped, DTCs alter signaling pathways to ensure metabolic homeostasis. For instance, DTCs in BCa display altered lipid metabolism and elevated ROS[159]. Alternatively, in tumor mass dormancy, the growth of micrometastases is inhibited by a balance between proliferation and apoptosis. This balance is regulated by the tumor microenvironment through signals that inhibit proliferation, through restricted blood supply or by the action of the immune system, although again, tumor cells can undergo reprogramming to maintain the tumor mass dormant. One example of the proliferation-apoptosis balance is the release of factors such as VEGF by tumor cells in micrometastases to induce angiogenesis and sustain proliferation while the tumor microenvironment compensates with high expression of anti-angiogenic factors such as thrombospondin 1 (TSP1)[160,161]. The proliferation-apoptosis equilibrium is also strictly regulated by the immune system. T cells and NK cells are recruited to the metastatic site to control the spread of the tumor mass. However, as a counterbalance, tumor cells evade the action of these immune cells by expressing several inhibitory signals[158].

The duration of tumor cell dormancy varies between cancer types, being extremely short for the most aggressive types and lasting several years in the least aggressive. In lung cancer, metastasis can appear a few weeks after diagnosis whereas breast and prostate cancer cells can remain dormant for decades[158].

2.1.7 Overt colonization

As mentioned before, EMT favors invasion, migration, intravasation and extravasation, but the ability of cancer cells to grow when they reach distant organs is likely to be regulated by the reversion of EMT, known as mesenchymal-to- epithelial transition (MET). Factors that induce EMT are associated with reduced cell proliferation- Therefore, DTCs reverse the EMT program in order to enhance proliferation and favor the establishment of macrometastases[162]. Additionally, stromal and autocrine mediators, physical contact with stromal cells and genetic and epigenetic alterations activate signaling pathways in DTCs to support metastatic growth. These pathways include stem-cell support pathways such as Wnt, TGFβ, bone morphogenetic protein (BMP), Notch and signal transducer and activator of transcription 3 (STAT3), cell metabolism and survival pathways such as PI3K-AKT, mitogen-activated protein kinase (MAPK) and hypoxia-inducible factor (HIF), positional and mechanical pathways such as

Hedgehog and Hippo, and inflammatory pathways such as nuclear factor kappa-light-chain-enhancer of activated B cells (NFκB) and STAT1[163].

Eventually DTCs acquire the necessary traits to overcome the pressure of the microenvironment or reprogram to reverse the slow cycling features of dormancy and they are able to give rise to macrometastases.

2.2 Epigenetic drivers of cancer metastasis

Cancer cells from a primary tumor harbor a wide variety of genetic mutations that could allow the acquisition of the traits necessary to successfully metastasize[164]. However, only less than 0.01% of the primary tumor cells that reach the circulation are able to metastasize, and mutations that specifically drive metastasis have not yet been identified[165]. Several studies have described molecules that endow cancer cells with features to achieve distant tissue colonization. However, the mechanisms by which these cells acquire these specific metastatic traits remain poorly understood[163,166]. Interestingly, it is believed that some metastatic features might be reversible and therefore not driven by mutations. Evidence shows that metastatic cells can be reprogrammed by specific microenvironments and that they can reverse their metastatic phenotype, thus becoming less aggressive[167,168]. Moreover, when primary tumors are compared to metastases, there is little heterogeneity in driver mutations but considerable epigenetic reprogramming, and it has been reported that changes in chromatin accessibility favors distant tissue colonization during metastatic progression[169]. Hence, metastasis seems to be a complex process involving many pathways, some of which might be activated by mutations, while others are controlled by epigenetic alterations. To illustrate this statement, there is evidence that tumor initiation mutations in VHL activate the hypoxia-inducible transcription factor 2A (HIF2A) transcription factor, but this does not automatically translate into acquisition of metastatic features. Nevertheless, in later stages of cancer progression, epigenetic modulation of HIF2A target genes leads to the emergence of metastatic traits[170].

To comprehend epigenetic modulation, it is essential to understand DNA structure and packaging (Figure 6). DNA is packed into chromosomes in order to fit in the nucleus of each cell but also to prevent DNA damage during cell division. A DNA molecule wraps around the core histone proteins H2A, H2B, H3 and H4, forming a nucleosome. The linker histone H1

binds to the nucleosome in order to lock DNA into place and allow nucleosome coiling into compact fibers called chromatin, which will in turn loop and fold to form chromosomes. Chromatin structures are divided into euchromatin and heterochromatin and they have important implications in gene expression. Euchromatin refers to decondensed genome regions where DNA is more accessible to transcription factors and transcription machinery and genes can be actively transcribed. In contrast, heterochromatin corresponds to highly condensed regions of the genome that are hardly accessible to molecules involved in the transcription process, and therefore transcriptionally inactive[171].

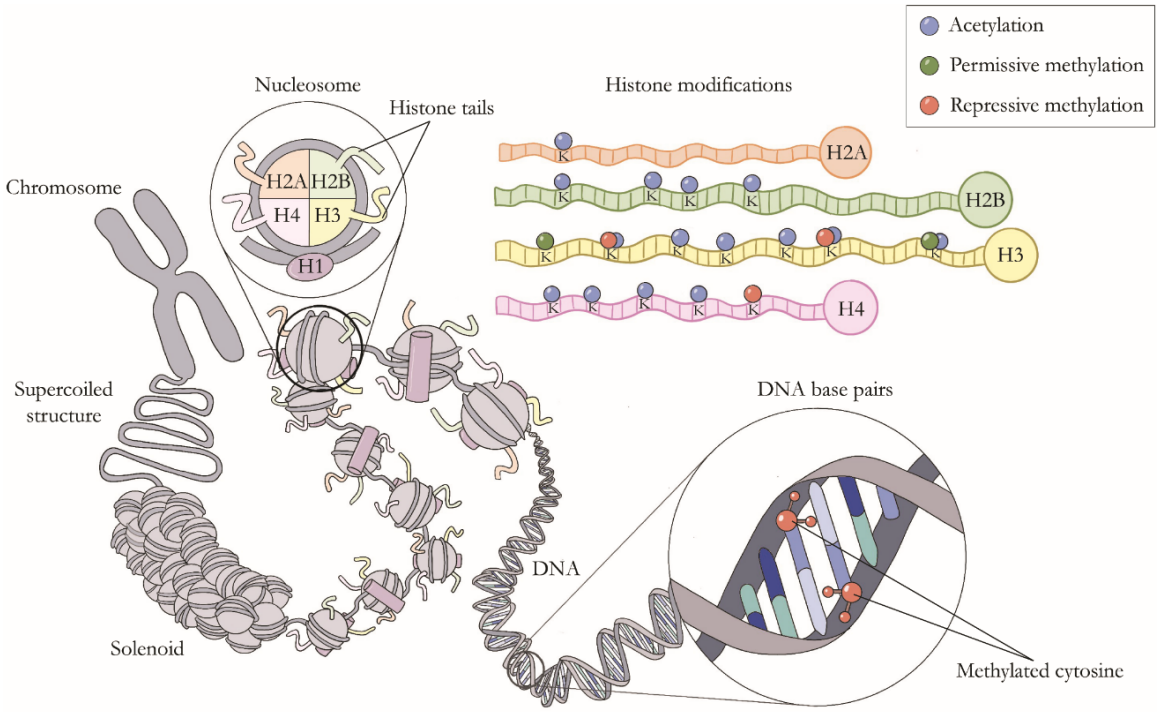


Figure 6. Schematic representation of DNA packaging and epigenetic modifications. DNA wraps 1.65 times around a histone octamer formed by four pairs of histone H2A, H2B, H3 and H4 to form nucleosomes, the basic repeating unit of chromatin. Adjacent nucleosomes are connected through linker DNA, bound to the histone H1. Nucleosomes fold into a 30nm chromatin fiber structure named solenoid which is further compacted into chromosomes. Histones have N-terminal tails protruding from the nucleosome that can undergo post-translational modifications such as acetylation and methylation and DNA can be methylated at cytosine bases of 5'-cytosine-phosphodiester-guanine (CpG) dinucleotides. These epigenetic modifications result in modulation of the chromatin structure that translate in genes expression changes. Genes within tightly packed chromatin regions are silenced whereas those in decondensed areas can be actively transcribed. The most common epigenetic modifications of histone lysine (K) residues are depicted. Adapted from Weaver *et al.* (2017).

Chromatin structure can change due to epigenetic alterations, which include DNA methylation and histone modifications. Epigenetic alterations are defined as heritable changes that modify chromatin structure without altering the DNA sequence, resulting in modification of DNA accessibility and subsequent changes in gene expression. Chromatin-remodeling proteins are responsible for epigenetic alterations and are highly relevant in normal gene regulation, as well as in pathological processes. In cancer, deregulation of chromatin due to aberrant expression or epigenetic modulation of chromatin-remodeling proteins confers cancer cells the ability to reprogram their genome and develop oncogenic traits.

DNA methylation occurs at cytosine bases of cytosine-phosphodiester-guanine (CpG) dinucleotides. CpG-rich regions, called CpG islands, are located mainly at the 5' end of the regulatory region of many genes, and their methylation is associated with closed chromatin and transcriptional silencing. DNA methylation is controlled by DNA-methylating enzymes, called DNA methyltransferases (DNMTs), and DNA demethylating enzymes. Evidence shows that alterations in DNA methylation patterns are linked to tumorigenesis and metastasis[172]. In BCa, the presence or absence of coordinated hypermethylation at specific genes determines an epigenomic profile associated with low metastatic risk, and DNA methylation aberrations at these specific sites correlate with poorer outcome[173]. Moreover, silencing of metastasis suppressor genes such as cadherin 11 (CDH11) by DNA methylation is found in lymph node metastases but not in primary tumors of melanoma or in head and neck cancer cells[174].

Histones can be modified at the N-terminal flexible tails as well as in the core domain, to regulate the compactation state of chromatin. Common histone modifications include acetylation (ac) and methylation (me) of lysines (K) and arginines (R), phosphorylation of serines (S) and threonines (T), ubiquitylation, and sumoylation of lysines. These posttranslational modifications not only have an effect on the histone-DNA interaction but also generates docking sites or modulate the affinity of nuclear chromatin-associated proteins for chromatin. Posttranslational modifications are reversible and are regulated by many different enzymes.

The acetylation of histone lysines neutralizes the positive charge of histone tails, thereby altering histone-DNA interactions, and it is generally associated with transcriptional activation. The process of histone acetylation is highly reversible and is tightly controlled by HATs and HDACs. The deregulation of HAT and HDAC expression is also associated with tumorigenesis and poor prognosis. For instance, the HDAC Sirtuin 1 (Sirt1) is involved in the

acquisition of metastatic traits by cooperating with EMT transcription factors and subsequent EMT induction[175].

The implications of histone methylation depend on the type of residue - lysine or arginine - and the specific position of this residue in the histone tail. For instance, methylation at H3K4, H3K36 and H3K79 is associated with transcriptional activation, whereas methylation at H3K9, H3K27 and H4K20 is linked to transcriptional repression. Lysine and arginine can be monomethylated or dimethylated but lysine can also be trimethylated. Histone methylation is regulated by histone methyltransferases (HMTs) and histone demethylases (HDMs) and their deregulation is associated with pathological processes. For instance, lysine specific demethylase 1 (LSD1), also known as KDM1A, is upregulated in multiple types of cancer. The association of LSD1 and the estrogen or the androgen receptor illustrates an example of the coordinate interaction between chromatin-remodeling proteins with specific transcription factors. LSD1 interacts with these nuclear receptors to form chromatin-associated complexes, where it removes repressive marks by demethylation of H3K9 in order to de-repress nuclear receptor target genes[176,177]. LSD1 inhibitors have been reported to attenuate tumor progression *in vitro* in many types of cancer and are currently being tested in clinical trials for small-cell lung carcinoma (SCLC) and acute myeloid leukemia (AML)[178].

Some histone modifications are associated with DNA regulatory elements. H3K9ac and H3K9me3 are two common modifications at promoters, which are regulatory elements where transcription is initiated, and H3K4me1 and H3K27ac at active enhancers, which are regulatory elements farther away from the start site formed by clusters of transcription factor binding sites that amplify the transcription initiated at promoters. Epigenetic alterations at promoter and enhancer regions regulate gene expression and their alteration is associated with the acquisition of metastatic traits[172].

Most chromatin-modifying enzymes interact with coactivators and corepressors to form functional complexes that increase their activity. An example of these functional complexes are the two repressive PcG protein complexes, polycomb repressive complex 1 (PRC1) and PRC2 and the antagonistic trithorax group (TrxG) protein complexes, associated with the activation of gene expression[179]. Other chromatin remodeler complexes require energy from ATP hydrolysis for their functions and include the following families: SWI/SNF, imitation switch (ISWI), INO80, and chromodomain-helicase DNA-binding (CHD), the latter including the Mi-2/ nucleosome remodeling and deacetylase (NurD) complex[180,181]. Alterations in

chromatin-remodeling complexes are frequent in cancer. For instance, inactivating mutations in AT-rich interaction domain 1A (ARID1A), a subunit of the SWI/SNF chromatin remodeling complex, are frequent in ER+ BCa and loss of ARID1A expression is associated with loss of luminal cell identity and resistance to hormonal therapy[182].

In summary, chromatin structure is rigorously controlled by chromatin remodelers and its alteration is extremely relevant in cancer metastasis, as it can trigger the activation of pathways required for metastatic progression such as EMT. Activation of EMT depends on multiple signals from the tumor microenvironment or on paracrine signaling factors. This signaling triggers the expression of several transcription factors such as TWIST, SNAIL, SLUG and zinc finger E-box binding homeobox 1 (ZEB1), which are known to control EMT programs. However, the acquisition of mesenchymal traits is not permanent, as cells need to reverse to an epithelial phenotype through MET once established at a distant tissue. This reversibility requires the reprogramming of gene expression, which involves several epigenetic regulators. For instance, EMT transcription factors can interact with PcG proteins at the E-cadherin promoter, which will result in histone methylation and E-cadherin transcriptional repression. The aforementioned EMT transcription factors can also interact with HDACs from the NuRD repressive complex or histone demethylases such as LSD1 in order to orchestrate EMT activation or repression[183]. Moreover, a recent example of acquisition of metastatic traits by chromatin deregulation is L1CAM expression by cancer cells. When cancer cells detach from the epithelium in order to reach the blood or lymphatic circulation, the loss of epithelial cell-cell contacts triggers L1CAM expression by transcriptional derepression through the downregulation of the epigenetic regulator Repressor Element-1 Silencing Transcription factor (REST). Upon extravasation, cancer cells use LCAM1 to adhere to the surface of blood capillaries and activate essential pathways for metastatic outgrowth. LCAM1-dependent growth is a mechanism used for wound healing and it is leveraged by metastatic cells[184].

Epigenetic alterations in cancer have potential clinical use. Detection of hypermethylated promoters of tumor suppressor genes or specific histone-modification patterns in biological fluids and tissue biopsies can be used as biomarkers for early detection, prognosis and therapy response prediction, although its use in routine clinical practice requires cost-effective techniques[185,172]. For instance, 90% of primary prostate tumors have hypermethylation of regulatory sequences at the glutathione S-transferase pi (*GSTP1*) gene and evidence shows that this hypermethylation can be detected in urine samples from prostate cancer patients[186,187].

Hypermethylation of death-associated protein kinase (DAPK), $p16^{INK4a}$ and epithelial membrane protein 3 (EMP3) are associated with poor outcome in lung, colorectal and brain cancer respectively, which makes these epigenetic alterations interesting prognosis biomarkers[188]. Moreover, hypermethylation of MGMT predicts favorable response to chemotherapy in patients with gliomas[189,190].

Cancer is a disease of altered gene expression and it occurs as a result of dysregulation of transcription factors as well as altered epigenetic regulation. Transcription factors are crucial for cancer progression. However, their inhibition is challenging because they do not have enzymatic activity and it is hard to inhibit protein-protein interactions using drugs. Conversely, chromatin remodelers are ideal drug targets as they either have enzymatic activity or recognize DNA or histone covalent modifications[191]. For this reason, inhibitors targeting chromatin-remodeling proteins are being explored in multiple clinical trials, and seven epigenetic drugs, targeting DNMTs or HDAC, have already been approved by the US Food and Drug Administration (FDA) for the treatment of hematological malignancies[192].

2.3 Metastatic organ tropism

The spread of distant metastases to specific organs is known as metastatic organ tropism and it is not an aleatory process (Table 4). Some cancer types spread mainly to one particular site, for example prostate cancer to bone or pancreatic cancer and uveal melanoma to liver, while other types such as breast and lung cancers or melanoma can colonize many different organs. Moreover, some types of cancer metastasize at different organs simultaneously, while others do this in a sequential manner. Colorectal cancer, for example, frequently metastasizes first to the liver and after to the lungs and the brain[193]. Metastases are rarely found in some tissues such as the skin, the ovaries and the spleen, whereas other tissues like the bone, the liver, the lungs and the brain are more frequently colonized by cancer cells[193].

In 1889, Steven Paget proposed the "seed and soil" theory based on the observation that different types of cancer have a preference to spread to specific organs. This theory hypothesized that colonization of distant tissues depends on how favorable the host microenvironment of a certain organ (soil) is to a metastatic cancer cell (seed)[194]. It is now known that this hypothesis is not completely accurate, as all distant tissues are harsh for DTCs.

In fact, the most welcoming tissue for a circulating cancer cell is the same organ from which the primary tumor arose[195]. Thus, metastatic organ tropism is driven by a combination of multiple factors, such as the molecular features of cancer cells, the immune cell composition of the host tissue, and interactions between cancer cells and local cells from the host microenvironment. Additionally, organ-specific metastases are controlled by blood and lymphatic circulation patterns, as well as by barriers of the host organ, such as the anatomy of the capillaries, which ranges from the permissive fenestrated capillaries of the liver or the bone marrow to the strictly protected brain capillary walls[196].

Table 4. Metastatic organ tropism for solid tumors. Adapted from the National Cancer Institute 2020, Common sites of metastasis, accessed 6 September 2020, <https://www.cancer.gov/types/metastatic-cancer>.

Cancer Type	Main Sites of Metastasis
Bladder	Bone, liver, lung
Breast	Bone, brain, liver, lung
Colon	Liver, lung, peritoneum
Kidney	Adrenal gland, bone, brain, liver, lung
Lung	Adrenal gland, bone, brain, liver, other lung
Melanoma	Bone, brain, liver, lung, skin, muscle
Ovary	Liver, lung, peritoneum
Pancreas	Liver, lung, peritoneum
Prostate	Adrenal gland, bone, liver, lung
Rectal	Liver, lung, peritoneum
Stomach	Liver, lung, peritoneum
Thyroid	Bone, liver, lung
Uterus	Bone, liver, lung, peritoneum, vagina

2.4 Diversity of breast cancer metastasis

In BCa, metastasis can occur years or even decades after removal of the primary tumor. The duration of this latency period as well as the patterns of metastatic spread, are dependent on BCa subtypes. ER- tumors are associated with early relapse. In contrast, ER+ tumors usually recur after longer periods, which indicates that these tumors need to accumulate colonizing

metastatic features under selective pressure of the distant tissue. Regarding BCa organ tropism, the bone accounts for 70% of BCa metastases, followed by the liver and the brain, which represent around 30% and 20%, respectively[197]. The bone is the preferred metastatic site for all BCa subtypes except for basal-like tumors, which show higher relapse in the lung, brain and distant lymph nodes[198]. Among all the subtypes prone to bone metastasis, the luminal subtypes metastasize to the bone at a much higher rate (80.5%) than HER2-enriched tumors (55.6%)[199]. Unfortunately, adjuvant BCa treatment can have an impact on metastasis patterns in specific BCa subtypes. For instance, adjuvant chemotherapy in ER- patients leads to a decrease in bone metastasis but an increase in visceral metastasis[200].

Most alterations of the metastatic lesion are already present in the primary tumor and can define the affinity of BCa cells for certain organs[201]. For instance, the expression of a specific set of genes, called lung metastasis signature, in the primary breast tumor correlates with an increase in the development of lung metastasis and a decrease in bone and liver metastasis among basal-like tumors. This signature includes genes that collectively promote growth, survival and angiogenesis, including epiregulin, cyclooxygenase 2 (COX2), the transcriptional regulator inhibitor of DNA binding 1 (ID1), MMPs and Retinoic acid receptor responder protein 3 (RARRES3)[202,203]. In contrast, the expression of a different set of genes that include trefoil factor (TFF) 1 and TFF3, cell adhesion proteins and proteins involved in the FGFR-MAPK signaling pathway is specifically associated with bone metastasis[204,205].

Despite significant advances in our understanding of metastatic BCa, this condition is generally considered incurable. Therefore, further research in this field is required to improve therapies. Given that the bone is the foremost BCa metastatic site, affecting a vast number of patients, the research of this thesis focuses on unraveling the molecular mechanisms involved in BCa bone metastasis.

3. BONE METASTASIS

The bone is one of the preferred metastatic sites for cancer cells. Bone metastases occur in 70% of breast and prostate cancer patients and in 15-30% of lung, thyroid or kidney cancer patients[206]. Interestingly, bone metastasis is more common than primary bone tumors, especially in adults[207].

Once bone metastases have been established, there is typically a short-term prognosis. In BCa, 80% of the patients die within the first 5 years after diagnosis of bone metastasis. Although rarely cured, bone metastases can be treated in order to slow down their growth and palliate symptoms[206]. Bone metastases cause severe pain, impaired mobility, pathological fracture, spinal cord or nerve root compression, bone marrow infiltration and hypercalcemia, which can have fatal consequences such as dysfunction of the kidneys, gastrointestinal tract and even the central nervous system[208]. These complications are collectively called skeletal-related events (SREs).

Given the high number of patients suffering from bone metastases as well as the associated costs, there is an urgent need to find new tools to identify those at risk of distant bone metastasis and to develop new therapeutic options to prevent this condition.

3.1. Structure of the bone

The bone has several essential functions in the human body. It plays a crucial role in structural support and movement, as it protects vital organs and provides muscle attachments. However, it also has a metabolic function, serving as a reserve of minerals and energy. Moreover, the bone contains the bone marrow, which is the main site of hematopoiesis[209].

A variety of cell types, vessels and hydroxyapatite crystals are embedded in a matrix of collagen fibers and other non-collagenous proteins, forming the porous bone structure. The human skeleton has two types of bone, namely cortical and trabecular (Figure 7A). These are identical in chemical composition but differ in structure. The dense and compact cortical bone accounts for 80% of skeletal mass, forming the outer part of all skeletal structures. Cortical bone structure confers resistance to bending and tension and makes this type of bone ideal for structural and protective functions. In contrast, trabecular bone accounts for 20% of the

skeleton and is found inside long bones and inner portions of large flat bones. This type of bone is less dense and more elastic and has a major metabolic function[210,211].

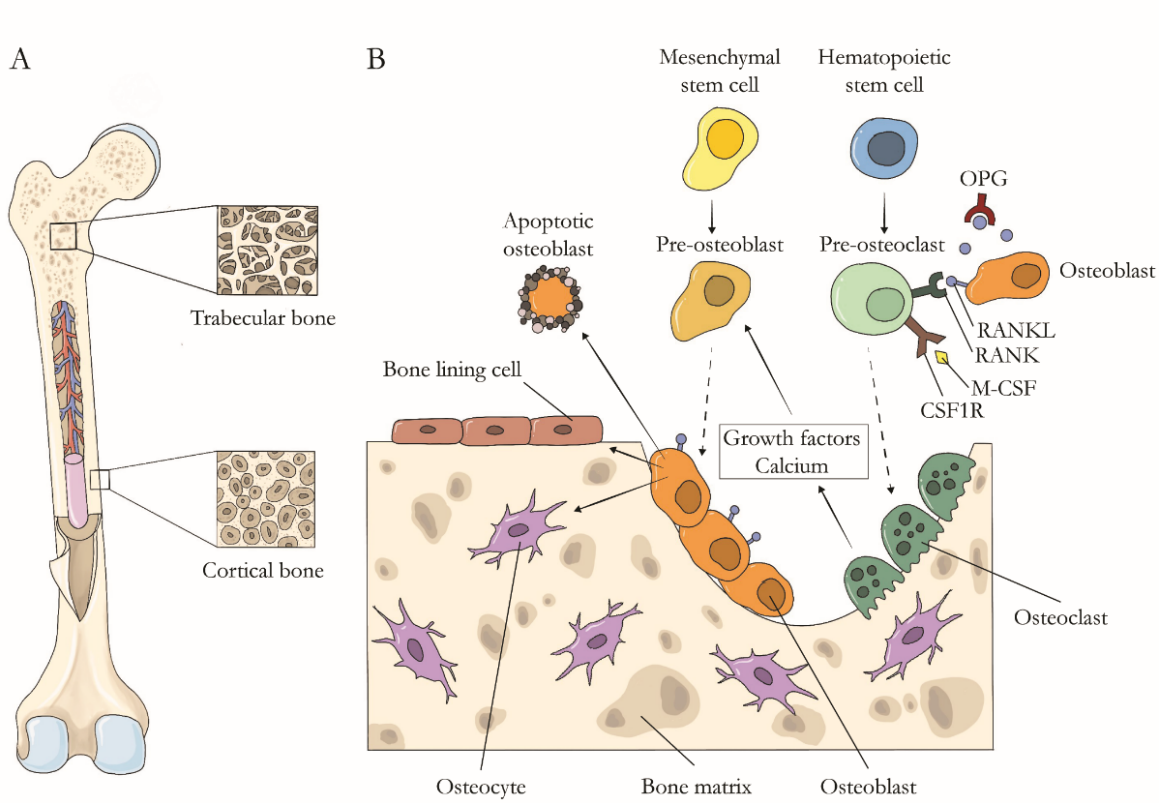


Figure 7. Bone anatomy and remodeling. (A) Schematic overview of bone structure. Two types of bone tissue form the human skeleton: the compact and dense cortical bone and the porous and light trabecular bone. **(B)** Physiological bone remodeling. When osteocytes sense old or damaged bone areas, osteoclasts, derived from hematopoietic stem cells, are recruited to resorb bone. Osteoblasts, osteocytes and stromal cells trigger osteoclast differentiation by signaling through different molecules. Bone resorption results in growth factors and calcium release, thereby contributing to osteoblast differentiation. Osteoblasts, derived from mesenchymal stem cells, are recruited to replace the bone removed by osteoclasts and, after the new bone has been synthesized, osteoblasts either differentiate into osteocytes if buried in the newly formed bone matrix, become bone lining cells, or undergo apoptosis. M-CSF, macrophage colony-stimulating factor; CSF1R, colony-stimulating factor 1 receptor; RANK, receptor activator of nuclear factor-κB; RANKL, RANK ligand; OPG, osteoprotegerin. Adapted from Iaquinta *et al* (2019) and Weilbaecher *et al.* (2011).

The bone is a dynamic organ that is continuously remodeled by the coordinated activity of four specific cell types: osteoblasts, bone lining cells, osteoclasts and osteocytes (Figure 7B). In addition, the bone microenvironment is formed by other cell types, such as adipocytes,

fibroblasts, reticulocytes, chondrocytes, endothelial cells, pericytes and nerve cells, as well as hematopoietic and mesenchymal stem cells[209].

Osteoclasts are large multinucleated cells with a unique ability to resorb bone. These cells derive from the myeloid lineage of hematopoietic precursors of the bone marrow and their formation and activation are regulated by cytokines and hormones. Macrophage colony-stimulating factor (M-CSF) and receptor activator of nuclear factor-κB ligand (RANKL), produced by osteoblasts, osteocytes or stromal cells, are required for osteoclastogenesis. Osteoclast precursors display the colony-stimulating factor 1 receptor (CSF1R) and the receptor activator of nuclear factor-κB (RANK) on their surface, which induce osteoclasts formation and activation upon M-CSF and RANKL binding, respectively. Both M-CSF and RANKL are critical for the differentiation of osteoclast precursors but M-CSF is further required for osteoclast proliferation, survival and the necessary cytoskeletal rearrangements for bone resorption[209,212]. Osteoclasts adhere to the bone matrix and resorb bone by secreting hydrogen ions and proteases such as cathepsin K. Hydrogen ions are released to acidify the extracellular space beneath osteoclasts and dissolve the mineral component of the bone matrix while proteases are secreted to digest the protein component of the matrix[206,212].

Osteoblasts derive from mesenchymal stem cells (MSCs) in the bone marrow. MSCs differentiate to osteoprogenitors that give rise to preosteoblasts and eventually to osteoblasts. All these differentiation processes are regulated by signaling of cytokines and hormones such as parathyroid hormone (PTH), prostaglandin, interleukin-11 (IL-11), IGF-1, TGF-β, fibroblast growth factor (FGF), platelet-derived growth factor (PDGF), Wnt and bone morphogenetic proteins (BMPs), and expression of transcription factors such as runt-related transcription factor 2 (Runx2), osterix (Osx) and activating transcription factors (ATFs)[206,213]. Osteoblasts have large nuclei and enlarged Golgi apparatus and are responsible for the synthesis of new bone matrix to replace the bone removed by osteoclasts and maintain structural integrity and bone homeostasis. Osteoblasts synthesize the osteoid matrix, the unmineralized portion of the bone whose main component is collagen, and produce alkaline phosphatase, osteopontin (OPN) and osteocalcin to mineralize bone. Moreover, they secrete other factors such as RANKL and M-CSF, which have a direct impact on osteoclastogenesis[209,212].

Bone lining cells are quiescent flattened osteoblasts that cover the bone surface where resorption or bone formation is not required. These cells, which are joined by adherens junctions, are located close to osteoblasts, and can reacquire their secretory activity depending on the physiological status of the bone. The functions of bone lining cells are poorly defined but evidence shows that they regulate osteoclast differentiation and prevent the interaction between osteoclasts and the bone matrix when bone resorption is not required[212,213].

Osteocytes are the most abundant cells of the bone and derive from osteoblasts that suspend their activity once trapped in newly secreted bone matrix. These cells are embedded in mineralized bone and function as stress sensors that regulate bone metabolism and homeostasis in response to mechanical and hormonal signals. Osteocytes are connected to each other and to the bone surface through extensive filopodia and are active during bone resorption, inhibiting osteoblast differentiation and recruiting osteoclasts where bone remodeling is required[212,213]. Osteocyte apoptosis is regulated by multiple factors that influence bone remodeling, such as bone damage, hormone levels and mechanical stimuli and is tightly associated with osteoporosis[209]. An example of hormonal control of osteocyte survival is E2, whose deficiency has been associated with osteocyte apoptosis and subsequent osteoporosis[214].

The bone matrix comprises inorganic salts and an organic matrix and provides structural support for all the aforementioned cell types. The organic matrix is formed mainly by collagenous proteins, predominantly type I collagen, but also by noncollagenous proteins such as osteocalcin, osteonectin, OPN, fibronectin, BMPs and growth factors. The inorganic component of the bone matrix consists mostly of hydroxyapatite crystals, formed by calcium and phosphate, which are used for metabolic functions and maintenance of serum homeostasis. Hydorxyapatite crystals deposit on the organic matrix, conferring the characteristic stiffness of bone tissue[212,213].

3.2. Bone remodeling

Bone remodeling refers to the removal of bone matrix and subsequent replenishment and is essential for maintaining the structural integrity of the bone, as well as for mineral homeostasis. This process occurs by the sequential and cooperative action of osteoblasts, bone lining cells, osteocytes and osteoclasts, which group together in temporary anatomical structures called

basic multicellular units (BMUs). Bone remodeling is structured in a cycle of stages: quiescence, activation, resorption, reversal, formation, and return to quiescence[215].

Normally, bone surfaces are in a quiescent state if remodeling is not required. When osteocytes recognize old or damaged bone areas, the bone remodeling process is activated and osteoclast precursors are recruited for differentiation into mature osteoclasts and subsequent bone resorption. Osteoclastogenesis requires mediators such as M-CSF and RANKL, produced by osteoblasts, osteocytes and stromal cells, that bind to their respective receptors CSF1R and RANK at the surface of osteoclast precursors and trigger its proliferation, differentiation and fusion into active multinucleated osteoclasts. Mature osteoclasts resorb bone matrix until suppressive signals are released by osteoblasts, osteocytes or lining cells during the reversal stage. For instance, osteoblasts produce osteoprotegerin (OPG), a decoy receptor that prevents RANKL binding to RANK, thus inhibiting osteoclast differentiation and activation. The ratio of RANKL to OPG is critical for osteoclast activation. Furthermore, osteocytes and lining cells produce prostacyclin (PGI_2), a potent bone resorption inhibitor. Following the cease of bone resorption, differentiation of osteoblast precursors and accumulation of new osteoblasts is mediated by growth factors and chemotactic agents and the resorption cavity is filled with new bone matrix in a process called formation. After synthesis and mineralization of the new bone matrix, mature osteoblasts undergo apoptosis, differentiate into osteocytes or undergo morphologic and functional transformation to lining cells and the bone returns to its original state of quiescence[215–217].

Skeletal integrity is maintained by a balance between bone resorption and bone formation and is controlled by many local and systemic factors. Deregulation of this mechanism results in skeletal diseases such as osteoporosis, caused by excessive bone resorption, or osteopetrosis, resulting from excessive bone formation[211]. The local factors that regulate bone remodeling include cytokines, growth factors and prostaglandins produced in an autocrine or paracrine manner by bone cells as well as factors of the bone matrix released during bone resorption. Alternatively, systemic factors include PTH, PTH-related protein (PTHrP), calcitonin, 1,25-dihydroxyvitamin D3 (calcitriol), IGF-1 glucocorticoids, androgens and E2[213]. The latter has a critical role in bone homeostasis by decreasing the responsiveness of osteoclast progenitors to RANKL and by reducing the levels of osteoclastogenic cytokines such as IL-1, IL-6, IL-11, TNF-α, TNF-β, and M-CSF. Moreover, E2 stimulates osteoblast proliferation and upregulates

the local production of OPG, IGF-1 and TGF-β[210]. The decrease in E2 at menopause is the main cause of bone loss and osteoporosis[218].

3.3. Cancer cell interactions with the bone microenvironment

The discontinuous capillaries and constant remodeling of the bone make this organ highly permissive and supportive for cancer cell colonization and growth. Cancer cells interact with the bone microenvironment and benefit from the plethora of growth factors present in the osteogenic niche. Primary tumors can condition the bone marrow through the production of circulating factors that target cells in the bone microenvironment, forming a pre-metastatic niche. For instance, BCa cells from the primary tumor can secrete the enzyme heparanase in order to cleave heparan sulfate from the bone marrow, resulting in bone resorption and the release of several factors that support bone colonization and metastatic growth[219]. Calcium released during bone resorption is detected through the calcium-sensing receptor (CaSR) and works as a chemoattractant for cancer cells[220,221]. Furthermore, OPN secreted by cancer cells promotes the recruitment of bone marrow cells and MMPs secretion by osteoblasts, thereby supporting bone osteolysis in prostate cancer[222,223]. PTHrP is upregulated in many types of cancer and promotes bone resorption and the production of supporting factors by the bone microenvironment[224].

Regarding bone homing, DTCs can interact with the bone endothelium in order to extravasate and compete with hematopoietic stem cells (HSCs) for occupancy of osteogenic niches (Figure 8). Many of the adhesion molecules and chemokines that cancer cells use for extravasation are involved in physiological mechanisms used by HSC homing to bone. For instance, the expression of RANK by BCa cells mediates attachment to RANKL in osteolytic areas in the bone[225]. CXCR4 at the surface of DTCs from different types of cancer binds to its ligand CXCL12, expressed by osteoblasts, mesenchymal cells and pericytes at the bone, and inhibition of this interaction sensitizes cancer cells to standard chemotherapy[226–228]. Cancer cells also express integrins on their surface that interact with ECM proteins such as bone

sialoprotein (BSP), fibronectin, OPN, VCAM1 or type I collagen in the bone, allowing bone colonization[229–231].

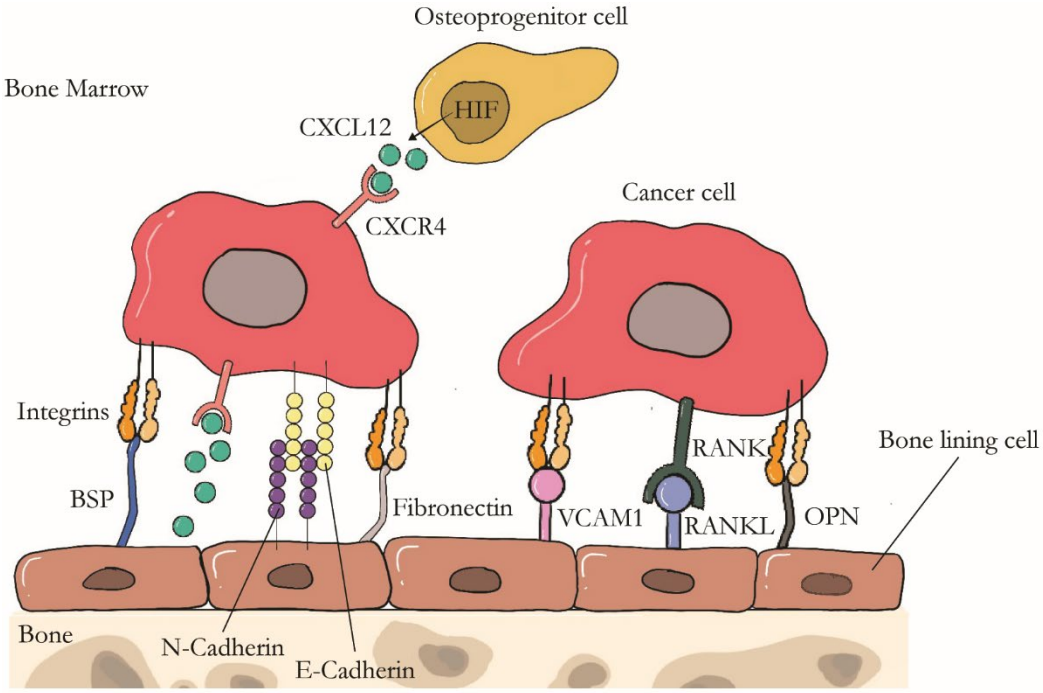


Figure 8. Molecules involved in cancer cell bone homing. Different ligand-receptor pairs are used by cancer cells to enter the bone tissue. CXCL12, C-X-C Motif Chemokine Ligand 12; CXCR4, C-X-C Motif Chemokine Receptor 4, BSP, bone sialoprotein; VCAM1, vascular cell adhesion molecule 1; RANK, receptor activator of nuclear factor-κB; RANKL, RANK ligand; OPN, osteopontin. Adapted from Salvador et al (2019).

DTCs in the bone marrow may remain in a latent state as solitary cells or as a dormant tumor mass for long periods. During this time tumor cells are not fully adapted to the bone microenvironment, as they are unable to grow, but they are resistant to treatments aimed at cell proliferation. Cell-autonomous mechanisms and angiogenic and immunological processes maintain the latent state of DTCs or micrometastases[232] (Figure 9). Stromal signals such as hypoxia, TGF-β and BMP can trigger DTC quiescence[233–235]. Particularly, TGF-β signaling confers a dormant state by increasing p38 MAPK signaling in DTCs[234]. Mitogen-and stress-activated kinase 1 (MSK1), a downstream target of p38 MAPK, regulates chromatin marks that modulate the expression of genes involved in luminal cell differentiation, thus facilitating dormancy by suppressing metastatic features[236]. Cancer cells can express the Wnt inhibitor dickkopf 1 (DKK1), which leads to a self-imposed slow-cycling state, as well as downregulation of NK cell ligands and evasion of immune surveillance[237]. The binding of

prostate cancer cells to annexin II at the surface of osteoblasts triggers the expression of growth arrest-specific 6 (GAS6) receptors, well-known inducers of dormancy in HSCs. GAS6 produced by osteoblasts is sensed by these receptors, resulting in a reduction in cell cycle progression[238]. Moreover, DTCs inhibit PI3K signaling under nutritional stress to induce quiescence and autophagy[239,240]. Tumor mass dormancy is sustained by a balance of proangiogenic and antiangiogenic factors, such as VEGF and TSP1, and a balance of immune cell presence and tumor cell evasion of their action[158].

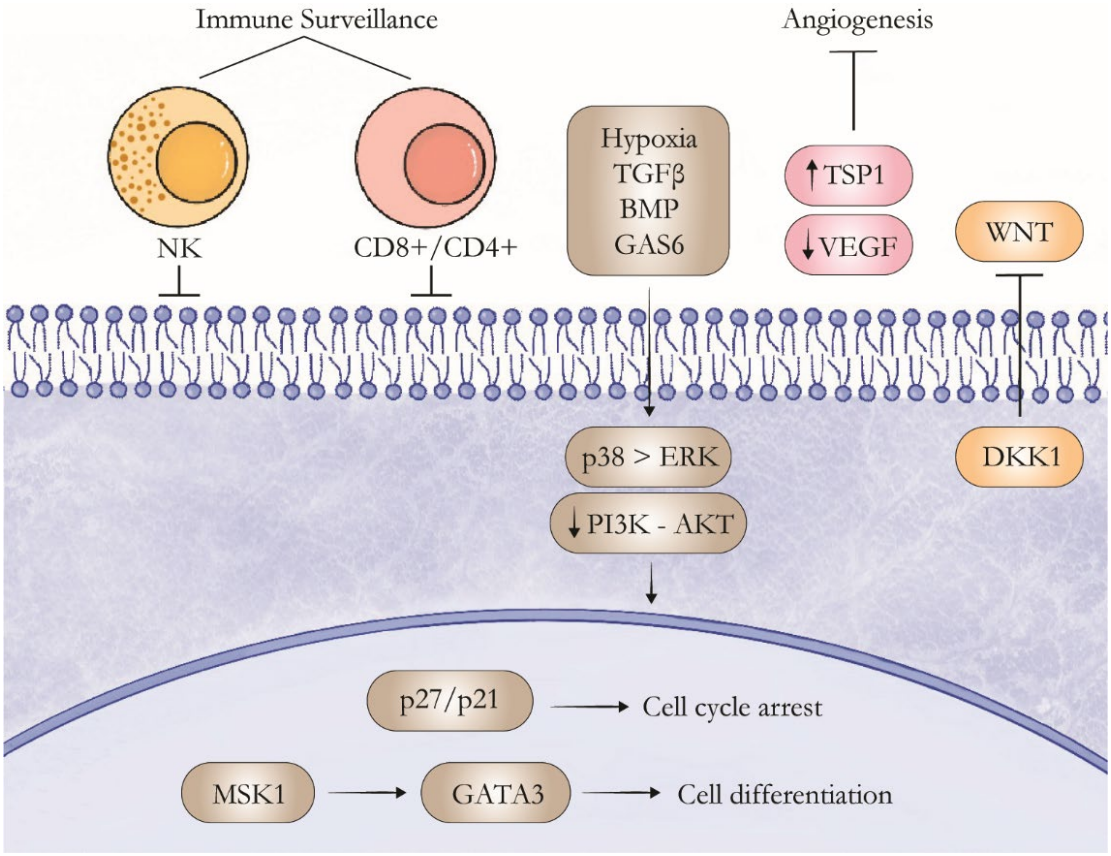


Figure 9. Mechanisms of metastasis dormancy. Bone metastatic cells can remain in a latent state through cell-autonomous mechanisms and interactions with the bone microenvironment. A balance between proliferation and apoptosis is maintained during metastatic dormancy through regulation of angiogenesis and the action of the immune system. In cancer cells, multiple signaling molecules trigger the activation of pathways that result in cell cycle arrest and cell differentiation. NK, natural killer; CD8+/CD4+, cytotoxic T-cell; TGFβ, transforming growth factor-β; BMP, bone morphogenetic protein; GAS6, growth arrest-specific 6; TSP1, thrombospondin 1; VEGF, vascular endothelial growth factor ; ERK, extracellular signal-regulated kinase; PI3K, Phosphoinositide 3-kinase; DKK, dickkopf 1; p27/p21, cyclin-dependent kinase inhibitors; MSK1, mitogen-and stress-activated kinase 1; GATA3, GATA binding protein 3. Adapted from Salvador et al (2019).

Acquiring the macrometastatic growth capacity in the bone depends on mechanisms such as DTC evasion of immune surveillance or attenuation of antiproliferative signals from the bone microenvironment, as well as on the establishment of new interactions between DTCs and the bone microenvironment[232]. For instance, VCAM1 expression by cancer cells promotes the transition from micrometastasis to overt metastasis by recruiting osteoclast progenitors and increasing osteoclast activity[241]. The growth of metastatic bone lesions involves aberrant bone remodeling through paracrine crosstalk between cancer cells and the bone microenvironment. Depending on the mechanism of interference with physiological bone remodeling, macrometastases can be classified as osteolytic, osteoblastic or mixed lesions.

3.3.1. Osteolytic metastasis

Osteolytic metastases are characterized by the destruction of the normal bone and are typical in multiple myeloma (MM), adult T cell leukemia, and breast, lung and renal cancer[216]. Osteolysis is caused by osteoclast stimulation, not by the direct action of cancer cells on the bone tissue.

Cancer cells secrete osteolytic factors such as PTHrP, IL-6, IL-8, IL-11, VEGF and TNF-α, which enhance osteoclast formation by increasing the ratio of RANKL to OPG[216] (Figure 10). Evidence shows that neutralizing antibodies to PTHrP reduce tumor cell growth and are effective in the treatment of hypercalcemia[224]. Moreover, cancer cells can also produce MMPs that increase the amount of active RANKL either by processing it to an active soluble form or by cleaving epidermal growth factor (EGF)-like ligands from tumor cells to activate epidermal growth factor receptors (ERBBs) and suppress the expression of OPG in osteoblasts, resulting in osteoclast activation[223,242]. Denosumab, a monoclonal antibody that targets RANKL are an effective treatment option in patients with bone metastasis[243]. Transcription factors such as GLI family zinc finger 2 (GLI2) and RUNX2 in cancer cells are associated with osteolysis by inducing expression of PTHrP and MMPs, respectively[244,245].

Upon activation through cancer cell-derived molecules, osteoclasts attach to the bone surface and release hydrogen ions and proteases to resorb bone, which facilitates cancer cell invasion through the bone stroma. Bone resorption involves the release of several factors stored in the matrix into the bone microenvironment. These factors include TGF-β, IGFs, BMPs and calcium ions and they support cancer cell growth while inhibiting apoptosis. Moreover, stromal

TGF-β and calcium stimulate PTHrP production by cancer cells, among other osteolytic factors, further promoting bone resorption, cancer cell growth and subsequent release of bone matrix factors[246–248]. Therefore, there is a "vicious cycle" in osteolytic metastasis where cancer cell-driven osteoclast activation and bone resorption results in the release of bone matrix factors, which in turn support cancer cell growth and lead to further bone resorption.

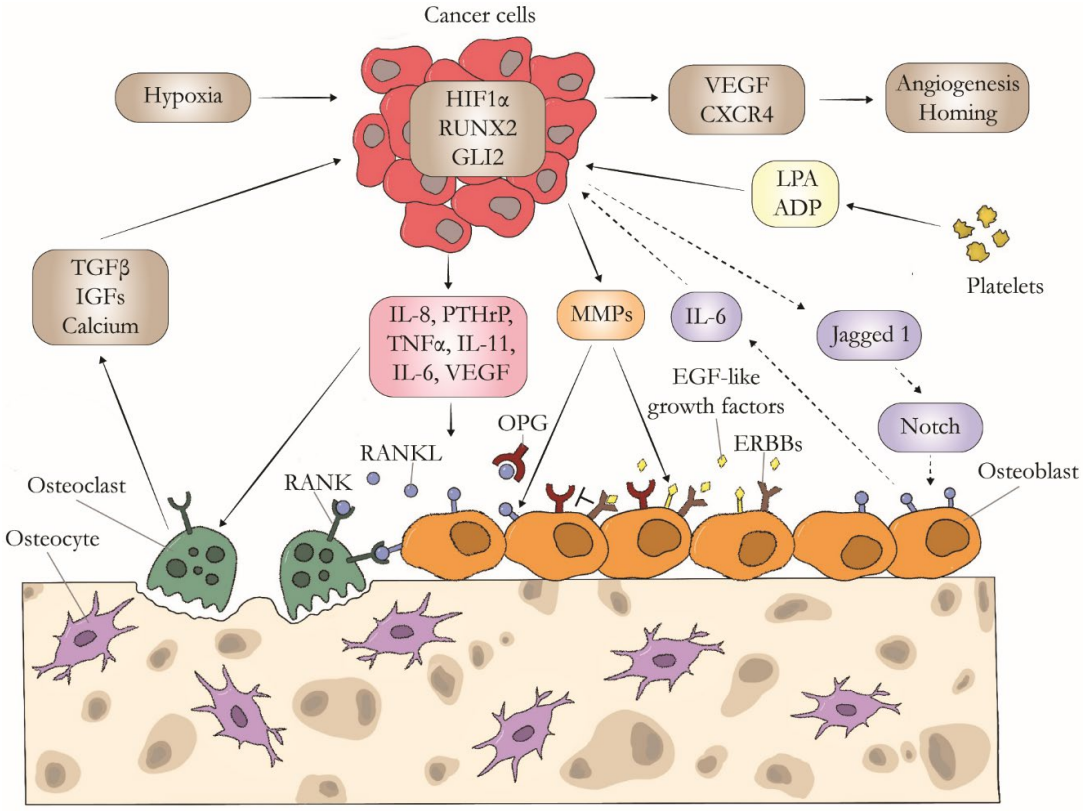


Figure 10. The vicious cycle of osteolytic bone metastasis. Cancer cells secrete numerous factors that facilitate bone homing and induce angiogenesis and osteoclastic bone resorption. Osteoclasts, directly (solid arrows) or indirectly (dashed arrows) activated by the aforementioned molecules, resorb bone and cause the release of factors from the bone matrix. These released factors further promote tumor growth and bone resorption, perpetuating tumor activity. HIF1α, hypoxia-inducible factor 1α; GLI, family zinc finger 2; RUNX2, runt-related transcription factor 2; VEGF, vascular endothelial growth factor, CXCR4, C-X-C Motif Chemokine Receptor 4; LPA, lysophosphatidic acid; ADP, adenosine diphosphate, TGF-β, transforming growth factor-β; IGFs, insulin-like growth factors; IL-6/8/11, interleukin-6/8/11; PTHrP, parathyroid hormone-related protein; TNFα, tumor necrosis factor; MMPs, matrix metalloproteinases; RANK, receptor activator of nuclear factor-κB; RANKL, RANK ligand; OPG, osteoprotegerin; EGF, epidermal growth factor; ERBB, epidermal growth factor receptor. Adapted from Weilbaecher et al (2011).

Additionally, other signaling pathways such as HIF1α and Jagged 1 (JAG1)-Notch support osteolytic bone metastasis[249–251]. HIF1α, together with TGF-β, induce the expression of VEGF and CXCR4, promoting angiogenesis and facilitating bone homing. JAG1 binds to Notch in osteoblasts and activates a signaling pathway that results in IL-6 release and osteoclasts activation[249]. In patients with bone metastasi, chemotherapy can induce JAG1 expression in osteoblasts, which provides a pro-survival niche for cancer cells. Pre-clinical data shows that the use of antibodies against JAG1 in combination with chemotherapy impairs tumor growth[252].

Finally, platelets can further contribute to osteolytic metastasis by inducing cancer cell production of osteolytic factors such as IL-6 and IL-8 through the release of lysophosphatidic acid (LPA) and adenosine diphosphate (ADP)[253].

3.4.2. Osteoblastic metastasis

Osteoblastic metastases are caused by cancer cell production of factors that stimulate osteoblast proliferation, differentiation and bone formation not necessarily preceded by bone resorption. This types of bone metastasis is present in prostate cancer, small cell lung cancer, Hodgkin lymphoma and medulloblastoma[207].

Although the mechanisms that lead to osteoblastic metastases are less understood than for osteolytic metastases, some molecules are associated with osteoblast growth and excessive bone formation. BMP, Wnt and TGFβ produced by cancer cells stimulate MSC recruitment and differentiation into pre-osteoblasts, which will differentiate into osteoblasts owing to the action of several tumor-derived growth factors such as endothelin 1 (ET-1), platelet-derived growth factor (PDGF), BMPs, IGFs, FGFs and urokinase (uPA)[216,254,255] (Figure 11). ET-1 levels are increased in prostate cancer patients with osteoblastic metastasis and preclinical evidence shows that blockade of the endothelin receptor reduces the growth of this type of metastasis[256,257]. Alternatively, insufficient bone resorption due to the action of cancer cells can also lead to excessive bone formation and osteoblastic metastasis. For instance, prostate cancer cells produce prostate-specific antigen (PSA) to cleave the amino-terminal portion of PTHrP, thereby impairing the interaction with the PTH receptor and decreasing bone resorption[258].

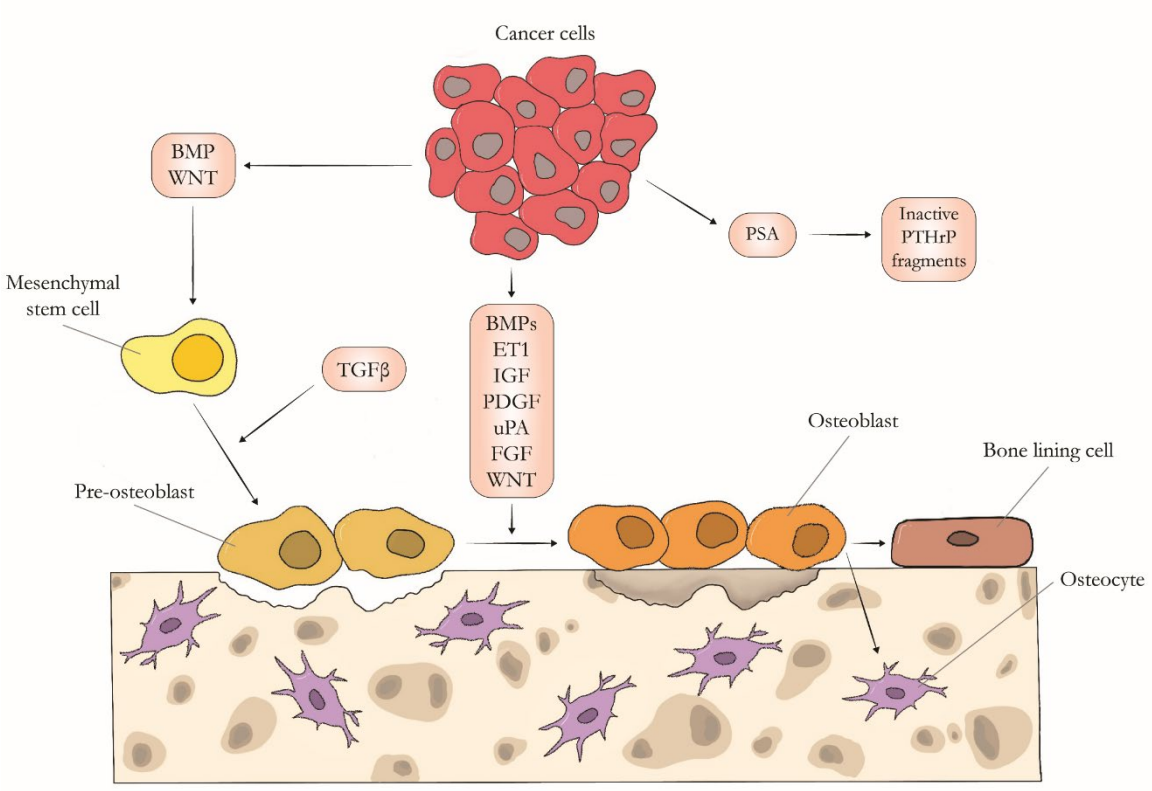


Figure 11. Mechanisms of osteoblastic bone metastasis. Cancer cells can cause excessive bone formation either by promoting osteoblast differentiation and growth through the release of several growth factors or by inactivating osteolytic cues. BMP, bone morphogenetic protein; TGF-β, transforming growth factor-β; ET-1, endothelin 1; IGFs, insulin-like growth factors; PDGF, platelet-derived growth factor (PDGF); uPA, urokinase; FGF, fibroblast growth factor, PSA prostate-specific antigen; PTHrP, parathyroid hormone-related protein. Adapted from Weilbaecher *et al* (2011) and Mundy (2002).

3.4.3. Mixed metastasis

Deregulation of the normal bone remodeling process during cancer progression leads to excessive bone resorption in osteolytic metastasis and excessive bone formation in osteoblastic metastasis. However, this distinction is not absolute and patients can present mixed bone metastasis by having osteolytic and osteoblastic lesions or one lesion with both osteolytic and osteoblastic components. Mixed bone metastases are present in breast, gastrointestinal and squamous cancers, although most bone metastases in BCa patients are osteolytic.

3.5. Bone metastasis treatment

The presence of bone metastasis is associated with SREs, which have an important impact on patient's quality of life and survival. Bone metastasis management is focused on SRE prevention and symptom palliation, and it requires a multidisciplinary approach that includes oncologists, radiotherapists, a pain control team, radiologists, endocrinologists, surgeons and psychologists[207]. The major treatments for bone metastasis are directed to the bone microenvironment and are used in combination with systemic therapies for advanced cancer to reduce tumor burden at skeletal and extra-skeletal sites[259]. Bone-targeting treatments include local approaches and systemic approaches.

3.5.1. Local treatments

Local treatments comprise radiation therapy, ablation techniques and orthopedic surgery, and they are used to prevent fractures and relieve symptoms such as pain.

Radiation therapy is used to relieve pain and improve patient's quality of life, as well as to reduce the need for analgesics. 50% of the patients experience full relief from pain within the first 2 weeks of radiation therapy treatment[259,260].

Ablation techniques are based on the action of cold, heat or chemicals on the bone tissue to reduce tumor burden. The main ablation techniques used for bone metastasis treatment are radiofrequency ablation and cryoablation[207].

Orthopedic surgery is performed to fix pathological fractures or to stabilize high-risk lesions. Surgery depends on life expectancy and metastasis site[207,261].

Osteoplasty consists of the injection of polymethyl methacrylate (PMMA), also called bone cement, into a bone lesion, and it is used to reduce the risk of pathological fracture, relieve pain and improve mobility. Less invasive than surgery, this technique is used in lesions that are not eligible for surgery, lesions refractory to radiotherapy and chemotherapy and after ablation techniques to fill the space left after tumor reduction[262].

3.5.2. Systemic treatments

Systemic approaches include inhibitors of bone resorption, osteoblast modulators and radiopharmaceuticals, and they are aimed at restoring the abnormal bone remodeling that occur in bone metastases[259] (Figure 12).

3.5.2.1. Inhibitors of bone resorption

Inhibitors of bone resorption impair osteoclast formation and activity and include the approved drugs bisphosphonates and denosumab and other molecules that are still not licensed such as cathepsin K inhibitors and c-Src inhibitors. Moreover, evidence shows that some systemic cancer treatments such as mTOR inhibitors and proteasome inhibitors also have an impact on osteoclastogenesis[261].

Bisphosphonates are pyrophosphate analogs that either inhibit essential enzymes for osteoclast activity and survival or are metabolized to cytotoxic compounds that induce osteoclast apoptosis. These drugs are used to modify the bone microenvironment by targeting osteoclasts instead of directly targeting cancer cells. However, evidence suggests that bisphosphonates might have a direct effect on other cell types such as immune cells, osteoblasts, endothelial cells or even cancer cells[263–266]. Bisphosphonates are a major therapeutic option for SRE prevention in metastatic BCa patients and they include zoledronic acid (ZOL), ibandronate, pamidronate and clodronate, all of them being equally effective[267]. Clinical data from the ABCSG-12 and ZO-FAST clinical trials showed that adjuvant ZOL significantly improves PFS and OS in BCa patients, although the beneficial effects of this drug seem to be dependent on a low E2 environment[268,269]. Moreover, the AZURE clinical trial showed that ZOL reduced the development of bone metastasis and, in women who were more than 5 years postmenopausal, improved disease outcomes[270]. These results prove that the endocrine environment affects the anticancer activity of ZOL, although the mechanisms underlying this observation are still under investigation. Close monitoring of patients treated with bisphosphonates is required, as there is a risk of severe adverse effects such as kidney damage or osteonecrosis of the jaw. Other common but mild adverse effects include fatigue, fever, nausea and anemia[261].

Denosumab is a high affinity antibody to RANKL that prevents its interaction with RANK at the osteoclast surface and inhibiting osteoclast maturation and activity[243]. Clinical trials have demonstrated that denosumab is superior to ZOL in delaying the first SREs in breast and prostate cancer patients, although it does not improve PFS or OS[271,272]. The adverse effects associated with this drug include nausea, diarrhea, weakness and osteonecrosis of the jaw. However, it does not affect kidney function[261].

Cathepsin K inhibitors, which include odanacatib, dutacatib and balicatib, act by blocking the function of a lysosomal protease secreted by osteoclasts in order to cleave collagen and degrade the bone matrix. These drugs have been tested for osteoporosis and osteoarthritis treatment and odanacatib in particular has proven to reduce bone resorption markers in BCa patients with bone metastasis[273]. Unfortunately, long-term treatment with odanacatib was associated with atrial fibrillation and stroke, which lead to the discontinuation of its development[274,261].

c-Src inhibitors block the action of the proto-oncogene tyrosine-protein kinase Src, which is involved in osteoclast activation and include bosutinib and dasatinib. Clinical trials have shown promising results in terms of drug safety and tolerance and on bone turnover markers[275–277]. However, a phase II study revealed that PFS is not significantly improved by treatment with a c-Src inhibitor[278].

mTOR inhibitors block the action of mTOR, which promotes osteoclast differentiation and inhibits apoptosis. Clinical trials with everolimus have shown improved PFS and beneficial effects on bone turnover and SREs[279–281].

Proteasome inhibitors inhibit osteoclast differentiation by preventing the degradation of inhibitor of kappa B (I-κB) and therefore blocking NF-κB signaling. Bortezomib and carfilzomib in combination with chemotherapy improve PFS and OS in patients with MM[282,283].

3.5.2.2. Osteoblast modulators

Osteoblast modulators are being explored for bone metastasis treatment, although none have been approved for the prevention of SREs to date. Osteoblast modulators include PTH, anti-

sclerostin antibodies, DKK-1 inhibitors, activin-A inhibitors, ET-1 antagonists and tyrosine kinase inhibitors.

PTH upregulates genes from the Wnt pathway and downregulates DKK-1 and sclerostin, promoting osteoblast differentiation. Pre-clinical data have shown that PTH improves bone mineral density; however, clinical data show that PTH is associated with risk factors in patients with MM[284,285].

Anti-sclerostin antibodies include romosozumab, blosozumab and BPS804 and they act by blocking the action of sclerostin, a Wnt inhibitor, to trigger osteoblast differentiation. Clinical trials show an increase in bone mineral density in osteoporosis patients treated with anti-sclerostin antibodies[286,287]. However, romosozumab is associated with cardiotoxicity[261].

DKK-1 inhibitors promote osteoblast differentiation by blocking the action of the Wnt inhibitor DKK-1 produced by cancer cells. Phase IB and phase II clinical trials with DKK-1 inhibitors report an increase in bone mineral density in patients with MM[288,289].

Activin-A inhibitors block the action of the activin-A cytokine, secreted by osteoblasts, osteoclasts and stromal cells and involved in osteoclast activation and inhibition of osteoblast differentiation. A clinical trial with sotatercept, an activin-A analog that blocks the activin-A receptor, showed an improvement in bone mineral density and bone pain in MM patients[290].

ET-1 antagonists block the effects of ET-1, involved in osteoblast differentiation. Although preclinical data showed that blockade of the ET-1 receptor reduces the growth of osteoblastic metastasis, clinical trials with the ET-1 antagonists zibotentan and atrasentan did not show improvement of OS or PFS[256,257]. However, atrasentan did have a beneficial effect on bone pain and SREs in prostate cancer patients[291].

Tyrosine kinase inhibitors impair osteoblast activation by blocking the VEGF receptor. Cabozantinib was shown to improve SREs in prostate cancer patients and OS and PFS in patients with renal cell carcinoma[292,293].

3.5.2.3. Radiopharmaceuticals

Radioisotope-labeled molecules can be selectively delivered to bone lesions to induce DNA damage and apoptosis. Radium-233 dichloride is an alpha particle-emitting radiopharmaceutical that works as a calcium mimetic, forming complexes with hydroxyapatite

in the bone and delivering cytotoxic radiation to bone metastasis sites. A clinical trial has demonstrated that radium-233 significantly improves OS and SREs in advanced prostate cancer patients[294].

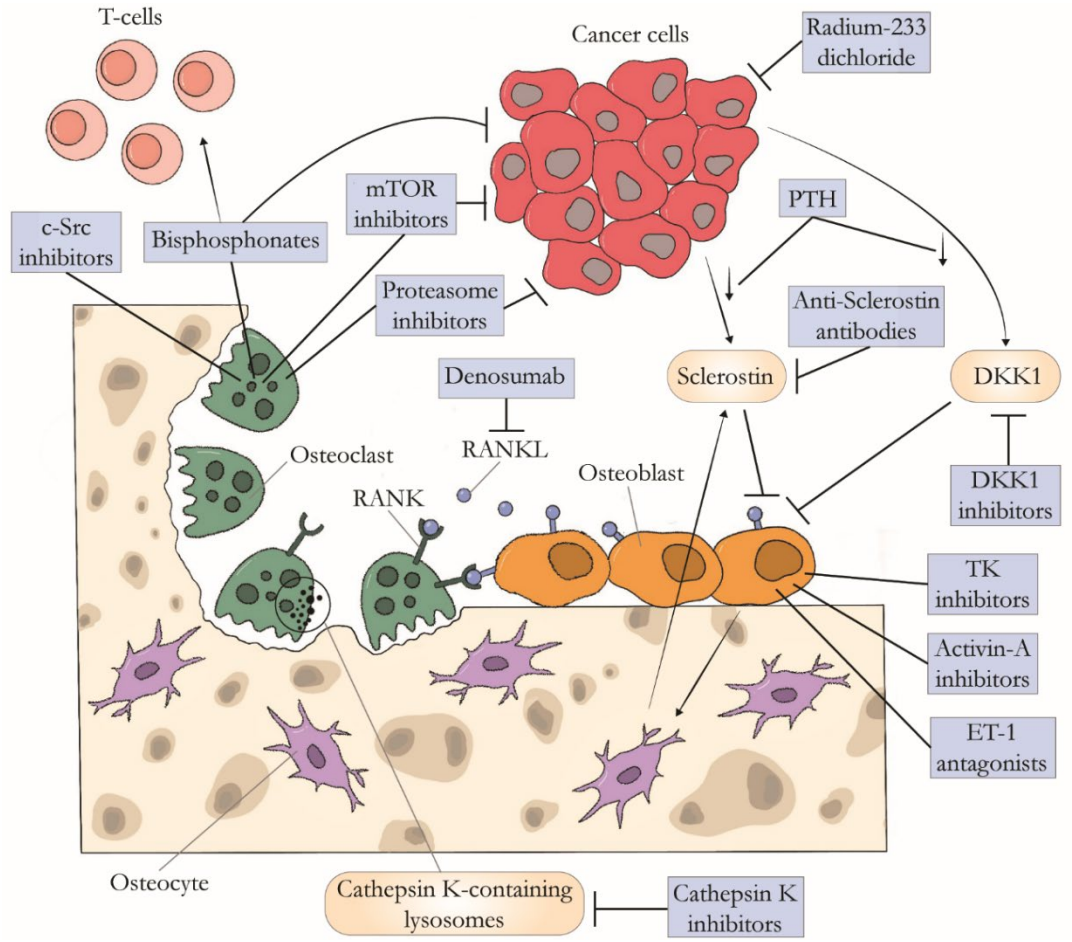


Figure 12. Mechanism of action of common and potentially novel therapeutic options for bone metastasis. Proto-oncogene tyrosine-protein kinase Src (c-Src) inhibitors, bisphosphonates, mammalian target of rapamycin (mTOR) inhibitors, proteasome inhibitors, denosumab and cathepsin K inhibitors compromise osteoclast activation, differentiation and action. Additionally, bisphosphonates, mTOR inhibitors, proteasome inhibitors as well as radium-233 dichloride also have a direct anti-tumor effect. In contrast, parathyroid hormone (PTH), anti-sclerostin antibodies, dickkopf 1 (DKK1) inhibitors, tyrosine kinase (TK) inhibitors, activin-A inhibitors and endothelin 1 (ET-1 antagonists) act by modulating osteoblast activity. Adapted from D'Oronzo *et al* (2019).

3.6. Predictive biomarkers for relapse and response to bone metastasis treatment in breast cancer

Despite recent therapeutic advances, bone metastasis is still a major complication with an important impact on patient's quality of life. Up to 50% of patients with bone metastasis treated with anti-resorptive drugs develop new bone metastases, skeletal complications and disease progression. These observations highlight the need to develop new therapies, as well as to identify novel biomarkers to predict relapse and response to existing bone-modifying agents[216].

Bone turnover biomarkers are currently used to identify bone metastatic patients at high risk of SREs or death. These biomarkers include type I collagen fragments such as C-terminal telopeptide of type I collagen (CTX), cross-linked C-terminal telopeptide of type I collagen (ICTP) and N-terminal telopeptide of type I collagen (NTX), products of collagen synthesis such as procollagen type 1 N-terminal propeptide (P1NP), and bone alkaline phosphatase (BALP)[232]. High NTX or BALP levels are associated with an increased risk of skeletal complications and disease progression in patients undergoing bone antiresorptive therapy[295,296]. Moreover, bone turnover biomarkers can provide prognostic information about future bone recurrence. For instance, high P1NP, CTX and 1-CTP serum levels in early BCa patients are associated with an increased risk of bone metastasis but not metastasis to other distant sites, thereby indicating the importance of bone turnover to provide a proper niche for BCa cells[297]. Other biomarkers such as dedicator of cytokinesis protein 4 (DOCK4), involved in cell adhesion processes, peroxiredoxin 4 (PRDX4), an osteoclastogenesis inducer and p21-activated kinase 4, a repressor of ER transcriptional activity, has also been suggested to identify BCa patients at high risk of bone recurrence[298–300].

Preventing metastasis is far better than treating it. The use of adjuvant bone-modifying treatment was tested in clinical trials and showed that not all patients benefit from treatment[301]. Thus, biomarkers that can predict which patients are most likely to benefit from this treatment represent an unmet medical need. Macrophage-capping protein (CAPG) and PDZ domain-containing protein GIPC1 (GIPC1) in primary breast tumors are associated with bone metastasis and poor OS but can also predict benefit from ZOL treatment. BCa patients with high CAPG and GIPC1 expression levels show up to 10-fold reduction of bone metastases and 2.5-fold reduction for death upon ZOL treatment compared to patients with low levels of

these biomarkers[302]. Additionally, it has been proved that the v-maf avian musculoaponeurotic fibrosarcoma oncogene homolog (MAF) transcription factor drives specifically bone metastasis in early-stage ER+ BCa patients by regulating the expression of a set of genes that support bone colonization[303]. Moreover, MAF status predicts likelihood of benefit from adjuvant ZOL treatment in BCa patients[304]. Considering these observations, this transcription factor has emerged as a promising biomarker for bone metastasis prevention and treatment decision-making for BCa patients.

4. MAF transcription factor, a novel biomarker for bone metastasis

The discovery of MAF as a metastasis driver, particularly to the bone, makes this transcription factor a powerful target for the prevention and treatment of BCa metastasis to bone and highlights the need of further exploring MAF function in BCa cells.

4.1. The MAF family of transcription factors

MAF transcription factors, together with Fos, Jun, CREB and activating transcription factor (ATF) families, are part of the activator protein 1 (AP-1) superfamily of basic leucine zipper (bZIP) proteins (Figure 13A). The MAF family consists of small and large MAF proteins. The former comprise MAFF, MAFG and MAFK, while large MAF proteins include NRL, MAFA, MAFB, and MAF (also known as c-maf), the latter alternatively spliced in two isoforms (short and long) that differ only in their carboxy termini[305] (Figure 13B).

Transcription factors from the AP-1 family have a basic domain, which allows binding to 12-*O*-tetradecanoyl phorbol 13-acetate-responsive element (TRE) or cAMP-responsive element (CRE) DNA sequences and a leucine zipper domain, which mediates homo or heterodimerization with other bZIP transcription factors required for DNA binding (Figure 13C). The MAF family has an additional domain called extended homology region (EHR) that allows the recognition of longer sequences called MAF-recognition elements (MAREs). MAREs are composed of a TRE or CRE core, recognized by the basic domain, and a flanking TGC sequence recognized by the EHR domain. TGC sequences are indispensable for MAF binding, but the TRE or CRE core can be more degenerated[305,306].

Both small and large MAF proteins have a basic domain, a leucine zipper domain, and EHRs. However, they differ in an N-terminal transactivation domain that contains binding sites for transcription coregulators, only present in large MAF proteins[305]. Small MAF proteins homodimerize to bind to MARE sites whereas large MAF proteins homodimerize or heterodimerize with other bZIP transcription factors *in vitro*, although no relevant partners *in*

vivo have yet been identified[307–309] . Notwithstanding, large MAF proteins do not heterodimerize with small MAF proteins[305].

MAF proteins can undergo post-translational modifications such as phosphorylation, sumoylation and acetylation. For instance, glycogen synthase kinase 3 (GSK3) phosphorylation of MAFA has a dual function, inducing its ubiquitylation and degradation but also increasing its transcriptional activity through the recruitment of the P/CAF coactivator[310]. Furthermore, MAFB sumoylation in myeloid cells results in inhibition of its transactivation capacity and impaired macrophage differentiation[311].

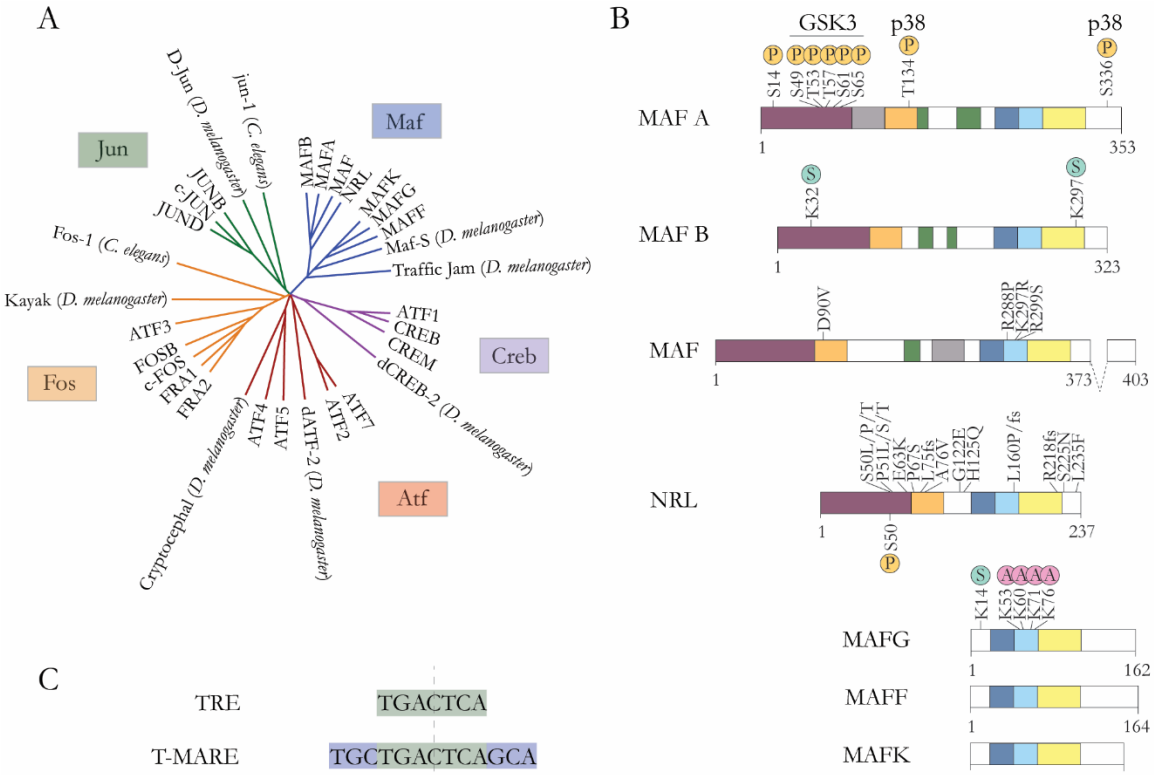


Figure 13. The MAF family of transcription factors, members of the AP-1 superfamily. (A) Phylogenetic tree of the activator protein 1 (AP-1) superfamily of proteins. **(B)** Schematic representation of MAF family member structures. MAF is alternatively spliced in two isoforms named short and long. Residues involved in post-translational modifications as well as the kinases responsible for some of these modifications are illustrated. Mutations related to cataract and retinitis are depicted in MAF and NRL proteins, respectively. **(C)** Recognition sequences of AP-1 superfamily members. MAF transcription factors can bind to a T-MARE, C-MARE, degenerated MARE or half MARE flanked by AT-rich sequences. TRE, 12-Otetradecanoyl phorbol 13-acetate (TPA)-responsive element; CRE, cAMP-responsive element; MARE, MAF recognition element. Adapted from Eychène *et al* (2008).

Large MAF proteins recruit coactivators such as p300, CRE binding protein (CREBBP), P/CAF or TATA binding protein (TBP) to activate transcription. In contrast, small MAF homodimers compete with large MAF proteins for binding to target gene promoters, thus repressing transcription[312,313]. Hence, transcriptional regulation through MAREs depends on a balance of large MAF-containing complexes and small MAF homodimers[314].

Large MAF proteins can regulate their own expression, as their promoters contain MARE sequences. This is important during development, where there is a cascade of expression of various large MAF proteins in which the expression of one member might be implicated in the expression of the next member in the cascade. Large MAF proteins are important in tissue specification and act as key regulators of terminal differentiation in many tissues, including bone, brain, kidney, lens, pancreas, retina and blood[306,315]. For instance, MAFA participates in insulin transcription and production in pancreatic β cells, MAFB regulates macrophage and podocyte differentiation and pancreatic endocrine cell commitment and maturation and MAF is involved in T-cell, lens and chondrocyte differentiation[305].

Mutations in the large MAF proteins NRL and MAF have been associated with diseases such as pulverulent cataract, clumped pigmentary degeneration and retinitis pigmentosa[306,315]. Furthermore, deregulation of MAF protein expression can lead to cancer development and osteolytic metastasis[303,305].

4.2. MAF function in cancer cells

Small MAF members are not considered oncogenes but have been reported to participate in antioxidant responses that might contribute to cancer development. In contrast, large MAF proteins have been directly implicated in carcinogenesis[305,316].

Large MAF members are considered oncogenes due to their ability to transform primary cells. The first MAF family member discovered was *v-maf*, identified in the AS42 avian retrovirus, which induces musculo-aponeurotic-fibrosarcoma in chickens[317,318]. Additionally, transgenic mice with MAF overexpression in the T-cell compartment develop T-cell lymphoma[319], and in humans, 50% of MM and 60% of angioimmunoblastic T-cell lymphomas (AITLs) express high levels of MAF, which directly contribute to cancer progression[319,320,321]. Similarly, translocations involving MAFA and MAFB are also found in MM[322,323].

MAF, MAFA and MAFB can transform fibroblasts *in vitro* and this capacity relies on high expression levels rather than on activating mutations[309,324,325]. MAF and MAFA are more effective at transforming cells, whereas MAFB displays less oncogenic activity[324,326]. In contrast, NRL has no transforming activity and, accordingly, its expression has not been found to be deregulated in human cancer[305]. In summary, high expression levels of unmutated MAF have oncogenic activity in cell culture, animal models and human cancer.

As transcription factors, MAF proteins regulate a series of genes that might be responsible for malignant transformation. Various studies based on gene expression profiling of MM patients have reported a MAF gene expression signature in this disease[327,328]. Some of these MAF target genes include C-X-C chemokine receptor type 1 (CXCR1), CXCL12, cyclin D2, integrin β7 and AMPK-related protein kinase 5 (ARK5) and they have been found to be up-regulated in MM and AITL with MAF overexpression as well as in MAF transgenic mice[319–321].Thus, MAF governs important processes in MM cells by regulating the expression of genes involved in cell proliferation, such as cyclin D2, genes involved in cell migration and invasion processes, such as CXCL12 and ARK5, and genes involved in cell-cell interactions, such as integrin β7 and CXCR1[321,326,329]. However, the specific MAF transcription program and its role in other cancer types have yet to be defined.

4.3. MAF in breast cancer metastasis to bone

Since bone metastasis is a major complication in BCa patients, the discovery of biomarkers that specifically predict risk of bone metastasis exhibits a great surge of interest in the field. Thus, an experimental xenograft mouse model and three rounds of intracardiac injection of MCF7, human ER+ BCa cells, and subsequent isolation after bone metastasis establishment was used in our laboratory for the generation of a derived cell line with enriched bone metastatic capacity, named BoM2[303]. Comparative genomic hybridization of parental MCF7 and BoM2 cells allowed the detection of copy number aberrations (CNAs) and the identification of a substantial 16q23 chromosomal gain in BoM2 compared with the parental population. Consistently, association of 16q13 chromosomal gain with bone metastasis and poor outcome was clinically validated in independent BCa datasets with annotated clinical follow-up. After examination of the 16q23 region, the MAF transcription factor was identified

as a potential bone metastasis mediator. The association of high MAF expression levels and incidence of bone metastasis but not metastasis, to extra-skeletal sites such as the brain and lungs was validated in historical datasets from ER+ BCa patients (Figure 14A).

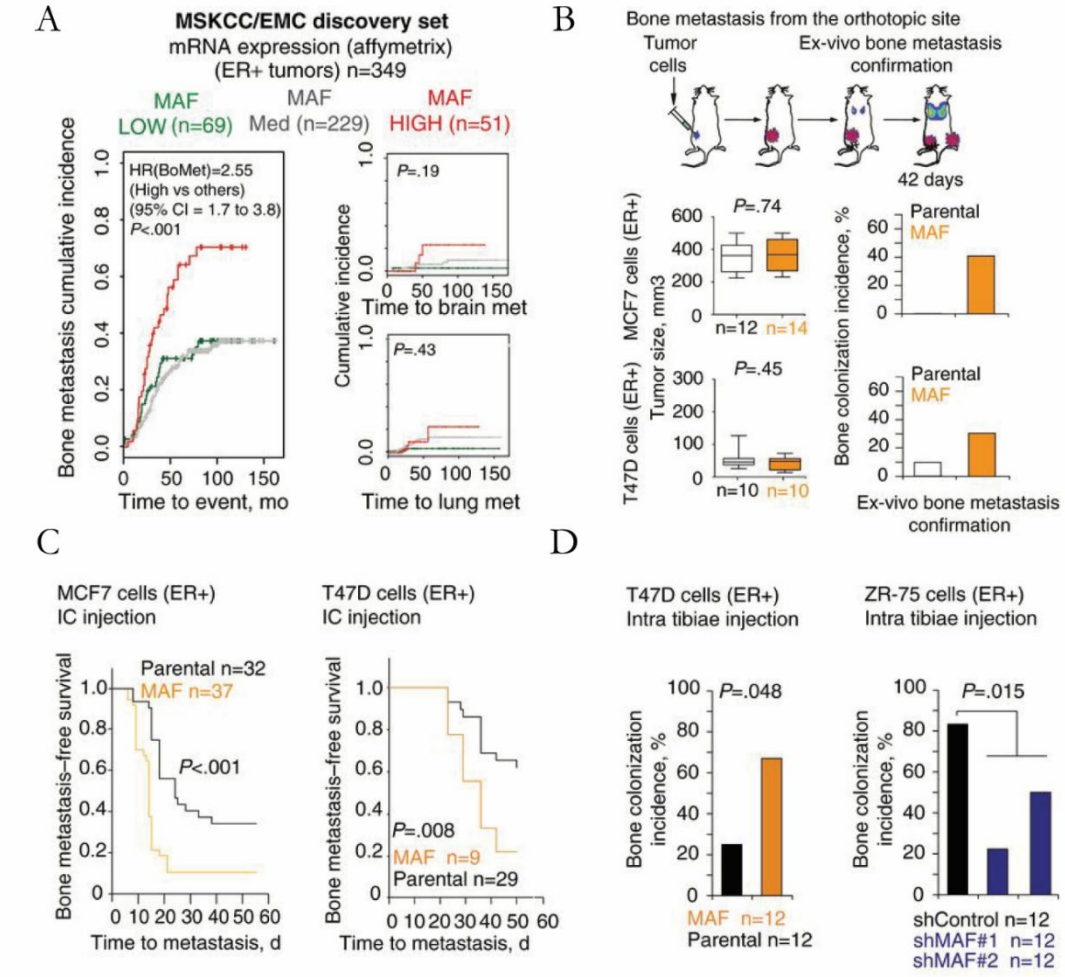


Figure 14. MAF promotes breast cancer bone metastasis. (A) Bone, brain and lung metastasis cumulative incidence plot of estrogen receptor-positive (ER+) primary breast cancer (BC) patients. High MAF expression in the primary tumor is associated with high incidence of bone metastasis but not brain or lung metastasis. **(B)** Bone colonization incidence of mice with MAF-overexpressing and control parental MCF7 and T47D BC cells implanted in the mammary fatpad. Mice bearing primary tumors with similar size were selected for bone metastasis evaluation. **(C)** Kaplan-Meier curves of bone metastasis-free survival for MAF-overexpressing and parental MCF7 and T47D BC cells inoculated through intracardiac injection. **(D)** Bone colonization incidence of MAF-overexpressing and parental T47D and parental and MAF short hairpin-carrying ZR-75 BC cells inoculated through intratibial injection. Adapted from Pavlovic *et al* (2015).

Furthermore, *in vivo* experiments based on orthotopical injection into the mammary gland as well as intracardiac or intratibial injections of different BCa cell lines, showed that MAF-overexpressing cells had an enhanced capacity to metastasize to the bone, while this capacity was abrogated upon MAF depletion[303] (Figure 14B, C and D).

Gene expression analyses in MAF-overexpressing cells suggested that MAF controls the expression of a set of genes that collectively support functions required for bone metastasis. *PTHLH,* the PTHrP-coding gene, was identified among these MAF-regulated genes and its expression correlated with increased osteoclast activation, which is essential for BCa bone metastasis[303]. Of note, PTHrP expression in T4 primary breast tumors in combination with positive nodal status is associated with an increased risk of bone metastasis; however, it fails to predict bone metastasis in early-stage BCa[330,331]. Together, these results suggest that PTHrP is not sufficient for bone metastasis development, which requires a certain degree of transformation in BCa cells. MAF-overexpression in tumor cells might provide the acquisition of necessary functions for bone metastasis development and, in this context, PTHrP expression would become an advantage for cancer cell colonization of the bone. Taken together, these results shed light on the potential use of MAF as a therapeutic target for the prevention or treatment of bone metastasis.

4.3.1. MAF as a predictive biomarker for adjuvant zoledronic acid treatment in breast cancer metastasis to bone

Meta-analyses of individual patient data from randomized clinical trials have demonstrated that adjuvant bisphosphonates, especially ZOL, reduce the incidence of bone metastasis and improve survival only in postmenopausal patients with early BCa[332]. However, no clear biological reason explains why treatment benefit is restricted to women with low E2 expression levels.

Although some biomarkers have been defined as predictors for response to bone-modifying agents, they are still not used in clinical practice, and patients are selected for adjuvant bisphosphonate treatment based only on menopausal status. To fine-tune the selection of patients, a retrospective analysis of the phase III AZURE clinical trial was performed to investigate whether MAF amplification in primary breast tumors could predict the outcomes

of treatment with adjuvant bisphosphonates. This study revealed that, indeed, MAF status predicts response to ZOL treatment (Figure 15A).

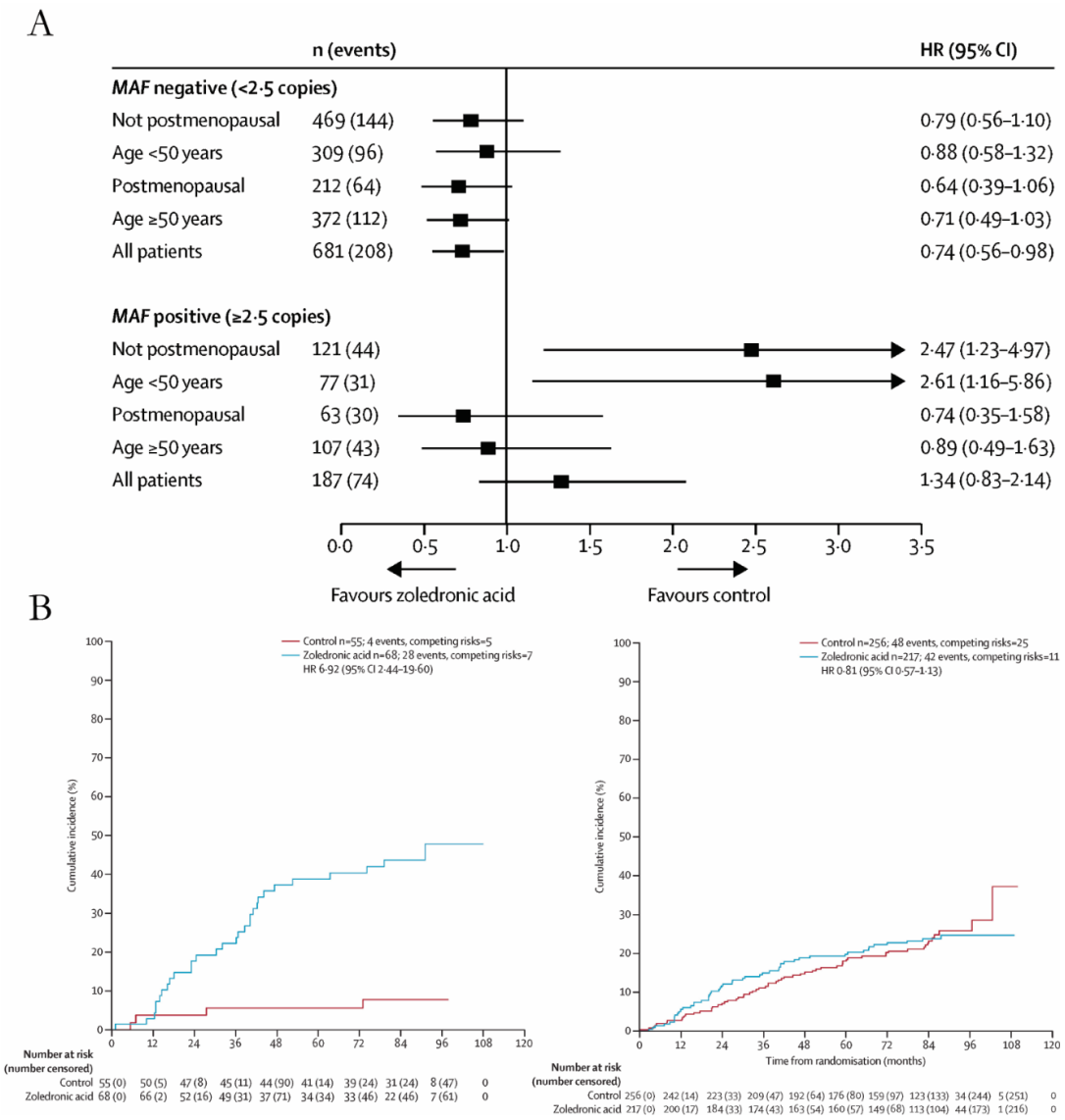


Figure 15. Effect of MAF expression on treatment outcomes with adjuvant zoledronic acid in early breast cancer patients. (A) Effect of High MAF status on association between adjuvant zoledronic acid (ZOL) treatment and metastasis-free survival, stratified by age and menopausal status. **(B)** Cumulative risk of extraskeletal metastasis in patients with MAF-positive tumors (left) and MAF-negative tumors (right) upon control or adjuvant ZOL treatment. HR, hazard ratio. Adapted from Coleman *et al* (2017).

Patients with MAF-negative tumors treated with adjuvant ZOL showed improved disease outcomes and, importantly, this beneficial effect was independent of their menopausal status.

This observation suggests that adjuvant bisphosphonate treatment could be extended to premenopausal women with MAF-negative tumors. In contrast, MAF-positive tumors treated with ZOL were associated with increased adverse outcomes, extraskeletal recurrence and mortality in premenopausal women, demonstrating that this subgroup of patients should be excluded from adjuvant bisphosphonate treatment (Figure 15B). This can be explained by the fact that ZOL treatment does not affect tumor cell viability[333]. Specifically, the primary effect of ZOL acts on the bone stroma, thus impairing bone resorption. Thus, MAF-positive tumors become more aggressive in a high E2 environment and, although ZOL treatment stabilizes the bone, aggressive cells are still able to metastasize to extraskeletal sites.

Collectively, these data highlight the importance of using MAF as a diagnostic tool for ZOL treatment decisions in early BCa patients and emphasizes the need for additional mechanistic studies to understand the effect of bisphosphonate treatment on disease outcomes depending on MAF and hormonal status[334].

4.3.2. Potential therapeutic strategies to target MAF-driven bone metastasis in breast cancer

Collectively, the ability of MAF to transcriptionally regulate genes that support bone metastatic functions and the poorer prognosis associated with MAF expression in premenopausal patients upon ZOL treatment makes MAF an attractive potential molecular target for the prevention and treatment of BCa metastasis to bone. However, the nuclear localization of MAF, as well as its intrinsically disordered structure and lack of a catalytic domain, make it a remarkably challenging therapeutic target. For this reason, further knowledge of the molecular mechanisms underlying this transcription factor is crucial to identify potentially targetable proteins required for bone metastasis.

In MM, MAF has been reported to increase the interaction between tumor cells and the bone marrow stroma by regulating genes such as integrin β7. Integrin β7 can efficiently bind to E-cadherin expressed by bone stromal cells and this interaction induces the expression of the pro-angiogenic cytokine VEGF, which favors tumor cell proliferation and survival[321]. Moreover, other integrins such as integrin β8 are deregulated in MAF-overexpressing cells and mouse models knocked out for different large MAF proteins show defects in cell-cell interactions in pancreatic islets, as well as between somatic and germline cells during gonad

morphogenesis[310,335,336]. Together, these data suggest that disruption of integrin signaling is one MAF function in oncogenesis. Thus, targeting this specific MAF function might help to overcome the challenge that targeting MAF itself poses. Integrins under MAF transcriptional control, among other membrane proteins, are interesting potential candidates for bone metastasis treatment, as they are involved in tumor cell-bone stroma interactions and are located at the plasma membrane, which makes them easily accessible pharmacological targets.

Additionally, MAF-transforming activity has proven to be context-dependent. Evidence shows that the transforming capacity of some large MAF proteins is dependent on culture conditions and cell type[326]. These findings demonstrate the existence of MAF-binding partners that might positively or negatively modulate MAF-transforming activity. Moreover, MAF physiological targets are mostly inducers of terminal differentiation during development and do not explain the oncogenic properties of MAF. None of the physiological targets of MAF have been found to be deregulated in MAF-overexpressing cancer cells. Therefore, MAF oncogenic activity is a result of a switch of target genes that allow cancer cells to acquire novel functions. This switch can be explained by modifications in MAF dimerization and MAF interacting partners. Therefore, deciphering the MAF-interacting network is crucial to understand MAF oncogenic activity and to identify potentially targetable proteins that are necessary for MAF-driven metastasis.

MAF has proven to be a relevant factor for the development of bone metastasis and response to bisphosphonate treatment in BCa. Nonetheless, further knowledge is required in order to make a breakthrough in the treatment of bone metastasis, a fatal complication in BCa patients.

Objectives

OBJECTIVES

Although MAF has emerged as a potential molecular target for the treatment and prevention of BCa metastasis to bone, it is a challenging pharmacological target. Thus, to throw some light on new therapeutic opportunities to prevent bone metastasis, there is a need to thoroughly dissect the molecular mechanisms of MAF transcription in cancer cells, as well as the role of MAF target genes. To address the unsolved questions regarding the complex MAF-driven malignant transformation in BCa cells, the aims of the present thesis are:

1. To identify potentially targetable membrane proteins whose expression is regulated by MAF in metastatic BCa cells

2. To identify MAF-interacting partners and characterize their relevance and contribution to BCa bone colonization

3. To explore the molecular mechanisms of MAF-dependent transcription

Results

Chapter I: Optimization of a One-bead-one-compound combinatorial peptide library screening for the identification of potentially targetable MAF-regulated membrane proteins

After the discovery of the association of MAF with high risk of bone relapse in patients with early-stage BCa, this transcription factor arose as a potential molecular target to prevent bone metastasis. Nevertheless, as a nuclear protein with no enzymatic activity and an intrisically disordered structure, MAF became a considerably difficult therapeutic target and brought to the forefront the potential of MAF downstream targets to become therapeutically actionable. Among MAF downstream targets, we consider membrane proteins of particular interest due to their involvement in metastatic colonization through direct interactions between cancer cells and the host tissue and, importantly, because they can be therapeutically targeted with ease. Therefore, we believe that the identification of membrane proteins under MAF transcriptional control may pave the way for blocking MAF-driven bone metastasis.

One-bead-one-compound (OBOC) combinatorial peptide library screening methods have been previously used to identify novel ligands for protein inhibition as well as unique ligands against cancer cells[337–341]. This technology is based on resin beads that bear unique synthetic peptide ligands on their surface displayed multiple times and offers the possibility to simultaneously screen many different compounds against membrane proteins by incubation with living cells[342–344]. The main advantage of this technique is the possibility of using non-natural components such as D-amino acids, cyclic, turned or branched ligands, which are resistant to proteolytic degradation and facilitate *in vivo* applications[339]. Additionally, the possibility to perform the screening in living cells makes OBOC technology ideal for the identification of membrane protein ligands, as the presence of the plasma membrane and binding partners may be required for the correct folding and for the display of relevant epitopes.

Formerly, hits from OBOC screenings were isolated through manual techniques in a complicated and inefficient process. Thus, in order to increase the throughput, the Complex Object Parametric Analyzer and Sorter (COPAS) large particle biosorter has been used to isolate hits after incubation with fluorescently labeled living cells[342]. For an efficient deconvolution of peptide hits, COPAS sorting can be coupled to a matrix-assisted laser

desorption ionization time-of-flight (MALDI-TOF) mass spectrometry (MS) approach. This method is based on the cleavage of the peptide from the beads followed by transfer to the MALDI target, peptide desorption ionization and sequence determination based on the fragmentation pattern[342]. Together, these techniques can be used to identify peptides that efficiently bind membrane proteins of a specific cell population.

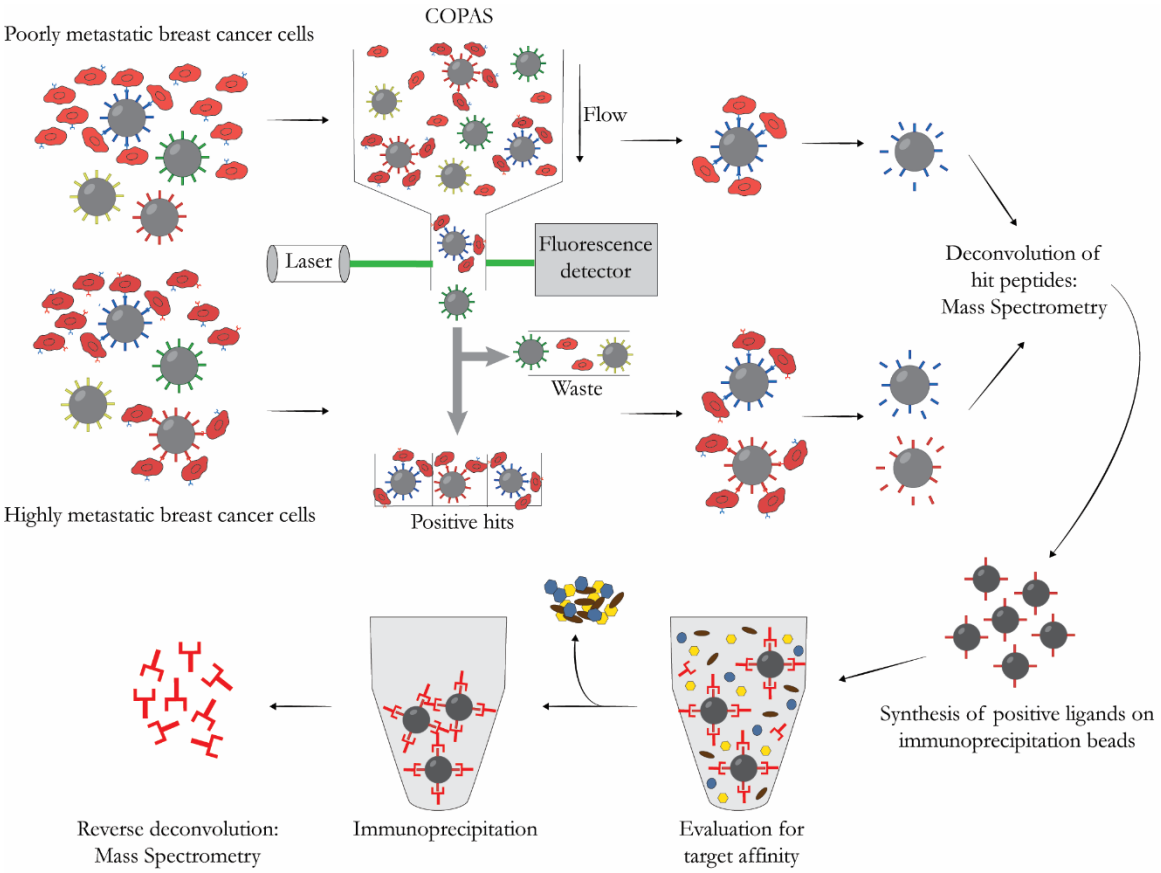


Figure 156. Schematic representation of a cell-based screening of One-bead-one-compound (OBOC) combinatorial peptide libraries for the identification of i) high affinity ligands to bone metastatic cell populations, ii) the identity of the membrane proteins recognized by peptide hits. Beads can be incubated with fluorescently labeled highly and poorly bone metastatic cell populations and hits with strong interaction with cells can be sorted through a Complex Object Parametric Analyzer and Sorter (COPAS). After cleavage of the peptide hits from the beads, sequence can be determined through matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) mass spectrometry (MS). Hits can be resynthesized and conjugated to immunoprecipitation beads in order to precipitate membrane proteins out of bone metastatic cell extracts. The identity of the precipitated membrane proteins can also be revealed by MS. Adapted from Cho *et al.* (2013).

We hypothesize that the MAF transcription program includes membrane proteins involved in cancer cell-bone stroma interactions. Therefore, through a high-throughput screening of OBOC peptide libraries coupled to a MALDI-TOF MS approach we seek to identify new membrane proteins differentially expressed in highly versus poorly bone metastatic BCa cells, defined by their MAF expression levels. The screening will reveal specific synthetic peptide binders capable of recognizing membrane proteins overrepresented in human cancer cells with high MAF expression levels. Afterwards, through a reverse deconvolution based on immunoprecipitation we seek to unravel the identity of the membrane proteins recognized by peptide binder hits. To this end, peptide hits will be resynthesized and conjugated to immunoprecipitation beads and used to precipitate membrane proteins from MAF-expressing cell extracts. Finally, the identity of the membrane proteins will be revealed by mass-spectrometry (Figure 16).

Hence, this approach will allow both the discovery of MAF-regulated membrane proteins as well as the identification of specific peptide binders capable of their recognition. Notably, peptide binders offer the possibility to be conjugated to chemotherapeutic agents and open up new opportunities for systemic therapies for bone metastasis with site-specific drug delivery and reduction of off-target effects.

Proof-of-concept experiments for OBOC combinatorial peptide library screening optimization

Considering the potential of screening OBOC peptide libraries in living cells for the identification of membrane proteins involved in bone metastasis, we first set up the system in our laboratory. In order to select the optimal conditions for cell-bead binding and confirm the capacity of the cell sorting instruments available at IRB Barcelona to isolate positive hits we focused on integrins and integrin-binding peptides, as they represent well-defined molecules frequently used in ligand binding assays.

α6β1 integrin expression in the ER+ BCa cell line MCF7 was confirmed by flow cytometry and different integrin-binding peptides were synthesized on 90μm-sized TentaGel beads with a 4-hydroxymethylbenzoic acid (HMBA) linker by 9-fluorenylmethoxycarbonyl (Fmoc) solid-phase peptide synthesis (SPPS)[345] (Figure 17A and B). TentaGel beads consist of a polystyrene

matrix on which polyethylene glycol is grafted and the HMBA linker serves as a support for the immobilization of carboxylic acids. The ester bond that results from the coupling of an amino acid to the HMBA linker is stable to the strong acids used during peptide synthesis but can be cleaved by nucleophiles like amines for sequence analysis after synthesis.

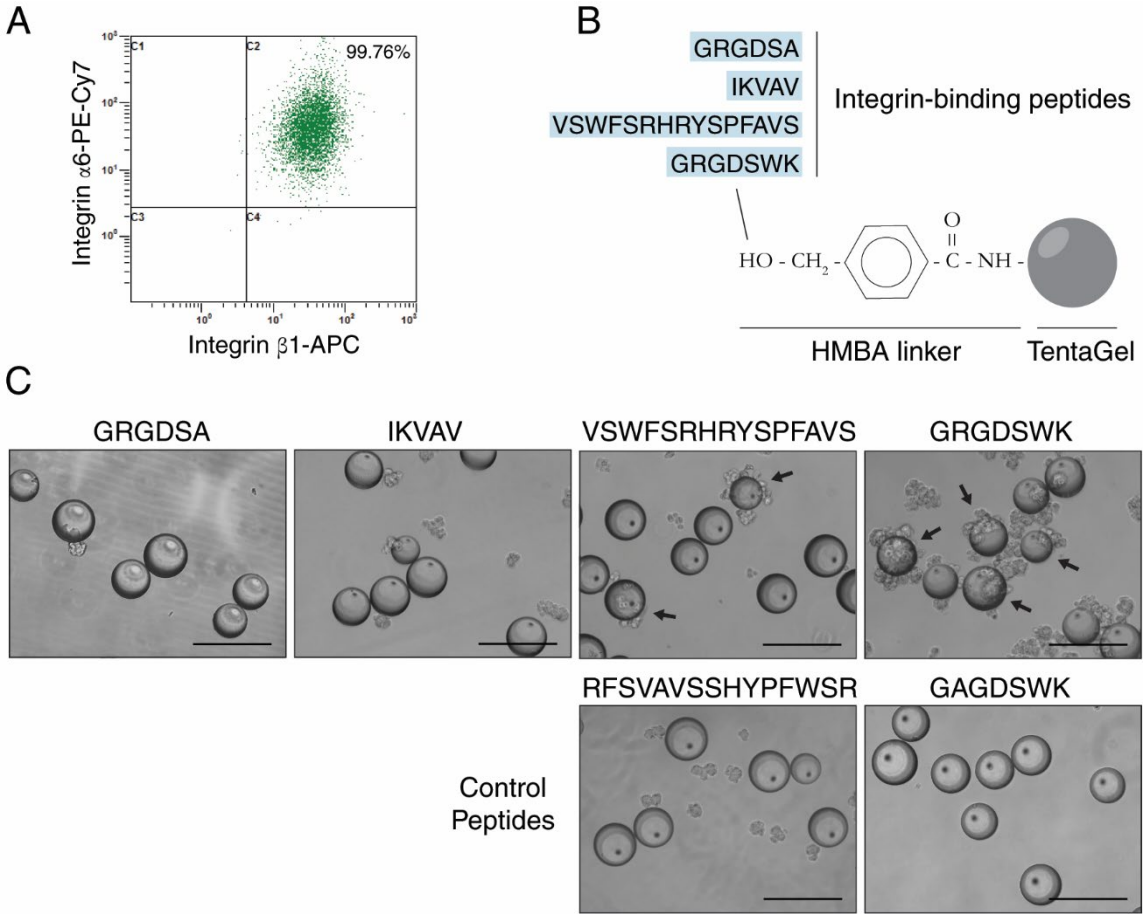


Figure 17. Integrin expression and integrin-binding peptides for One-bead-one-compound (OBOC) peptide library screening optimization. (A) Flow cytometric analysis for α6β1 integrin expression in MCF7 breast cancer cells. **(B)** Design of different integrin-binding peptides synthesized on TentaGel beads via a 4-hydroxymethylbenzoic acid (HMBA) linker **(C)** Bright-field microscopy images depicting association between MCF7 cells and TentaGel beads bearing different integrin-binding peptides after 1h incubation. Arrows show cell-bead binding. Scale bar, 250μm.

Since the α6β1 integrin is a well-characterized laminin receptor that specifically recognizes the widely occurring Arg-Gly-Asp (RGD) cell adhesion motif as well as the IKVAV binding domain on the α-laminin chain[346–348], GRGDSA (RGD-based) and IKVAV peptides were selected as integrin-binding peptides. Red fluorescent protein (RFP)-expressing MCF7 cells

were generated, properly dissociated to avoid aggregation and incubated with the beads bearing the newly synthesized integrin α6β1 peptide ligands. Unfortunately, only minimal association of beads and MCF7-RFP cells was achieved.

Considering that the α6β1 integrin binding domains of laminin might not be reproduced in an active conformation by short peptides, we continued with the synthesis of longer peptides to improve cell-bead association. We synthesized a VSWFSRHRYSPFAVS peptide, with high affinity for the α6β1 integrin, and its scrambled version RFSVAVSSHYPFWSR[349]. Additionally, we synthesized a longer RGD-based peptide with the sequence GRGDSWK, based on the addition of two amino acids that significantly increase binding to integrins, and a GAGDSWK negative control peptide with no RGD motive[342]. Beads with the newly-synthesized peptides were incubated with MCF7-RFP cells and subsequent microscopy revealed a significant increase in cell-bead association compared to the previous GRGDSA and IKVAV peptides, especially when beads were coated with the GRGDSWK peptide (Figure 17C). Of note, cell-bead binding was abrogated when beads were coated with control peptides, confirming the specificity of these integrin-binding peptides.

Upon achieving cell-bead binding, we attempted to sort the beads with cells attached by means of a COPAS, using control beads with no peptides on their surface to define the sort gate. Unfortunately, the COPAS was able to sort only a considerably small percentage of beads and subsequent microscopy revealed that most of the sorted beads were false-positives. We attempted to stabilize cell-bead interactions through chemical cross-linking prior to sorting in order to avoid cell dissociation from the beads while passing through the instrument, but it did not result in an improvement of COPAS efficiency (data not shown).

We next interrogated the capacity of a specialized instrument for fluorescence-activated cell sorting (FACS) to replace the COPAS. This instrument is designed for accurate single cell sorting and therefore it is not capable of sorting large particles such as the 90μm-sized beads regularly used for OBOC combinatorial libraries. However, the possibility of adapting a 130μm-sized nozzle to the FACS instrument offered the option of replacing COPAS purpose. Because 90μm-sized beads covered with a layer of cells could cause nozzle clogging problems, we scaled the synthesis of our strongest binding peptide GRGDSWK and the control peptide GAGDSWK to 30μm-sized beads and subjected the newly synthesized peptide-coated beads to incubation with fluorescent cells. Fluorescence microscopy confirmed binding of the cells

to GRGDSWK-coated beads and no association to GAGDSWK-coated beads. Further, we observed that the new 30µm-sized beads possessed blue autofluorescence, which would facilitate the process of sorting by the FACS instrument (Figure 18A).

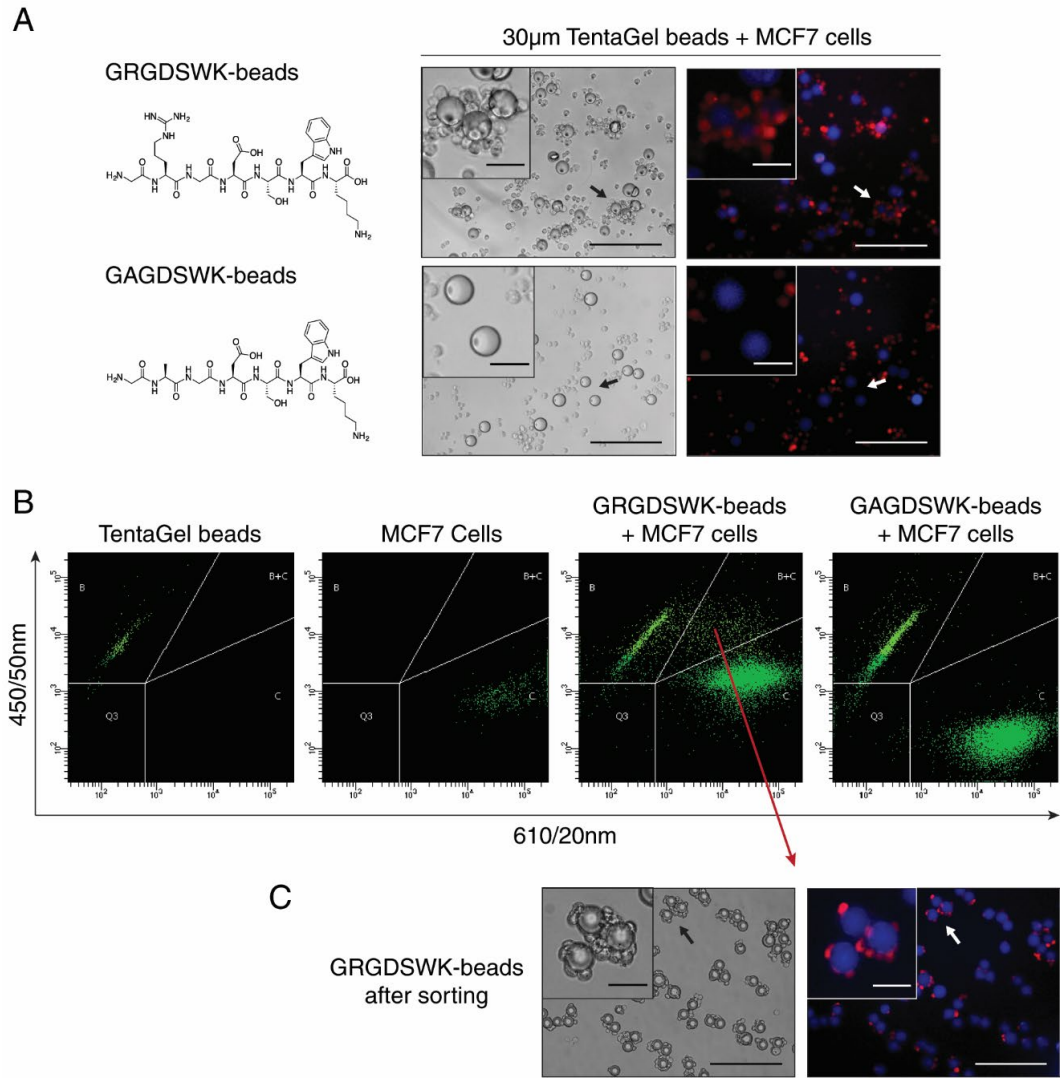


Figure 18. GRGDSWK-coated beads specifically bind to MCF7 cells and can be efficiently isolated by fluorescence-activated cell sorting (FACS). (A) Schematic representation of GRGDSWK and GAGDSWK (negative control) peptides followed by bright-field and fluorescence microscopy images illustrating association between MCF7 cells (red) and peptide-coated TentaGel beads (blue). **(B)** Flow cytometric analysis showing beads and cells distribution to define sorting gates followed by distribution of cell-bead mixtures. Horizontal axis represents red fluorescence intensity and vertical axis depicts blue fluorescence intensity. **(C)** Bright-field and fluorescence microscopy images showing GRGDSWK-beads coated with cells after sorting. Arrows indicate the location of magnified images. Scale bar, 250µm; inset scale bar, 50µm.

Washing steps after incubation and chemical cross-linking were removed in order to avoid partial cell dissociation as well as fixation of unbound cells onto the beads. Hence, the cell-bead mixture was loaded into the cell sorter and, by plotting red and blue fluorescence, the bead and cell populations were separated. Gating and thresholds were defined using a population of beads with no peptides on the surface and a population of red fluorescent cells.

After passing the GRGDSWK-coated beads and cells mixture through the cell sorter, significant association was confirmed and positive hits were successfully collected (Figure 18B). Subsequent microscopic examination revealed that all positive hits were coated with fluorescent cells (Figure 18C). In contrast, when cells were incubated with the control GAGDSWK-coated beads the percentage of positive hits was dramatically decreased, indicating the specificity of the GRGDSWK peptide and demonstrating the efficient functioning of our newly defined system.

Thus, we described a new methodology to isolate high-affinity ligands from OBOC libraries based on 30μm-sized resin beads sorted by a FACS instrument that offered a significant improvement in terms of accuracy and efficiency compared to COPAS sorting.

Design, synthesis and screen of OBOC peptide libraries

After proving that we could effectively sort cell-bound beads in our laboratory, our objective was to identify peptides that specifically distinguish between highly and poorly bone metastatic BCa cell populations. To this end, it was required to generate different OBOC combinatorial peptide libraries.

Random peptide libraries formed by large number of compounds are used for experimental screenings when the targets are unknown in order to identify possible binders. However, many challenging steps such as the synthesis scale to ensure the synthesis of all peptide permutations, the amount of all members, the sequence deconvolution and peptide structure elucidation after screening arise when using large peptide libraries[350]. The sequence deconvolution of peptide hits with similar physico-chemical properties or with the same amino acid composition but differential positioning in the sequence, which leads to identical masses, is not always possible[351]. To overcome this problem, multi-objective genetic algorithms can be used to

simplify the design of random peptide libraries in order to increase diversity of compounds while minimizing redundancy.

In computer science, a genetic algorithm is a method based on the idea of natural selection, where the best characteristics of individuals are selected for reproduction and passed to the next generation. In a genetic algorithm, a set of solutions for a problem is considered and the best options are selected following three biologically inspired rules: selection, crossover and mutation. The algorithm uses two fitness functions that give a score to each candidate solution, selects those candidates with the highest score to become *parents* and does a crossover of two *parents* to form the next generation. Every time that a generation is formed, random changes can be applied to individual *parents* to maintain diversity in a process called mutation[352]. This process is repeated until the algorithm finds the optimal solution.

In collaboration with Dr. Ernest Giralt's group, a genetic algorithm supported approach was used to design two different peptide libraries with optimal design to maximize the number of amino acid permutations and the number of peptides with unique mass while reducing library size to simplify the chromatographic and MS analysis of complex peptide mixtures[353] (Figure 19A). Libraries were made of hexa- and heptapeptides composed of representative D-amino acids from different categories that are considered relevant for protein-protein interactions[354]. Since most protein interactions occur between small domains, hexa- and heptapeptides are large enough to interact with receptors on the cell surface, offer sufficiently high diversity and are considerably easy to synthesize. Each library had different amino acid compositions and properties.

The first library (Library 1) consisted of 32 hexapeptides with unique masses that followed the sequence x_1-x_2-x_3-x_4-x_5-y. Peptide diversity was achieved through amino acid side chain diversity, having a representative amino acid from each category: *G* as an amino acid with no side chain, *w* as a hydrophobic-aromatic amino acid, *s* as a hydrophilic uncharged amino acid, *e* as a hydrophilic-negatively charged amino acid, *a* as a hydrophobic-aliphatic amino acid and *r* as a hydrophilic-positively charged amino acid. Thus, *G, w, s, e, a* and *r* were proposed for each x_i position and the genetic algorithm revealed that an optimal library with unique masses could be obtained by simplifying the input from *G, w, s, e, a, r* in each position x_i to {*G, w*} for x_1, {*s, e*} for x_2, {*G, r*} for x_3, {*a, r*} for x_4 and {*s, a*} for x_5.

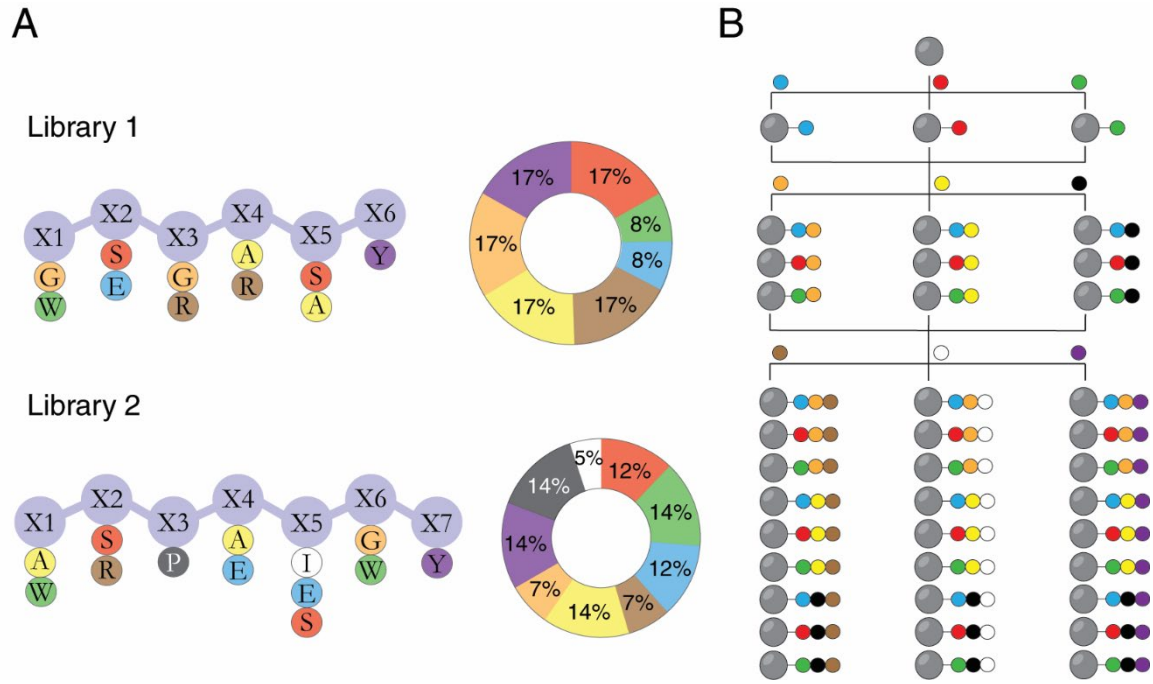


Figure 19. Design and synthesis of One-bead-one-compound (OBOC) combinatorial peptide libraries. (A) Schematic representation of two different OBOC peptide libraries followed by a diagram of the abundance of each amino acid **(B)** Schematic overview of the "split-and-mix" synthesis method used in OBOC peptide library synthesis. Beads are divided into equal portions and each portion is coupled to one amino acid. After the first reaction, all portions are pooled together, mixed and split again into different portions. Each new portion is coupled to a different amino acid, generating the second residue on the beads. These steps are repeated until the desired peptide combinations are obtained.

For the second library (Library 2), more complexity was introduced by using heptapeptides and adding *i* as a hydrophobic-aliphatic amino acid with a longer side chain and *p* as an amino acid without a primary amino group. This library followed the sequence x_1-x_2-p-x_4-x_5-x_6-y and the amino acids *G, w, s, e, a, r, i* were proposed for each variable position. The genetic algorithm produced an output were the variable positions were occupied by {*a, w*} for x_1, {*s, r*} for x_2, {*a, e*} for x_4, {*i, e, s*} for x_5 and {*G, w*} for x_6 to obtain a 48 heptapeptide library with 47 permutations unique by mass and one with mass overlapping.

Libraries were constructed by "split-and-mix" synthesis (Figure 19B) and the successful synthesis of all library components was confirmed by UPLC-MS, high resolution MS and nano

LC-MS/MS. The complete lists of library members are provided in supplementary information (Tables S1 and S2) and the full list of their MS spectra can be found in Kalafatovic *et al*, 2020.

After characterization of the libraries, beads were incubated with BoM2-RFP cells, the MCF7 derivative cell line with high MAF expression levels, to assess their capacity to bind to highly bone metastatic BCa cells. Following incubation, cell-bead mixtures were loaded into a FACS instrument in order to sort positive hits. Flow cytometric analysis revealed association between cells and beads from Library 1, however, few hits were detected when cells were incubated with beads from Library 2 (Figure 20).

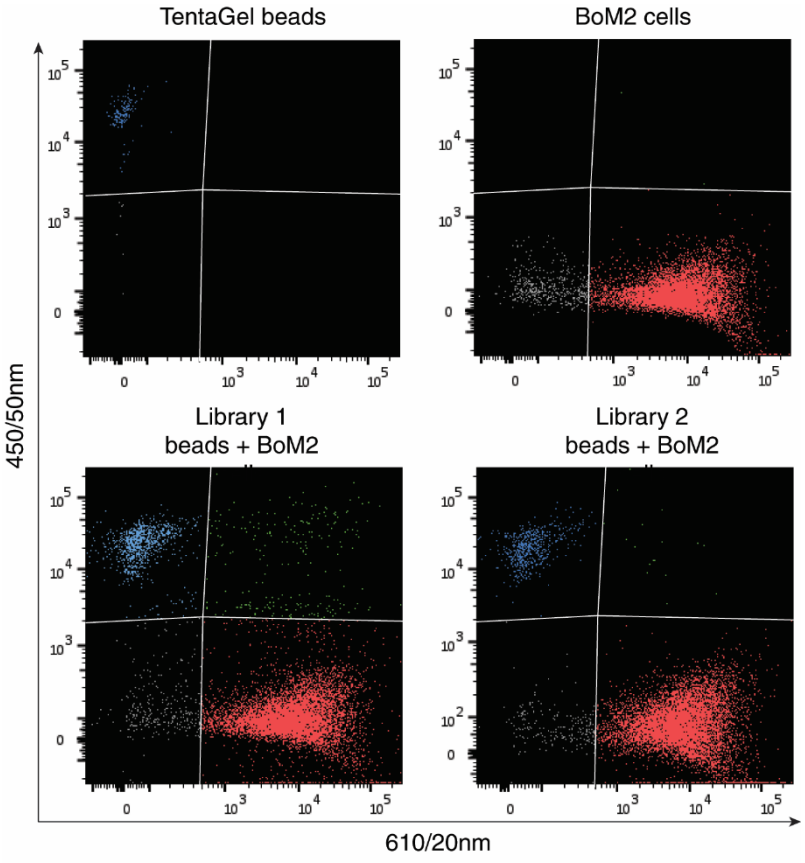
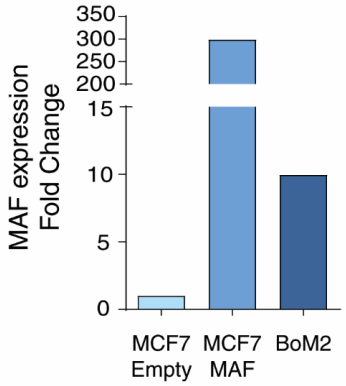


Figure 160. Beads from One-bead-one-compound (OBOC) peptide library 1 can associate with bone metastatic breast cancer cells. Flow cytometric analysis showing TentaGel beads and BoM2 cells distribution to define sorting gates followed by distribution of cell-bead mixtures. Horizontal axis represents red fluorescence intensity and vertical axis shows blue fluorescence intensity.

Since peptides from Library 2 did not seem to strongly bind to BoM2 cells we focused on the analysis of Library 1 peptides. We incubated Library 1 beads with highly and poorly BCa bone metastatic cells and screened for affinity interactions in order to select those peptides capable of specifically recognizing the highly bone metastatic population. MCF7 was selected as the poorly bone metastatic cell population and MAF-overexpressing MCF7 as well as BoM2 were selected as highly bone metastatic cell populations. We performed three different experiments

consisting of two rounds of incubation with cells and sorting in order to collect the necessary amount of positive hits for subsequent MS analysis. Interestingly, when Library 1 beads were incubated with BoM2 and MAF-overexpressing MCF7 cells, which show high MAF expression levels and therefore have more bone metastatic capacity than parental MCF7 cells, the number of sorted events was higher, indicating that bone metastatic BCa cells are more prone to associate to Library 1 beads (Figures 21A and B).

A



B

	Cell Line	Number of sorted events
Experiment 1	MCF7	1920
	BoM2	5185
Experiment 2	MCF7	3731
	BoM2	11835
Experiment 3	MCF7-Empty	1665
	MCF7-MAF	2799

Figure 21. The number of positive hits after incubation of beads from One-bead-one-compound (OBOC) peptide library 1 with breast cancer cell lines is higher when cells show high MAF expression levels. (A) MAF mRNA expression levels in MCF7 cells, MAF-overexpressing MCF7 cells and the bone metastatic derivative BoM2. **(B)** Number of sorted events after incubation of breast cancer cell lines with Library 1 beads.

Identification of peptide hits that target bone metastatic breast cancer cell populations

After screening and collection of positive hits, cells were eliminated from the beads surface by means of a highly concentrated NaCl solution. Then, since peptides where conjugated to the resin through an HMBA linker sensitive to nucleophilic reagents, peptides were cleaved with ammonia vapors prior to determining their sequence through UPLC-MS.

The optimized UPLC-MS method for the analysis of complex peptide mixtures that allowed the detection of all peptide sequences from the libraries after synthesis was used for sequence deconvolution of sorted hits. Unfortunately, the accuracy of sequence deconvolution was

compromised upon bead incubation with living cells and subsequent sorting. Even adding thorough washing steps, the purity of the sample was affected and led to MS spectra with high noise, which made peptide hits deconvolution extremely challenging. However, we were able to detect some masses by UPLC-MS and analyzed the peptide fragmentation pattern by high resolution MS in order to confirm peptide hits identity (Table 5).

Table 5. List of detected Library 1 peptides after incubation with breast cancer cells. UPLC-MS detected masses (m/z) at specific retention times after library synthesis and detected masses after beads incubation with cells and subsequent sorting of positive hits are shown. Blue, masses of hits detected by UPLC-MS whose MS/MS fragmentation pattern was confirmed by high resolution MS; Red, masses of hits detected by UPLC-MS whose fragmentation pattern did not fit with their expected fragmentation; Gray, masses of hits detected by UPLC-MS in samples where the background noise was too high to draw conclusions.

				Detected mass after incubation with breast cancer cells					
				Experiment 1		Experiment 2		Experiment 3	
Permutations	Monoisotopic mass	Retention time (min)	UPLC-MS detected mass	MCF7 1920 events	BoM2 5185 events	MCF7 3731 events	BoM2 11835 events	MCF7-Empty 1665 events	MCF7-MAF 2799 events
wsGaay	653,280975	7,03	653,14				653,21	653,21	653,21
Geraay	665,313375	1,83	665,15					665,34	
wsGasy	669,275875	6,9	669,21				669,27		
weGaay	695,291525	7,1	695,09				695,09		
Gsrray	708,3669	0,45	708,22				708,21		
weGasy	711,286425	7,03	711,22				711,28	711,28	
wsGray	738,34505	6,68	738,1		738,28				
wsraay	752,3607	6,73	752,23	752,35	752,28			752,45	752,41
wsGrsy	754,33995	6,58	754,23		754,28		754,22		
wsrasy	768,3556	6,68	768,23		768,16				
weGray	780,3556	6,73	780,17					780,3	
werasy	810,36615	6,73	810,18					810,3	810,24
wsrray	837,424775	6,4	837,37	837,37	837,17	837,18	837,29		837,18
wsrrsy	853,419675	6,4	853,06	853,25	853,17	853,37	853,29		853,31

We observed that low numbers of sorted events negatively impacted sequence deconvolution. Although we managed to detect some masses in samples with low sorted events, the background noise of MS spectra was very high and the fragmentation pattern of the corresponding peptides either did not fit the expected pattern or could not be assessed. In contrast, despite having some background noise as well, peptide masses of samples with higher number of sorted events were detected by UPLC-MS and their correct fragmentation pattern was confirmed by high resolution MS (Figure 22).

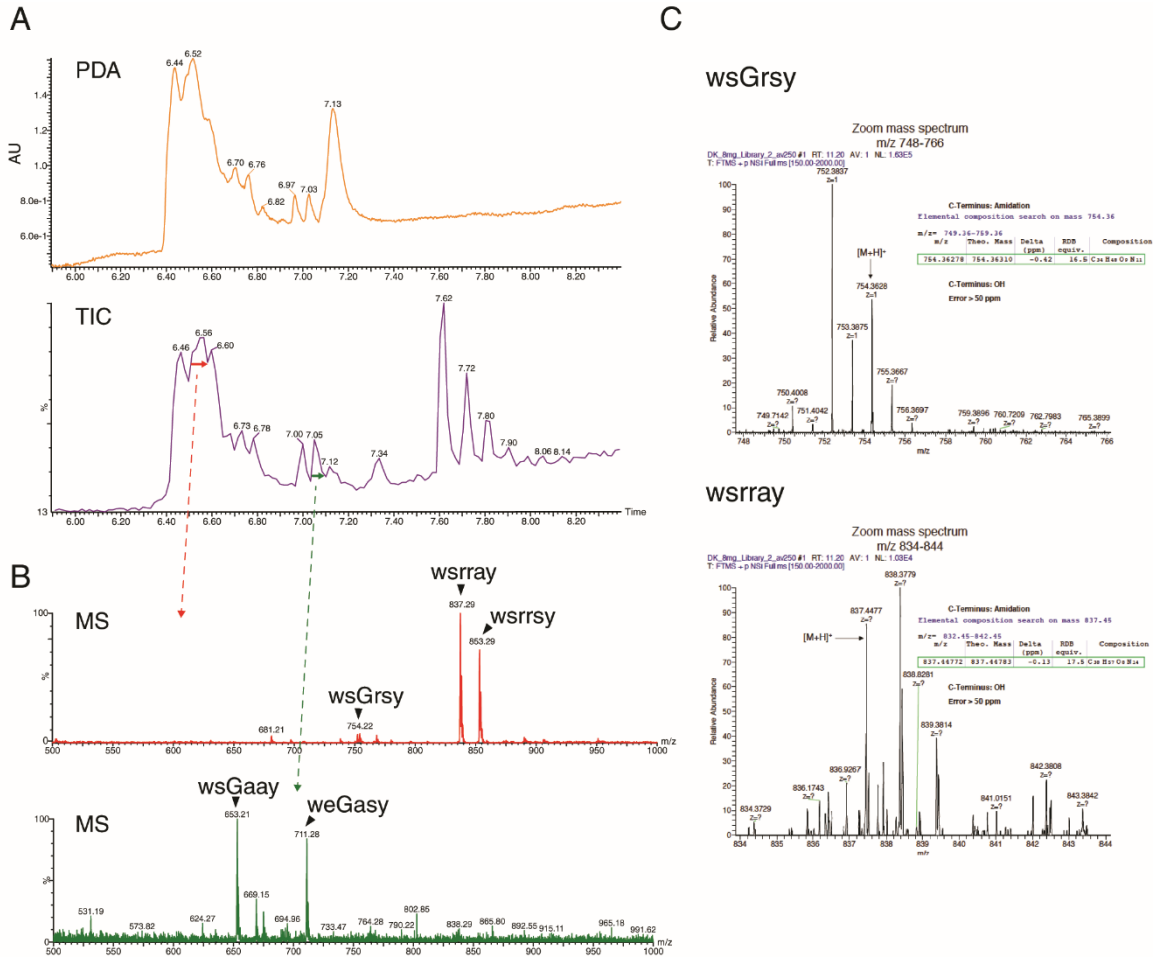


Figure 22. UPLC-MS and high-resolution MS data for Library 1 peptides after incubation with BoM2 cells and subsequent sorting (Experiment 2). (A) Photodiode array (PDA) chromatogram and total ion chromatogram (TIC) from 6 to 8.5min retention time (Rt) region. Arrows in TIC indicate the representative peaks for which MS spectra are shown. Detection wavelength 214nm. **(B)** Example of MS spectra at Rt=6.56min (red) and Rt=7.05min (green). Peptide masses found at Rt=6.57min, $[H+1]^+$: 754.22 (*wsGrsy*), 837.29 (*wsrray*) and 853.29 (*wsrrsy*); Peptide masses found and at Rt=7.05min, $[H+1]^+$: 653.21 (*wsGaay*), 711.28 (*weGasy*). **(C)** Example of high-resolution MS spectra corresponding to *wsGrsy* and *wsrray* sequences.

Although further experiments must be performed in order to endorse our observations, detected masses confirmed by high resolution MS suggested that *wsrray* and *wsrrsy* peptides were able to equally bind highly and poorly bone metastatic cells whereas *wsGray*, *wsGrsy*, *wsrasy*, *wsGasy*, *weGaay*, *Gsrray* peptides seemed to be specific of highly metastatic BoM2 cells. *wsraay*, *wsGaay* and *weGasy* peptides were also found in both highly and poorly bone metastatic cell populations, although their presence in MCF7 cells was not confirmed by high resolution MS.

In order to obtain samples with higher purity we tested different methods to remove bound cells from isolated beads. We attempted to remove cells by treating the beads with ethanol or boiling them in a 1% SDS solution for 5min but, unfortunately, we could not eliminate the background noise of the obtained MS spectra. Moreover, considering that the major part of the bead surface does not directly interact with cells, which leaves sufficient peptides on the bead surface for cleaving and deconvolution, we also attempted to sort beads in tubes with cellulose membranes. We expected that cellulose membranes ensured the necessary molecular weight cut-offs to keep cleaved peptides that were not directly involved in cell binding while discarding cells and beads without using any reagent for cell removal. Nevertheless, the background noise of the MS spectra was not reduced.

In summary, despite optimizing an efficient method to search for affinity interactions between several compounds and BCa cells and effectively sorting and collecting the positive hits, we faced technical difficulties to obtain reliable peptide deconvolution data due to sample complexity. Although we were able to generate some promising results, further technique optimization would be required in order to consider OBOC libraries viable for high-throughput cell-based screenings.

Chapter II: Characterization of the MAF interactome by proximity-dependent biotin identification (BioID)

The transforming activity of MAF has proven to be dependent on cell type and culture conditions. This finding suggests that MAF interacts with proteins that might modulate its oncogenic activity[1]. Furthermore, the expression of MAF physiological targets is not deregulated in cancer cells, which indicates that oncogenic MAF activity results from a switch of targets genes, probably due to modifications in MAF dimerization with other transcription factors and interacting partners[2]. This set of observations prompted us to explore the MAF interactome to gain some insight into the molecular mechanisms by which MAF regulates gene expression in BCa cells, thereby favoring metastasis. This bottom up approach represents a different strategy to identify potential targets to tackle MAF-driven bone metastasis, since it might reveal novel druggable proteins with the ability to modulate MAF-transforming activity.

Proximity-dependent biotin identification (BioID) is a unique method for the identification of physiologically relevant protein-protein interactions that occur in living cells. The BioID approach makes use of a promiscuous biotin ligase (*Escherichia coli*-derived BirA R118G*) fused in-frame to a bait protein to covalently label proximal proteins with biotin. Biotinylated proteins can be purified by conventional biotin capture methods and identified by MS^3 (Figure 23). Traditional approaches such as yeast-two-hybrid (Y2H) or co-immunoprecipitation (co-IP) are either performed in a non-natural cellular environment, often lacking the presence of associated binding partners and the machinery for post-translational modifications required for proper folding, or demand harsh conditions to solubilize bait proteins, resulting in the disruption of weak interactions with partner proteins due to stringent lysis and washings. The BioID technique allows to overcome these limitations and the subsequent loss of candidates by allowing the assessment of protein interactions under physiological conditions. Moreover, since proteins are labeled with biotin before solubilization, BioID allows for harsh lysis and yet offers the possibility to detect weak and transient interactions.

Thus, we believe that the description of the MAF interactome by means of a BioID screen will provide knowledge on the molecular mechanisms of MAF-driven bone metastasis and will open up new possibilities for the treatment of this condition.

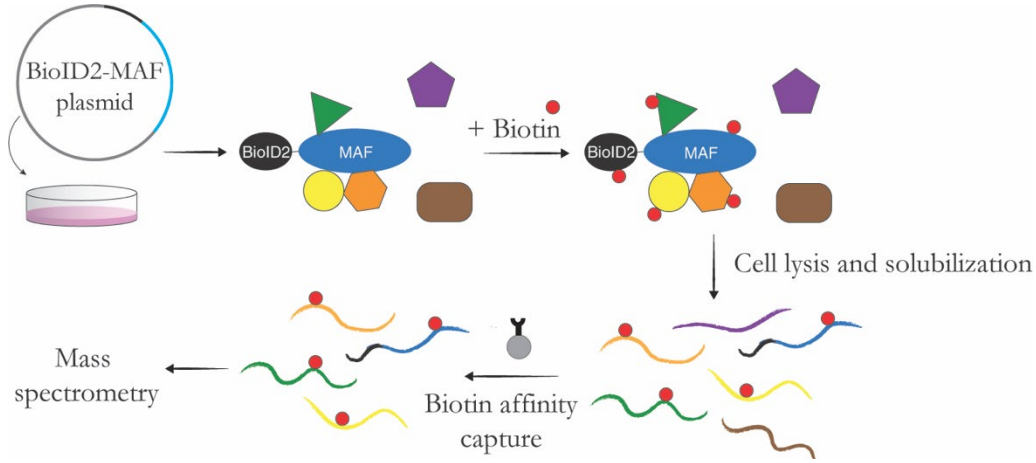


Figure 23. Schematic representation of the proximity-dependent biotin identification (BioID) method. Expression of a promiscuous biotin ligase (BioID2) fused to a protein of interest (MAF) in live cells leads to biotinylation of proximal and interacting proteins upon biotin addition. Following stringent cell lysis and protein denaturation, biotinylated proteins are captured by streptavidin-conjugated beads and identified by mass spectrometry. Adapted from Roux *et al.* (2012).

Generation and functional validation of BioID2-MAF fusion proteins

The first requirement for applying the BioID system was to fuse MAF (short and long isoforms) to a promiscuous biotin ligase. We decided to use a second-generation biotin ligase called BioID2, derived from *Aquifex aeolicus*, instead of the original ligase for BioID. BioID2 is very similar to the original BirA R118G* ligase but lacks the DNA-binding domain, which makes it smaller, thereby allowing a more-selective targeting of fusion proteins[4]. BioID2 requires less biotin supplementation than BirA R118G* to achieve similar biotinylation and enables enhanced labeling of proximate proteins. Thus, a tagged BioID2 ligase was fused in-frame to either the N-terminus (myc-BioID2-MAF) or the C-terminus (MAF-BioID2-HA) of the two MAF isoforms, short (S) and long (L). Since we had little information as to how N- or C-terminal fusions would affect MAF function and interaction with other proteins, we tried both fusions in parallel.

Another requirement for the correct functioning of the BioID method was to ensure that the newly generated fusion proteins were functional. Functional validation is challenging for proteins that do not have a well-known phenotypic outcome. Therefore, this validation is

typically approached by comparing the localization of the fusion protein to that of the endogenous protein by immunofluorescence (IF).

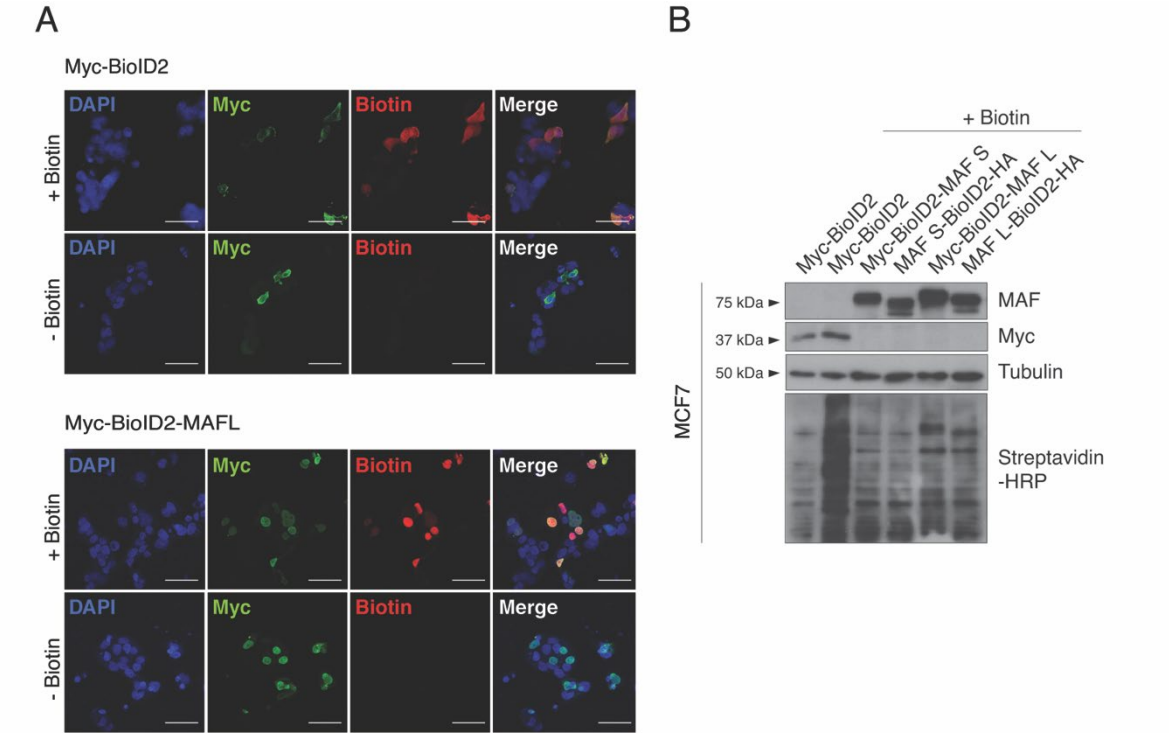


Figure 24. Functional validation of BioID2-MAF fusion proteins. (A) Representative immunofluorescence images of MCF7 cells transfected with the myc-BioID2 tag alone or the myc-BioID2-MAFL construct. The myc-BioID2 tag alone is distributed throughout the nucleus and cytoplasm (top), whereas the myc-BioID2-MAFL fusion protein localizes primarily in the nucleus (bottom). Biotinylated proteins, detected with fluorescently-labeled streptavidin, colocalize with BioID2 when cells are cultured with excess biotin (50M). DNA is labeled with DAPI. Scale bar, 50m. **(B)** Representative immunoblot showing the expression of either the myc-BioID2 tag alone and BioID2 fused to MAF (N- and C-terminal fusion, MAF short and long isoforms) in MCF7 cells analyzed after 24 h incubation in medium with or without excess biotin. Expression of the BioID2 biotin ligase leads to biotinylation of endogenous proteins, detected using HRP-conjugated streptavidin.

Hence, MCF7 cells were transfected with the newly generated expression plasmids and a control plasmid (myc-BioID2) for the expression of the biotin ligase without a fused protein of interest. After transfection, cells were cultured with or without supplemental biotin for 24 h prior to fixation and IF microscopy analysis to assess the localization of the control and fusion proteins as well as their ability to biotinylate proximal proteins upon biotin addition. Myc-BioID2 tag alone was distributed throughout the nucleus and cytoplasm, while BioID2-

tagged MAF showed a clear nuclear localization, thereby suggesting that the BioID2 tag did not hinder the biological function of MAF (Figure 24A). Moreover, biotinylated proteins were detected only after biotin supplementation and colocalized with BioID2, confirming the capacity of both myc-BioID2 tag alone or BioID2-tagged MAF proteins to biotinylate upon biotin addition.

Consistently, immunoblot analysis of MCF7 cell extracts after transfection with control and BioID2-tagged MAF and biotin addition confirmed the expression of the fusion proteins as well as biotinylation of endogenous proteins, probed by HRP-conjugated streptavidin (Figure 24B). Of note, myc-BioID2 alone biotinylated a wider range of endogenous proteins compared to BioID2-tagged MAF proteins. Fluorescence microscopy data indicates that myc-BioID2 biotinylates both nuclear and cytoplasmic proteins while BioID2-tagged MAF only labels nuclear proteins. It is worth to mention that the electrophoretic mobility, and therefore the identity, of most biotinylated proteins in myc-BioID2 and BioID2-tagged MAF differ.

Identification of the MAF interactome

Following validation of the N-terminal and C-terminal tagged MAF S and MAF L proteins, BioID was conducted on MCF7 cells. To this end, cells were transfected with the four plasmids for the expression of MAF fusion proteins but also with the myc-BioID2 control plasmid. The latter was used to identify proteins that may be randomly biotinylated or proteins that adhere to the beads used for purification, which is important for an accurate selection of interactors after the BioID pull-down. After transfection, cells were cultured in the presence of biotin for 24 h and biotinylated proteins were isolated with streptavidin-conjugated magnetic beads, washed and subjected to trypsin proteolysis. The released peptides were identified by nano-LC-MS/MS and high-confidence interactors were defined by comparing the myc-BioID2 control spectral counts with the BioID2-tagged MAF spectral counts of 2 independent biological replicates and filtering those with a Bayesian false discovery rate (BFDR) less than 0.02 and a 3-fold enrichment.

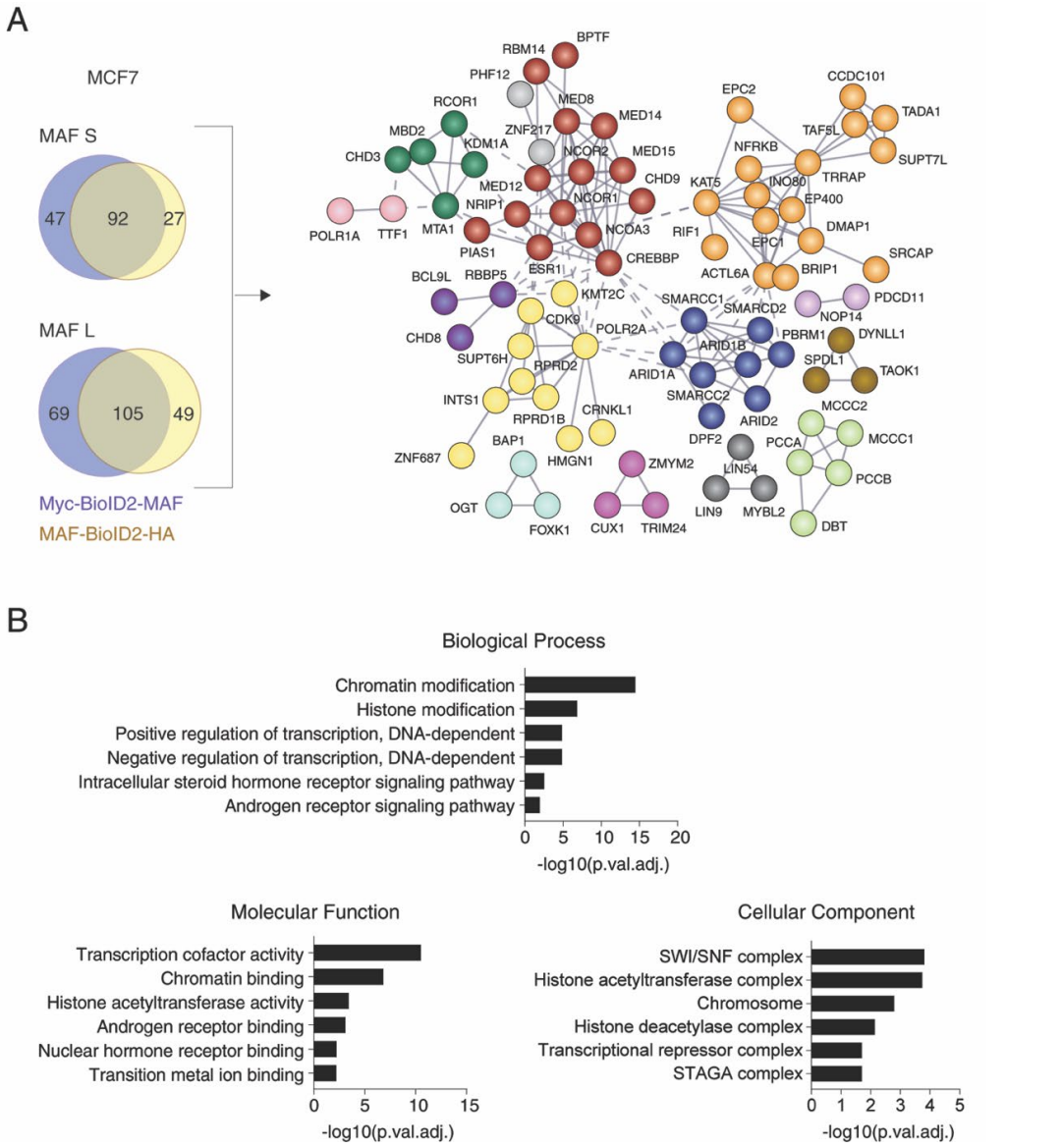


Figure 25. Characterization of the MAF interactome in MCF7 cells. (A) Network diagram of high-confidence MAF interactors (proteins identified with a Bayesian false discovery rate (BFDR) <0.02 and spectral counts at least 3-fold higher in BioID2-MAF fusion protein samples compared to the myc-BioID2 control) identified in MCF7 cells using the BioID system. Four BioID2-MAF fusion proteins (N- and C-terminal fusion, MAF short and long isoforms) were used as baits. Venn diagrams show MAF interactors discovered with each bait. Protein names were imported into the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) database for visual representation using publicly available protein interaction data. The thickness of the line in the network edges indicates the degree of confidence prediction of the interaction. Disconnected nodes in the network are hidden. The Markov Cluster (MCL) algorithm was used to cluster the proteins displayed in the network. n=2 biological replicates. Protein-protein interaction enrichment p-value < $1e^{-16}$. **(B)** Biological process, molecular function and cellular component gene ontology analysis of MAF interactors.

The analysis identified 139 high-confidence proximity interactors using the N-terminal-tagged MAF S isoform and 119 interactors using the C-terminal-tagged MAF S isoform (Figure 25A). The MAF L isoform was similarly analyzed and the experiment revealed 174 high-confidence interactors using the N-terminal-tagged isoform and 154 interactors using the C-terminal-tagged isoform. Since both MAF S and L isoforms were reported to equally promote bone relapse in BCa cells[303], we selected the high-confidence interactors identified in both N- and C-terminal-tagged MAF S (92 common interactors) and those identified in both N- and C-terminal-tagged MAF L (105 common interactors) to obtain a shortened list of 126 MAF-proximal proteins (71 common in the 4 conditions), thus reducing experimental noise. The complete list of MAF-proximal proteins in MCF7 cells is provided in Supplementary table 1.

CREBBP, a well-characterized MAF interactor[355], was found among the defined MAF-proximal proteins, which confirmed the correct functioning of the BioID experiment and a proper selection of MAF high-confidence interactors. Moreover, the set of 126 protein was strongly enriched (p-value < $1e^{-16}$) for known protein-protein interactions among themselves, which indicates the existence of biologically relevant complexes among MAF interactors. These included some of the major chromatin-remodeling complexes such as SWI/SNF, INO80, NurD and CoREST. Biological functions such as chromatin modification and transcriptional regulation are regulated through these complexes and their alteration has been implicated in cancer. Consistently, the group of MAF-proximal proteins was statistically enriched for well-known molecular functions that influence gene expression such as transcription cofactor activity, chromatin binding and histone acetyltransferase activity (Figure 25B). Thus, our data show that MAF associates with several protein complexes in the nucleus that may influence histone modification to generate transcriptionally active or repressive chromatin structures, thereby fine-tuning gene expression regulation.

Notably, gene ontology (GO) analysis revealed that a subset of the identified MAF interactors were involved in steroid hormone receptor signaling pathways (Figure 25B). Consistently, one of the statistically enriched molecular functions of MAF interactors was nuclear hormone receptor binding. In fact, ER itself emerged as a MAF interactor (Figure 25A).

Clinical data had demonstrated that patients respond differently to bisphosphonates treatment for metastatic BCa depending on their menopausal status as well as on MAF status. Tumors with high MAF expression levels become more aggressive after bisphosphonates treatment in pre-menopausal patients. Thus, given that this observation supports a potential role for E2,

the main ER ligand, in the modification of MAF-positive tumors behavior and that ER is the main driver of ER+ BCa, we considered the MAF-ER interaction of major interest. Our BioID data opened up the possibility of a cooperation between MAF and ER transcription factors. Additionally, interaction with known ER cofactors further supported the potential regulation of a joint transcriptional program that might contribute to BCa tumor aggressiveness.

Reciprocally, we wanted to investigate whether the presence of ER influences MAF interaction with other partners. Thus, we also explored the MAF interactome in an ER- setting. To this end, we transfected MDA-MB-231 cells with the previously used myc-BioID2 control and the four BioID2-tagged MAF plasmids, cultured them in medium supplemented with biotin for 24 h and validated the expression of the fusion proteins and their ability to biotinylate endogenous proximal proteins by immunoblot analysis (Figure 26A).

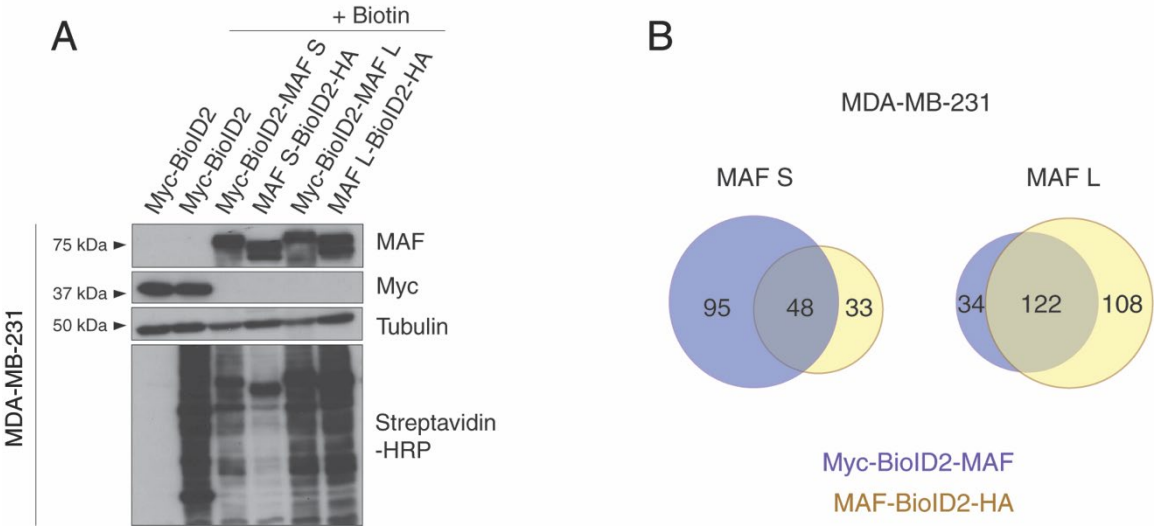


Figure 26. Characterization of the MAF interactome in MDA-MB-231 cells. (A) Representative immunoblot showing the expression of myc-BioID2 alone and BioID2-tagged MAF proteins (N- and C-terminal fusion, MAF short and long isoforms) in MDA-MB-231 cells analyzed 24 h after biotin supplementation. Expression of the BioID2 biotin ligase leads to biotinylation of endogenous proteins, detected using streptavidin-HRP. **(B)** Venn diagrams showing high-confidence MAF interactors (proteins identified with a Bayesian false discovery rate (BFDR) <0.02 and spectral counts at least 3-fold higher in BioID2-MAF fusion protein samples compared to the myc-BioID2 control) detected by BioID in MDA-MB-231 cells for each bait. n=1 biological replicate.

After validation, we performed a BioID pull-down that revealed 143 high-confidence proximity interactors using the N-terminal-tagged MAF S isoform, 81 interactors using the C-

terminal-tagged MAF S isoform, 156 high-confidence interactors using the N-terminal-tagged MAF L isoform and 230 interactors using the C-terminal-tagged MAF L isoform (Figure 26B). The complete list of MAF-proximal proteins in MDA-MB-231 is provided in Supplementary table 2.

Next, among the identified MAF-proximal proteins in MDA-MB-231, we analyzed the presence of representative components of the major chromatin-remodeling complexes that were found to interact with MAF in MCF7 cells as well as the presence of ER and some of its known coregulators.

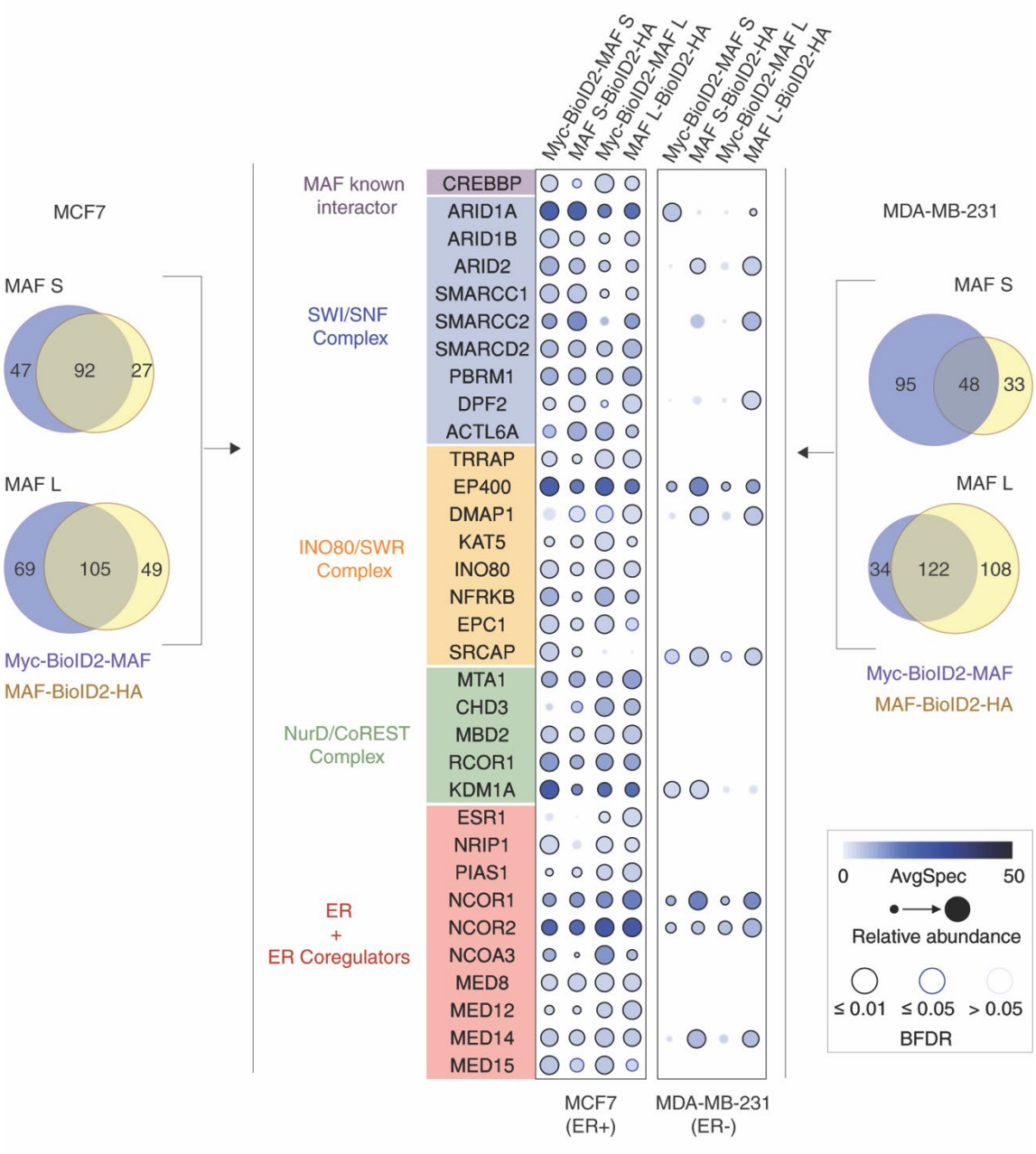


Figure 27. Bait-pray dot plot for selected MAF proximity interactors in MCF7 and MDA-MB-231 cells. Selected high-confidence MAF interactors discovered by BioID in the estrogen receptor-positive (ER+) cell line MCF7 are organized by protein complexes or according to their known functions and their abundance is visualized in both MCF7 and in the ER- cell line MDA-MB-231. Venn diagrams show MAF interactors detected by BioID in MCF7 and MDA-MB-231 cells for each bait. Dotplot was generated using ProHits-viz. Briefly, the color of the dots shows the average spectral count for each indicated interactor, the size of the dot represents the relative abundance of the interactor across the 4 different baits (MAF short and long isoforms, N- or C-terminal tag), and the edge color reflects the Bayesian false discovery rate (BFDR) value associated with each bait-pray interaction.

We observed that, in an ER- setting, MAF retained the ability to interact with some components of the SWI/SNF, INO80, NurD and CoREST chromatin-remodeling complexes. However, there was a significant loss of a subset of interactors described in MCF7 cells (Figure 27). As expected, no interaction with ER was detected in MDA-MB-231 cells.

Collectively, these data suggest that MAF interaction with specific partners is dependent on cell type and that the presence of ER might influence MAF association with other proteins in BCa cells.

Characterization of the MAF-ER interaction

To validate some of the interactions revealed by the BioID approach we employed co-immunoprecipitation (co-IP) in MCF7 cells with MAF overexpression. We tagged MAF S and L isoforms with a double hemagglutinin (HA) tag to facilitate MAF detection, isolation and purification. Both HA-tagged MAF S and L isoforms were captured and pulled down using magnetic beads coated with anti-HA-tag antibodies and co-immunoprecipitated selected members of important chromatin remodeling complexes such as ARID1A, a subunit of the SWI/SNF complex, and MTA1 and KDM1A, subunits of the NurD complex (Figure 28A). Moreover, both HA-tagged MAF isoforms also co-immunoprecipitated ER as well as its well-known cofactors NCoR1 and NCoA3.

Next, to further validate the selective association of MAF with ER, we performed a proximity ligation assay (PLA) in MCF7 cells. This technique allows *in situ* detection of protein interactions by using species-specific secondary antibodies conjugated to single-stranded

oligonucleotides which, when in close proximity, are ligated to form a circular DNA template. This circular DNA can be amplified and visualized by means of fluorescently labeled oligonucleotides and fluorescence microscopy.

Although the MAF-ER interaction seemed to be enriched for the MAF L isoform in the BioID experiment, this interaction was validated by Co-IP for both S and L isoforms. Yet, to simplify future experiments, we selected the L isoform.

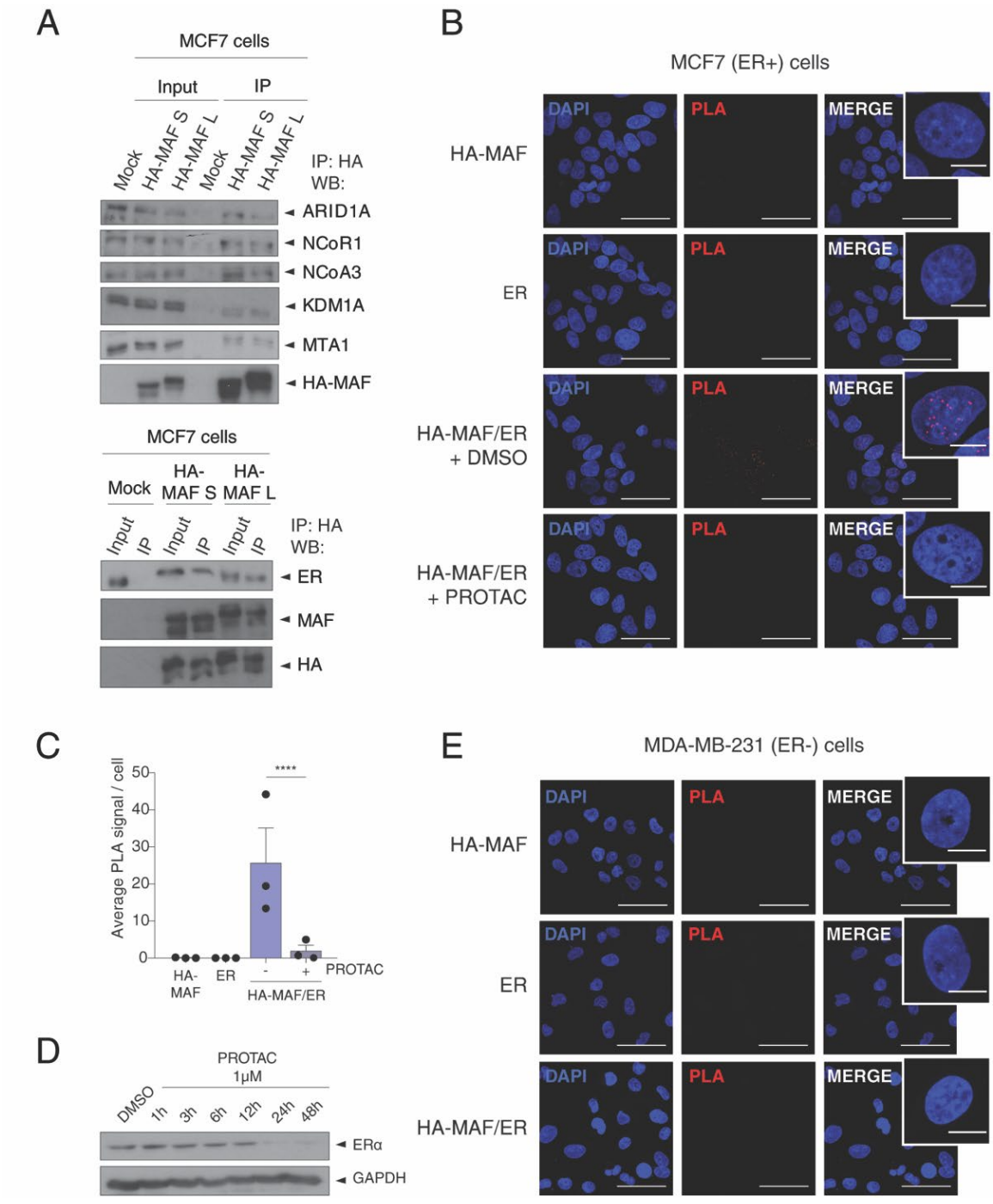


Figure 28. MAF interacts with the Estrogen Receptor (ER). (A) Immunoblot of HA co-immunoprecipitation of HA-tagged MAF (short (S) and long (L) isoforms) with endogenous ARID1A, NCoR1, NCoA3, KDM1A, MTA1 (top) and ER (bottom). **(B)** Proximity ligation assay (PLA) of HA or ER antibody alone or both antibodies together in stable MAF L-overexpressing MCF7 cells treated either with DMSO or a PROTAC against ER (1 M) 24 h prior to fixation . Representative confocal microscopy images for PLA red signal and DAPI nuclear staining with magnified inset are shown. Scale bar, 50m; inset scale bar, 10m. **(C)** PLA signal quantification. Each dot represents the average PLA signal from 123 to 201 nuclei per condition per biological replicate (n=3). Bars represent mean ± SD. Asterisks show statistical significance. **(D)** Representative immunoblot showing ER degradation by PROTAC (1M) in MCF7 cells at different timepoints. GAPDH was used as a loading control. **(E)** PLA of HA or ER antibody alone or both antibodies together in stable MAF L-overexpressing MDA-MB-231 cells. Representative confocal microscopy images for PLA red signal and DAPI nuclear staining with magnified inset are shown. Scale bar, 50m; inset scale bar, 10m.

Significantly higher fluorescence signal was detected in cell nuclei with both HA and ER antibodies compared to the single antibody controls, thus supporting an interaction between MAF and endogenous ER (Figures 28B and C). To verify PLA specificity, we treated the cells with a proteolysis targeting chimera (PROTAC) against ER prior to evaluating the MAF-ER interaction. PROTACS are an emerging tool for targeted elimination of specific proteins. These molecules consist of a target protein ligand attached to an E3 ubiquitin ligase recognition motif by a flexible linker, thereby mediating ubiquitination of the target protein and subsequent proteasome degradation[5,6]. In collaboration with Dr. Antoni Riera's group, we synthesized a PROTAC against ER and confirmed that it remarkably decreased ER protein levels after 24 h (Figure 28D). Consistently, PLA signal was abrogated in MCF7 cells treated with this ER-PROTAC (Figures 28B and C). These results validate PLA specificity and further support the MAF-ER interaction.

Additionally, the MAF-ER specific interaction was corroborated in the ER- cell line MDA-MB-231. Coherently, no fluorescent PLA signal was observed neither when cells were probed with specific antibodies against HA or ER alone nor after addition of both antibodies (Figure 28D).

The elimination of key protein domains is a useful tool to map the protein region that interacts with a specific partner protein. Thus, to further explore MAF interaction with endogenous ER, we generated different plasmids for the expression of truncated MAF proteins as well as a control plasmid for full-length MAF L expression. Truncated MAF proteins included MAF (ΔC-t), which lacks the DNA-binding domain, MAF L (ΔN-t 1), which lacks part of the transactivation domain and MAF L (ΔN-t 2), which lacks the entire transactivation domain (Figure 29A).

First, we assessed protein localization by IF to test whether MAF truncation affected subcellular localization. We observed that deletion of the MAF C-terminal domain led to improper localization. MAF (ΔC-t) truncated form was detected throughout the nucleus and cytoplasm, thereby suggesting the presence of a nuclear retention signal within the C-terminal domain (Figure 29B). In contrast, deletion of either part or the entire transactivation domain did not affect MAF nuclear localization.

Then, full-length and truncated MAF proteins were tested for interaction with ER using the PLA technique. Significant PLA signal was detected in all samples probed with both HA and ER antibodies compared to single antibody controls (Figures 29B and C). All MAF truncated proteins seemed to retain the ability to interact with ER in the nucleus, even the MAF (ΔC-t) truncated form, despite its expression was not confined to the nuclear compartment. However, although not significant, PLA signal in cells expressing truncated MAF proteins was not as strong as in cells expressing full-length MAF. This reduction in the detected fluorescent signal was more pronounced and consistent in cells expressing the MAF L (ΔN-t 2) truncated form, thereby indicating a potential requirement of the region between amino acids 85 and 120 of the transactivation domain for binding to ER.

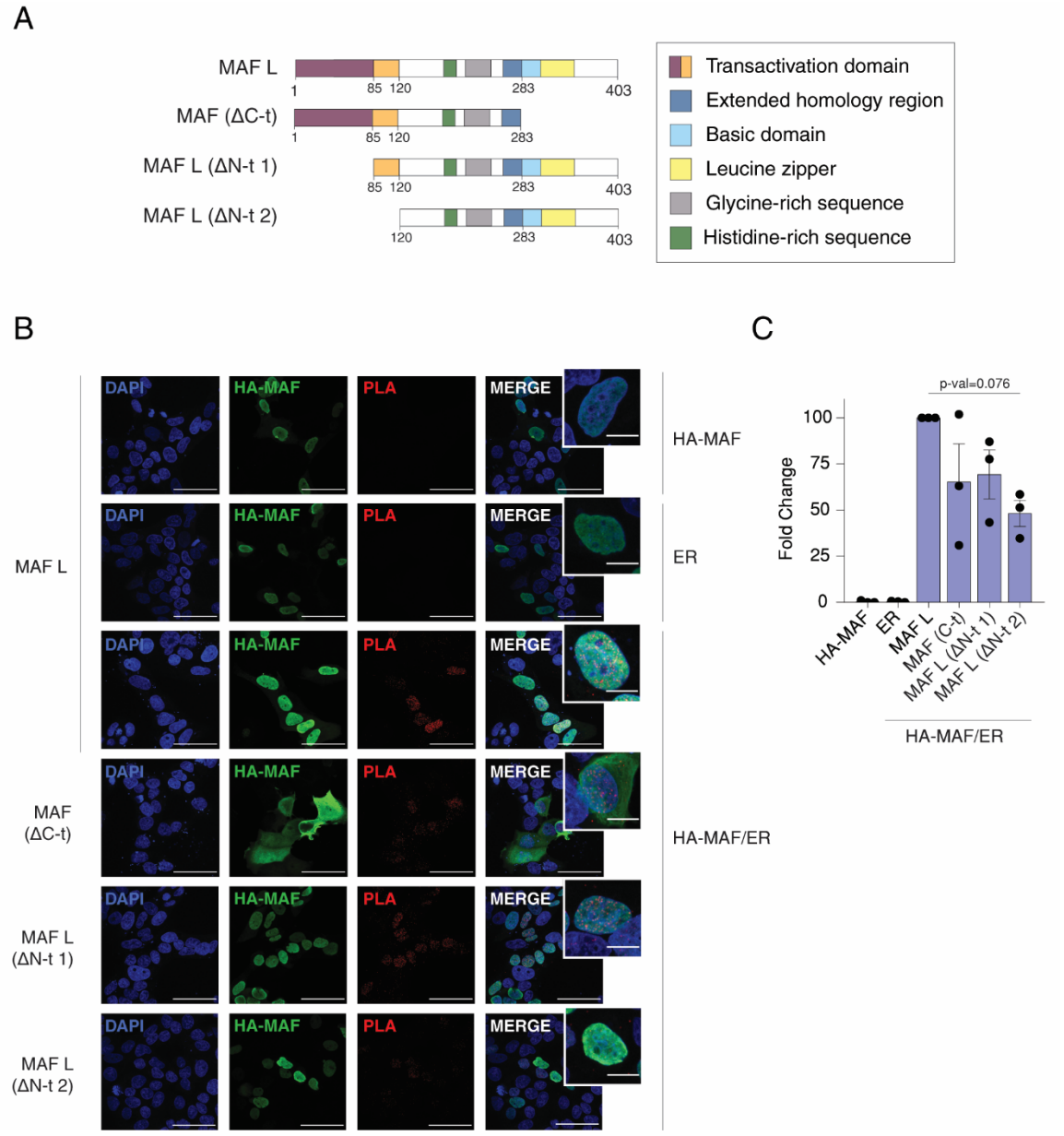


Figure 29. Mapping of the domain through which MAF interacts with the Estrogen Receptor (ER). (A) Schematic representation of full-length MAF (long isoform) and truncation constructs. **(B)** Proximity ligation assay (PLA) of HA or ER antibody alone or both antibodies together in MCF7 cells transiently transfected with HA-tagged MAF full-length and truncation constructs 24 h prior to fixation. Representative confocal microscopy images for HA-MAF expression (green), PLA signal (red) and DAPI nuclear staining with magnified inset are shown. Scale bar, 50m; inset scale bar, 10m. **(C)** PLA signal quantification. Each dot represents the average PLA signal from 66 to 129 nuclei per condition per biological replicate (n=3). Bars represent mean SD.

Chapter III: The molecular interplay of MAF and ER transcription factors in breast cancer cells

ER has a key role in the onset and progression of ER+ BCa by enhancing the expression of genes related to cell proliferation and survival[356]. This protein can regulate transcription through direct binding to DNA after E2 stimulation or by being tethered to DNA through interaction with other transcription factors[357]. Evidence show that ER can directly interact with members of the AP-1 family of transcription factors such as c-Fos and c-Jun and that genes containing AP-1 sites in their promoters are regulated by ER[358,359]. However, how these interactions affect transcriptional activity and how this supports cancer metastasis is not yet fully understood.

MAF, a transcription factor from the AP-1 family, acts in synergy with other transcription factors such as GATA3, STAT6 and Nuclear factor of activated T-cells (NFAT) to regulate the expression of specific genes in T-cells[360]. Moreover, MAF has been reported to physically associate with SRY-related HMG box (SOX) family members to bind and activate the promoter of certain genes[360]. Further MAF amplification was reported to trigger a genetic program to support BCa metastasis[303].

We revealed a novel interaction between ER and MAF, which suggests a cross-talk of both proteins in BCa cells. Since previous findings showed that AP-1 factors such as c-Fos and C-Jun are cooperating factors for ER, we hypothesize that MAF may also modulate ER function in BCa cells, triggering the expression of a genetic program that might have an impact on ER+ BCa progression and in bone metastasis development. To this end, a detailed understanding of the cooperation between MAF and ER is crucial to understand the mechanisms of E2-mediated cancer progression and might lead to the identification of potential new targets, thereby facilitating the discovery of approaches to ultimately target BCa metastasis.

Gene expression profiles in response to MAF overexpression and estrogen stimulation

We first assessed whether the presence of MAF modulates E2 response in BCa cells. Control (Mock) and MAF-overexpressing MCF7 cells were cultured in hormone-deprived media containing charcoal-stripped serum for 72h to deprive cells of E2 (Figure 30A). Then, cells were stimulated with 10nM E2 for 6 hours and gene expression changes were analyzed by RNA sequencing (RNA-seq).

Principal component analysis revealed that MAF-overexpressing cells diverged from control cells before and after E2 stimulation (Figure 30B). Consistently, different gene expression patterns were observed in MAF-overexpressing cells compared to control cells before and after E2 administration (Figure 30C). Some genes were up-regulated (cluster 1) or down-regulated (cluster 2) in MAF-overexpressing cells in comparison to control cells independently of E2 stimulation, thereby indicating that the expression of these genes is regulated by MAF. In contrast, two large sets of genes were up-regulated (cluster 3) or down-regulated (cluster 4) in response to E2 administration in both control and MAF-overexpressing cells. These groups reflect E2-dependent genes whose expression in independent of MAF. Interestingly, the presence of MAF and E2 had a synergistic effect on the up-regulation (cluster 5) or down-regulation (cluster 7) of two groups of genes. Moreover, some genes were up-regulated (cluster 6) or down-regulated (cluster 8) exclusively in the presence of both MAF and E2. These results show that there is a transcriptional modulation through the interactions of E2-ER with MAF. The complete list of the genes from each cluster can be found at Supplementary table 3.

Analysis of GO annotations revealed that the set of genes that were up-regulated in the presence of MAF (cluster 1), contained a significant overrepresentation of genes involved in hypoxia and epithelial to mesenchymal transition, which are important processes in cancer metastasis. Interestingly, genes down-regulated in the presence of MAF (cluster 2), were enriched for GO terms related to cell-cycle, including G2M checkpoint, E2F targets and mitotic spindle. These observations indicate that MAF overexpression confers metastatic traits to BCa cells but does not favor cell proliferation per se in this cellular context. Strikingly, genes involved in early and late estrogen response were also overrepresented, thereby indicating that, although in this specific cellular context these genes did not respond to E2 stimulation, MAF has the ability to control their basal expression levels (Figure 31A).

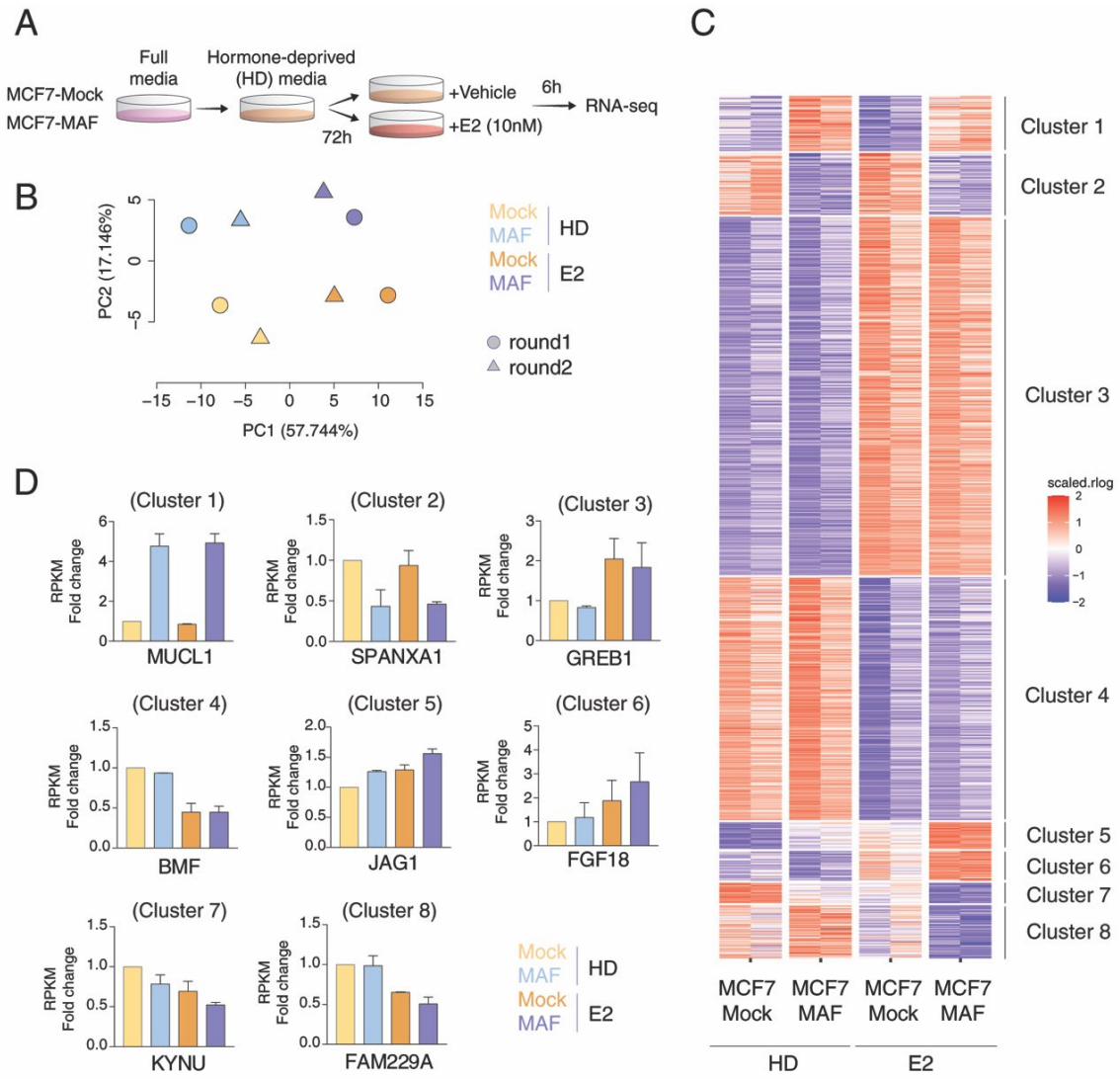
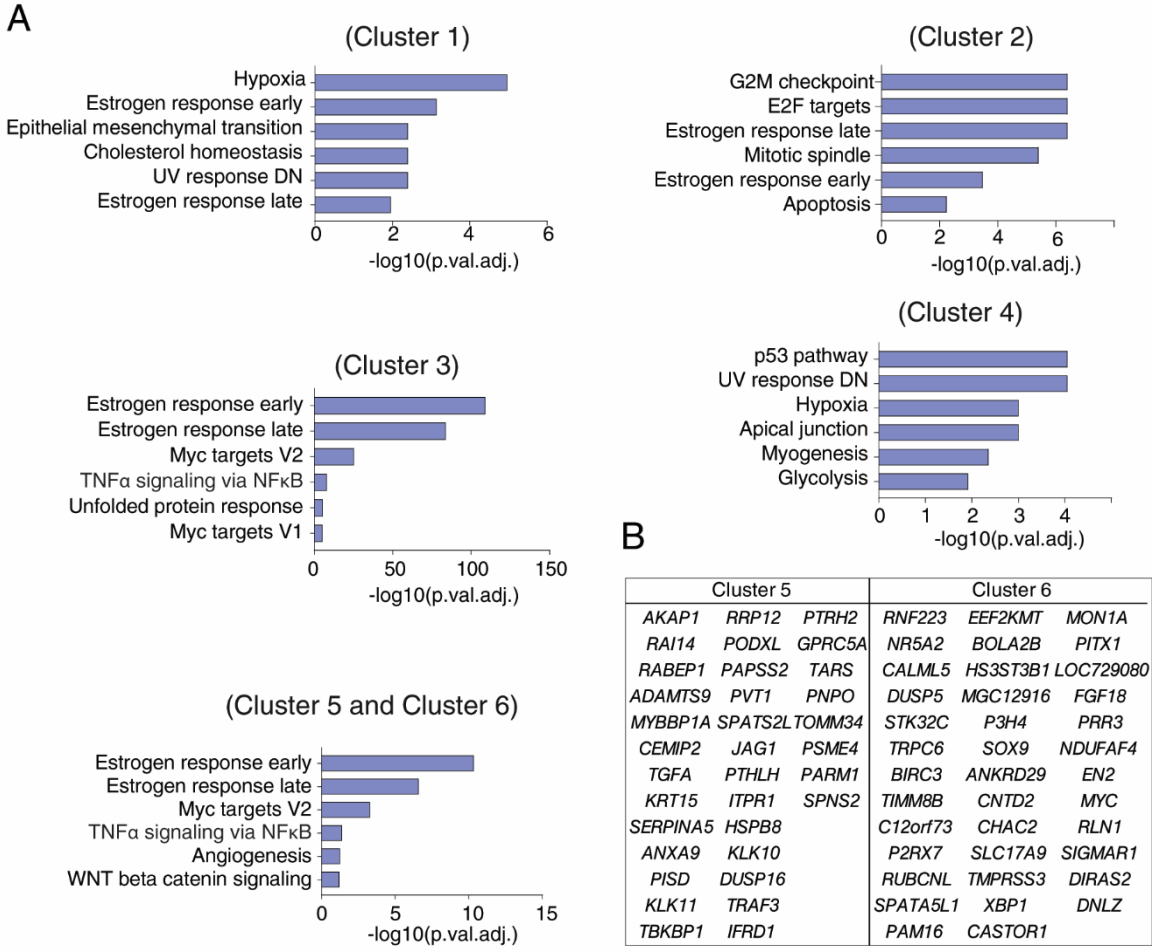


Figure 30. Gene expression profiles upon MAF overexpression and E2 stimulation. (A) Schematic representation of the experimental design. **(B)** Principal component analysis (PCA) on the rlog normalized expression matrix (showing first two components). Batch effect was adjusted gene-wise using a linear model. **(C)** RNA-seq heatmap of differentially expressed genes in MAF-overexpressing compared to control MCF7 cells upon administration of vehicle (HD) or 10nM estradiol (E2). Expression profiles were grouped in 8 clusters based on comparisons between the four conditions. Cluster 1, MAF up-regulated genes; cluster 2, MAF down-regulated genes; cluster 3, E2 up-regulated genes; cluster 4, E2 down-regulated genes; cluster 5, E2 and MAF up-regulated genes; cluster 6, up-regulated genes when E2 and MAF; cluster 7, E2 and MAF down-regulated genes; cluster 8, downregulated genes when E2 and MAF. Red indicates up-regulation and blue shows down-regulation. n=2 biological replicates. **(D)** RNA-seq expression of a selected gene from each cluster represented by fold change of RPKM values.

E2-induced genes (cluster 3) included several well-known ER target genes such as *GREB1*, *PGR*, *CCND1* and *CXCL12* (Figure 30C and D). This panel of genes were enriched for GO terms involved in E2 response pathways and were also myc targets. Consistently, genes down-regulated by E2 (cluster 4) included *BMF, IGFBP3, NR3C1* and *BRB7*, whose expression is typically reduced upon E2 addition in MCF7 cells[361]. Together, these data confirmed a successful E2 stimulation.

A



B

Cluster 5			Cluster 6		
AKAP1	*RRP12*	*PTRH2*	*RNF223*	*EEF2KMT*	*MON1A*
RAI14	*PODXL*	*GPRC5A*	*NR5A2*	*BOLA2B*	*PITX1*
RABEP1	*PAPSS2*	*TARS*	*CALML5*	*HS3ST3B1*	*LOC729080*
ADAMTS9	*PVT1*	*PNPO*	*DUSP5*	*MGC12916*	*FGF18*
MYBBP1A	*SPATS2L*	*TOMM34*	*STK32C*	*P3H4*	*PRR3*
CEMIP2	*JAG1*	*PSME4*	*TRPC6*	*SOX9*	*NDUFAF4*
TGFA	*PTHLH*	*PARM1*	*BIRC3*	*ANKRD29*	*EN2*
KRT15	*ITPR1*	*SPNS2*	*TIMM8B*	*CNTD2*	*MYC*
SERPINA5	*HSPB8*		*C12orf73*	*CHAC2*	*RLN1*
ANXA9	*KLK10*		*P2RX7*	*SLC17A9*	*SIGMAR1*
PISD	*DUSP16*		*RUBCNL*	*TMPRSS3*	*DIRAS2*
KLK11	*TRAF3*		*SPATA5L1*	*XBP1*	*DNLZ*
TBKBP1	*IFRD1*		*PAM16*	*CASTOR1*	

Figure 31. MAF modulates the estrogen (E2) response in MCF7 cells. (A) Associated gene ontology (GO) terms for genes present in the indicated clusters. Clusters 5 and 6 were considered a single group of E2-induced and MAF-dependent genes. **(B)** Name of the genes contained in clusters 5 and 6.

Genes in clusters 5 and 6 reached their maximum expression levels upon E2 stimulation and in a MAF-overexpressing context. Since clinical data revealed that MAF-positive tumors behave more aggressively in the presence of E2, we considered this panel of genes of major

interest to find a biological rationale for this observation. Interestingly, among the MAF-ER synexpression group, there were some genes whose expression in BCa cells is associated with enhanced migration and EMT, such as *FGF18*[362], maintenance of a stem cell-like state, such as *SOX9*[363], or metastatic colonization of the bone, such as *JAG1* and *PTHLH*[249,303] (Figure 31B).

ER and MAF coregulated genes were mainly enriched in E2 response pathways (Figure 31A). These results suggest that, in response to E2, MAF cross-talks with ER and modulates the transcriptional activity of this transcription factor either by triggering a switch of target genes or by potentiating the expression of specific E2-responsive genes.

GO analysis in clusters 7 and 8 did not identify overrepresented families of genes involved in any specific hallmark. However, the limited number of genes in these groups precluded any statistically significant analysis due to insufficient power.

Reverse transcription quantitative polymerase chain reaction (RT-qPCR) of selected target genes confirmed the RNA-seq results in MCF7 cells (Figure 32). The expression of *FGF18*, *JAG1* and *PTHLH* was significantly higher in MAF-overexpressing cells after E2 stimulation compared to control MCF7 cells. Although it did not reach significance, *SOX9* showed the same expression pattern in these cells.

Next, we interrogated whether our findings in MCF7 cells were cell-line specific or could occur in other ER+ BCa cell lines such as T47D or ZR75. To this end, we generated MAF-overexpressing T47D and ZR75 cell lines and examined the expression of ER and MAF-coregulated genes after 6h of 10nM E2 stimulation.

MAF expression in T47D and ZR75 cells showed a higher fold increase over basal levels compared to MCF7 cells. *FGF18*, *JAG1* and *PTHLH* were strongly dependent on MAF expression in both T47D and ZR75 cells lines. However, the effect of E2 on the expression of these target genes was not as clear as in MCF7 cells. *SOX9* expression levels were only slightly increased in MAF-overexpressing ZR75 cells. Together, these data show that high MAF expression is associated with an increase in *FGF18*, *JAG1* and *PTHLH* mRNA levels in three different BCa cell lines.

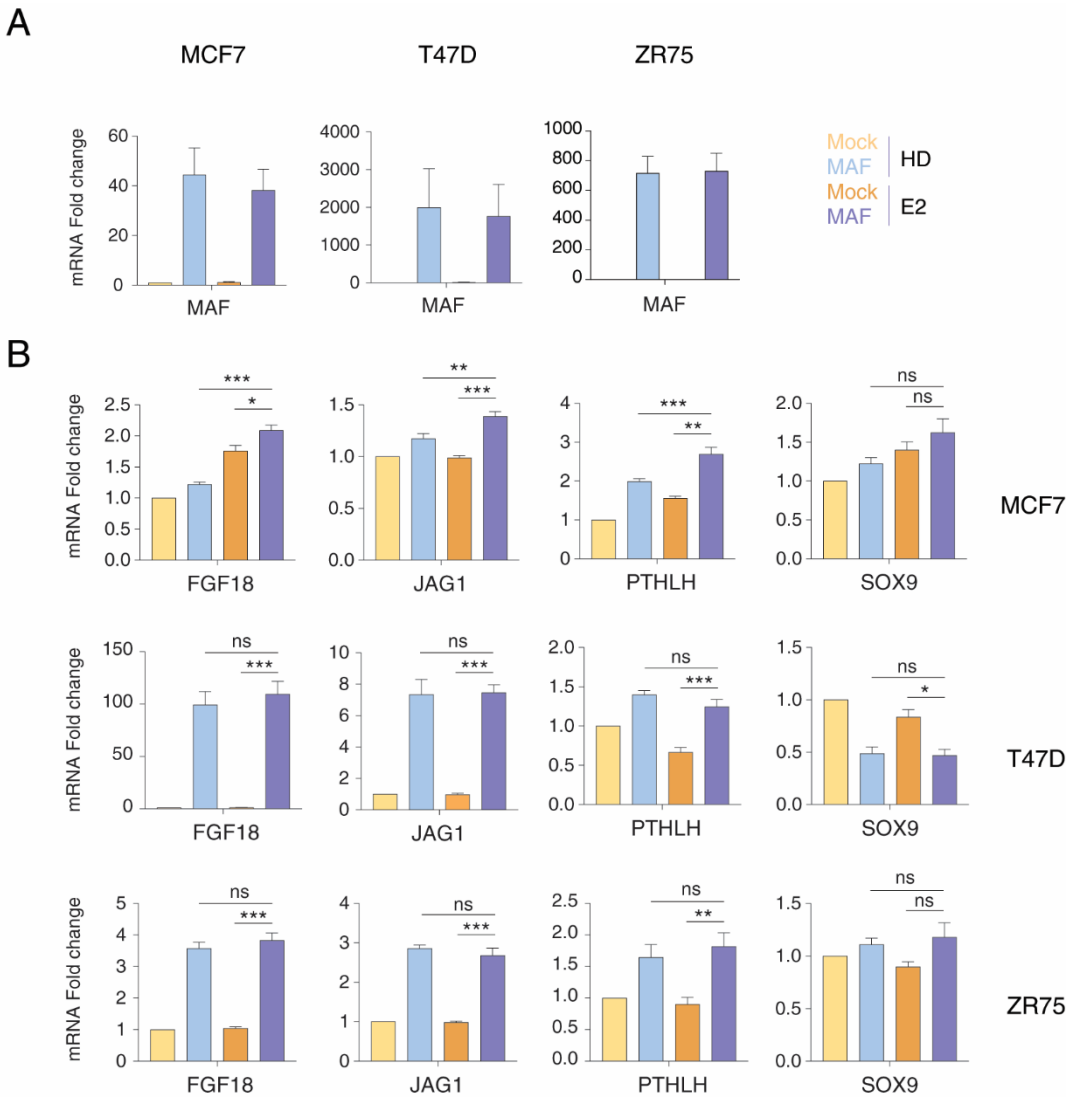


Figure 32. Reverse transcription quantitative polymerase chain reaction (RT-qPCR) validation of selected RNA-sequencing (RNA-seq) targets in MCF7 cells and its evaluation in the estrogen receptor-positive (ER+) breast cancer cell lines T47D and ZR75 (A) MAF expression analysis in MCF7, T47D and ZR75 cell lines before (HD) and after 10nM E2 administration. **(B)** Expression analysis of the MAF and E2-coregulated genes *FGF18*, *JAG1*, *PTHLH* and *SOX9* in MCF7, T47D and ZR75 cell lines. Expression was normalized to the housekeeping gene *GAPDH*. Data represent the average of 3 to 6 independent experiments. The symbols *, ** and *** indicate significant differences with $p < 0.05$, $p < 0.01$ and $p < 0.001$, respectively.

Changes in chromatin accessibility in response to MAF overexpression and estrogen stimulation

Our data supported that MAF is functionally involved in E2-mediated gene regulation in BCa cells. However, the mechanism by which MAF alters the E2 response remained unclear. Since BioID proximity labeling revealed that MAF interacts with several chromatin remodelers besides ER, we interrogated whether modulation of the E2 response in a MAF-overexpressing context was a consequence of changes in chromatin accessibility.

Because changes in chromatin architecture occur before differences in gene transcription, we stimulated hormone-starved control and MAF-overexpressing MCF7 cells with E2 for only 1 h and then mapped the chromatin landscape by assay of transposase accessible chromatin sequencing (ATAC-seq).

Data from three independent experiments revealed that chromatin was dynamic depending on both E2 induction and MAF overexpression (Figure 33A). Remarkably, chromatin accessibility was broadly increased in MAF-overexpressing compared to control MCF7 cells. Analysis of the distribution of differential peaks showed that chromatin opening occurred mainly at intronic and intergenic regions, but there was an overrepresentation of promoter regions for MAF-overexpressing samples compared to control samples (Figure 33B).

It has been shown that chromatin is largely remodeled upon E2 signaling[364]. Thus, to confirm a successful E2 induction, we interrogated ATAC-seq peaks ± 50 kb from genome wide transcription start sites (TSS) and assessed whether changes in chromatin accessibility correlated with gene expression. Indeed, the majority of ATAC-seq peaks associated to genes up-regulated by E2 identified in RNA-seq experiments became more accessible after E2 stimulation (Figure 33C). We also observed some chromatin opening in peaks associated to genes down-regulated by E2. However, closed chromatin regions were more representative in this set of genes.

Furthermore, we assessed whether MAF-dependent chromatin opening correlated with the expression of MAF-dependent genes. Strikingly, integrated RNA-seq and ATAC-seq analysis revealed that chromatin accessibility did not correlate with MAF-dependent gene expression. Both MAF up- and down-regulated genes correlated with increased chromatin opening (Figure 33D). This observation indicates that the regulation of MAF-dependent genes is not only

regulated by changes in chromatin accessibility but potentially requires the recruitment of coactivator and corepressor proteins.

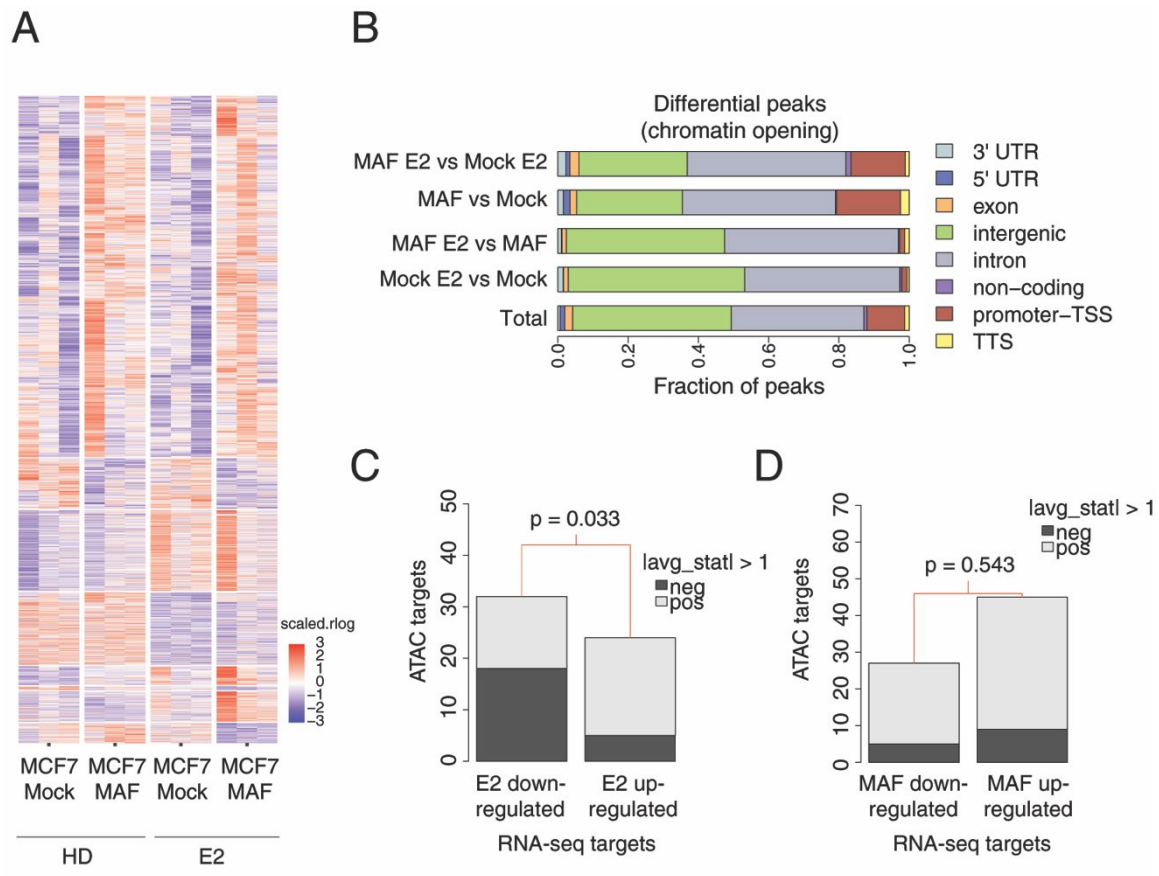


Figure 33. Accessible chromatin landscape induced by MAF and estrogen (E2). (A) Overview of ATAC-seq normalized data for most variable peaks. **(B)** Annotation of differential peaks (only chromatin opening) showing overrepresentation of promoter regions for MAF-overexpressing samples against Mock samples. **(C)** RNA-seq E2-responsive genes (both down- and up- regulated) enrichment in ATAC-seq peaks (distance to transcription start site (TSS) < 50kb). The number of E2-target genes with an average test statistic (for comparisons Mock E2 vs Mock and MAF E2 vs MAF) greater than 1 (in absolute value) in the ATAC-seq data is shown. E2-targets correlate moderately with chromatin opening intensities. Statistical significance was assessed via a gene randomization test. **(D)** RNA-seq MAF-responsive genes (both down- and up- regulated) enrichment in ATAC-seq peaks (distance to transcription start site (TSS) < 50kb). The number of MAF-target genes with an average test statistic (for comparisons MAF vs Mock and MAF E2 vs Mock E2) greater than 1 (in absolute value) in the ATAC-seq data is shown. MAF-targets do not correlate with chromatin opening intensities. Statistical significance was assessed via a gene randomization test.

Acetylation of histone H3 on lysine 27 (H3K27ac) has been shown to mark active enhancers and it is commonly used to map active chromatin regions[365]. Thus, we assessed the overlap of ATAC-seq peaks with previously published chromatin immunoprecipitation followed by sequencing (ChIP-seq) data of H3K27ac[366] to determine if chromatin opening was associated to enhancer regions. Interestingly, ATAC-seq peaks that overlap with H3K27ac present higher coverage in samples with MAF overexpression compared to control samples at promoters, introns and exons, thereby indicating that MAF-induced chromatin opening in these particular regions is stronger when peaks are located at enhancers (Figure 34).

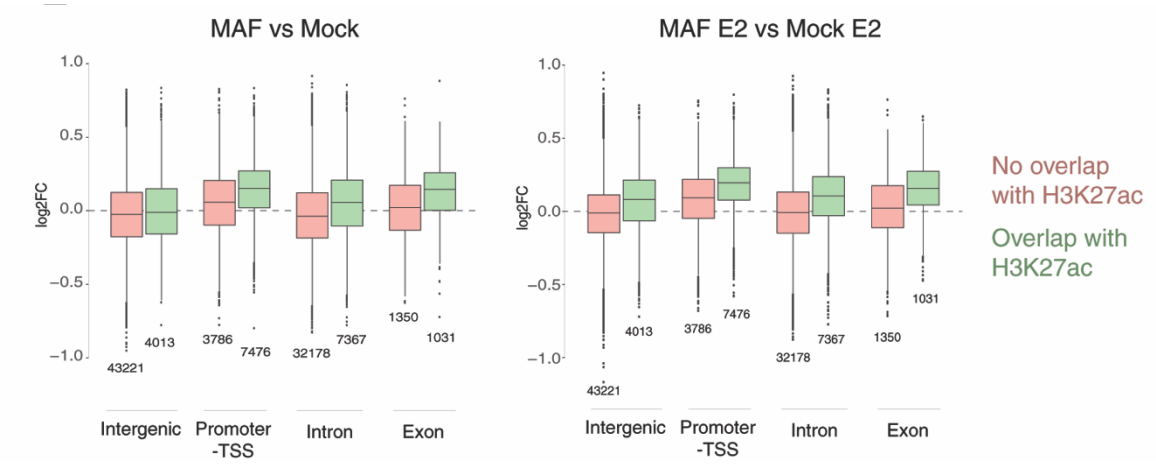


Figure 34. MAF-dependent chromatin opening is more pronounced at enhancer regions. MAF vs Mock (left) and MAF E2 vs Mock E2 (right) log2-fold change for peaks in intergenic, promoter-TSS, intron and exon regions. Peaks overlapping with H3K27ac (published data), a marker of active enhancers, present higher coverage in MAF samples than Mock samples, especially for promoters, introns and exons (in green).

Collectively, these results support a role of MAF in orchestrating chromatin accessibility to allow the recruitment of transcription factors and coregulatory proteins that may lead to the induction of an ER-dependent gene program in response to E2.

MAF potentiates ER recruitment to enhancers

Next, in collaboration with Dr. Luciano Di Croce's group, we investigated whether the presence of MAF was associated with changes in ER binding to chromatin. To address this question, we performed ChIP-seq on ER in control and MAF-overexpressing MCF7 cells

stimulated with E2 for 1 h after being cultured in hormone-deprived media containing charcoal-stripped ^{serum} for 72h.

Genome-wide mapping of ER identified 594 ER binding sites (ChIP-seq peaks) in hormone-deprived MCF7 cells and, after E2 stimulation, the number of ER binding sites rose to 3380 (Figure 35A). Most of the peaks found in hormone-starved cells were conserved after E2 administration (543 common peaks). Notably, there was a significant gain in ChIP-seq signal in these common peaks, thereby indicating that E2 largely increased ER recruitment to these regions. In contrast, only 51peaks were specific of hormone-starved MCF7 cells and disappeared upon E2 administration, which suggests that these peaks may be possible artifacts.

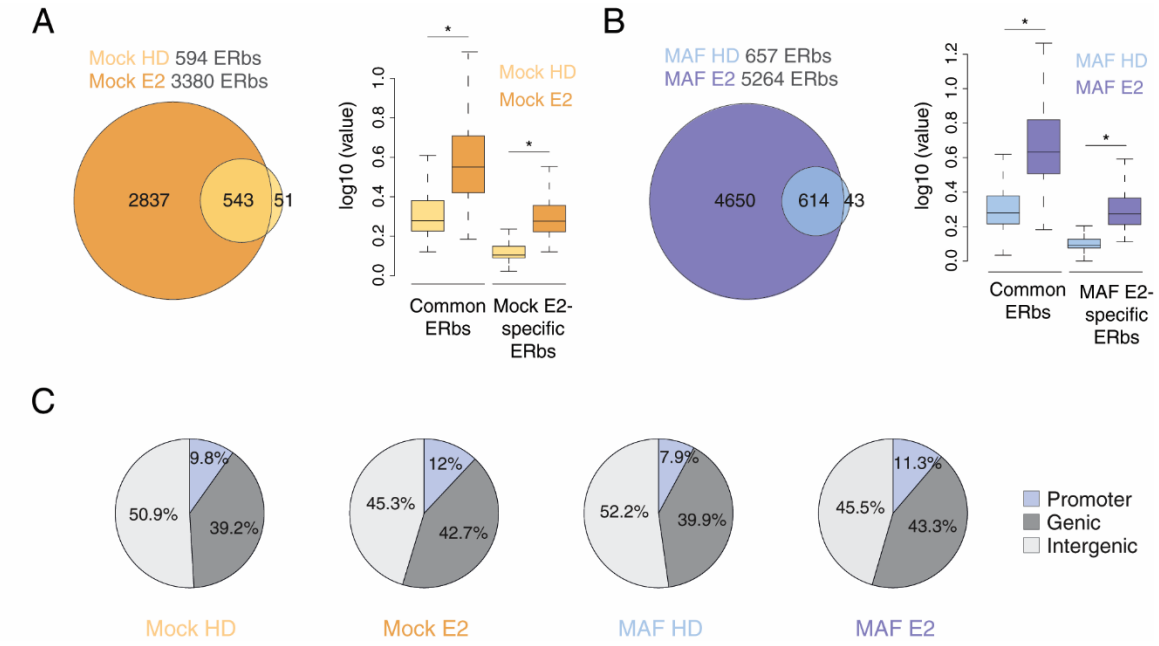


Figure 35. Estrogen receptor (ER) recruitment to chromatin after E2 administration (A) Venn diagrams showing ER binding sites (ChIP-seq peaks) identified in control MCF7 cells before (HD) and after 1 h E2 stimulation (left). Boxplots showing ER ChIP-seq signal intensity for common and new ER binding sites after E2 administration (right). Asterisks show statistical significance (p-value < 10^{-16}, Wilcoxon test, two-sided) **(B)** Venn diagrams showing ER binding sites identified in MAF-overexpressing MCF7 cells before and after 1 h E2 stimulation (left). Boxplots represent ER ChIP-seq signal intensity for common and new ER binding sites after E2 stimulation (right). Asterisks show statistical significance (p-value < 10^{-16}, Wilcoxon test, two-sided) **(C)** Pie charts showing the distribution of ER binding sites in the indicated genomic regions for each condition.

In MAF-overexpressing cells, E2 also triggered a global increase in ER recruitment, with ER binding sites rising from 657 to 5264 (Figure 35B). This large increase was substantially bigger than in control MCF7 cells, which suggests a role for MAF in facilitating ER recruitment to chromatin. Again, most ER binding sites found in hormone-deprived conditions were conserved after E2 stimulation (614 common peaks) and ChIP signal was strongly increased in these regions.

Analysis of peak distribution showed that, in all conditions, a small percentage of all peaks were located in promoters but the majority of ER binding sites appeared in intergenic or genomic regions, thereby suggesting that ER is mainly recruited to enhancer regions (Figure 35C). Consistent with our results, earlier research reported that about 95% ER binding sites are located at enhancer elements located at distant sites form the TSS of regulated genes instead of at promoter proximal regions[367]. Thus, to confirm ER binding at enhancers we assessed overlap of our defined ER binding sites with previously published ChIP-seq data of the histone acetyl transferase p300, a well-established marker of enhancer location[368]. p300 is a transcriptional coactivator that acetylates histone H3 at lysine 27, required for enhancer activity and expression of enhancer-driven transcriptional programs. ChIP-seq data from MCF7 cells treated with vehicle or E2 was downloaded, processed in the same manner as our samples and the number of p300 reads was counted in our sets of ER binding sites found in control and MAF-overexpressing MCF7 cells after E2 stimulation.

Notably, a significant enrichment of p300 was found in the set of 3380 ER peaks found in control cells as well as in the set of 5264 ER peaks identified in MAF-overexpressing cells after E2 administration (Figure 36A). In fact, almost all of the ER binding sites overlapped with p300, whose recruitment was also potentiated to these sites upon E2 stimulation (Figure 36B and C). The strong correlation between p300 and ER ChIP-seq peaks confirmed that the majority of the ER binding sites identified in our control and MAF-overexpressing MCF7 cells were located at enhancer regions.

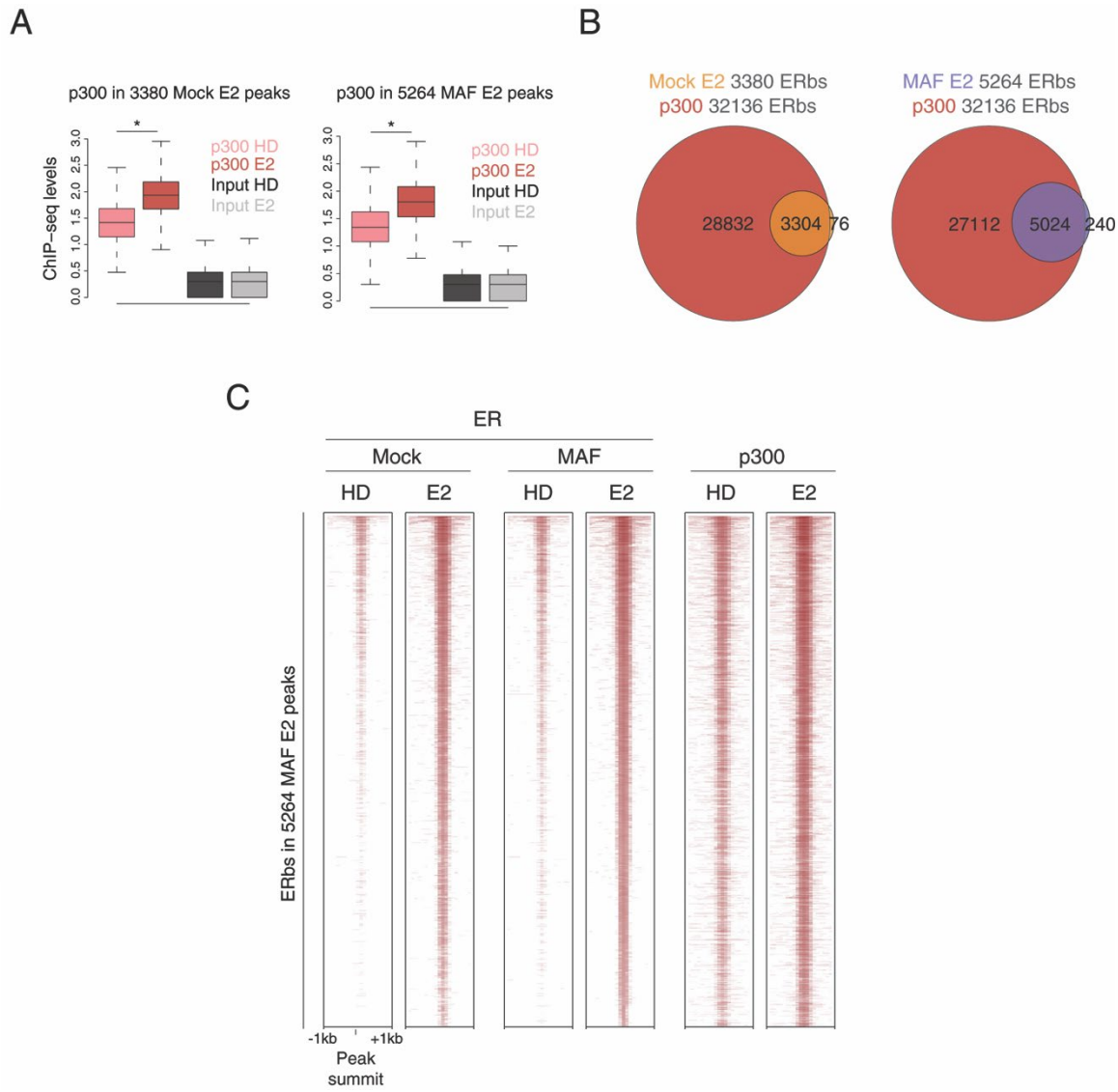


Figure 36. Estrogen receptor (ER) binds to enhancer regions (A) Boxplots showing enrichment of p300 (published data), a marker of active enhancers, in ER binding sites identified in control (left) and MAF-overexpressing MCF7 cells (right) after E2 administration. Asterisks show statistical significance (p-value < 10^{-16}, Wilcoxon test, two-sided) **(B)** Venn diagrams showing the overlapping of p300 and ER binding sites found in control (left) and in MAF-overexpressing MCF7 cells (right) upon E2 stimulation. **(C)** ChIP-seq heatmap showing the distribution of ER reads (in control and MAF-overexpressing MCF7 cells) and p300 on the 5264 ER binding sites found in MAF-overexpressing cells after E2 stimulation (peak summit ±1 kb). Enrichment levels were normalized for the total number of mapped reads of each sample. Peaks were ranked by the intensity of ER signal in MAF-overexpressing cells after E2 stimulation.

We next assessed the effect of MAF in ER recruitment upon E2 administration. Comparison of ER binding sites identified in the presence of E2 in control and in MAF-overexpressing

cells revealed that most ER binding sites (3101 peaks) found in control cells were conserved upon MAF overexpression (Figure 37A). However, 2163 new sites appeared exclusively in cells with MAF overexpression. Strikingly, the presence of MAF not only triggered a large expansion of ER binding sites but it also increased ER recruitment to common sites found in both control and MAF-overexpressing cells, as shown by an increase in ChIP-seq signal in these regions (Figure 37A and B). Considering that MAF was associated with increased chromatin accessibility, these observations strongly support a role for MAF as a pioneer transcription factor that creates accessible chromatin regions to allow ER recruitment.

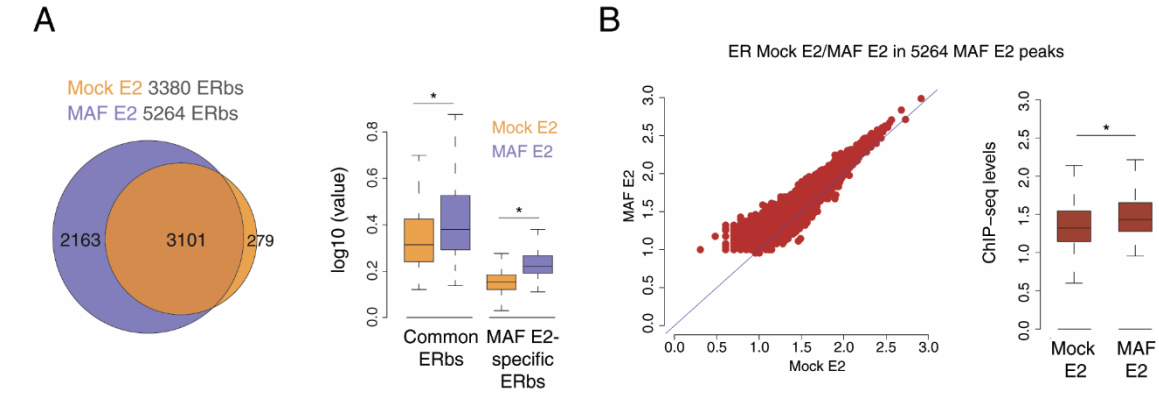


Figure 37. Estrogen receptor (ER) recruitment to chromatin is enriched in the presence of MAF
(A) Venn diagrams showing ER binding sites (ChIP-seq peaks) identified in control and MAF-overexpressing MCF7 cells after 1h E2 stimulation (left). Boxplots represent ER ChIP-seq signal intensity for common ER binding sites in both conditions and specific ER binding sites in MAF-overexpressing cells (right). Asterisks show statistical significance (p-value < 10^{-16}, Wilcoxon test, two-sided) **(B)** Scatter plot showing correlation of ER ChIP-seq reads in control versus MAF-overexpressing MCF7 cells. Each dot in the plot represents the number of ChIP-seq reads normalized by the total number of reads of each sample for each peak discovered in MAF-overexpressing cells after E2 stimulation (5264 peaks). Boxplots show the corresponding distribution values. A significant increase in ER signal is observed in MAF-overexpressing compared to control MCF7 cells (p-value < 10^{-16}, Wilcoxon test, two-sided) (right).

We then explored how changes in ER binding to chromatin correlate with the expression of genes from the MAF-E2 signature. Thus, we interrogated ChIP-seq peaks ± 50 kb from genome wide TSS and analyzed signal intensity in those peaks associated to ER and MAF up-regulated genes identified in RNA-seq experiments (clusters 5 and 6). Genes that were actively transcribed in the presence of E2 and MAF exhibited a significant increase in ER recruitment

in MAF-overexpressing cells after E2 administration (Figure 38). Interestingly, integrative analysis of ChIP-seq and RNA-seq data from ER and MAF down-regulated genes (clusters 7 and 8) also showed a milder increase in ER recruitment in peaks associated to those genes. These results suggest that MAF contributes to create ER-chromatin interactions and these interactions potentiate activating or repressing ER functions.

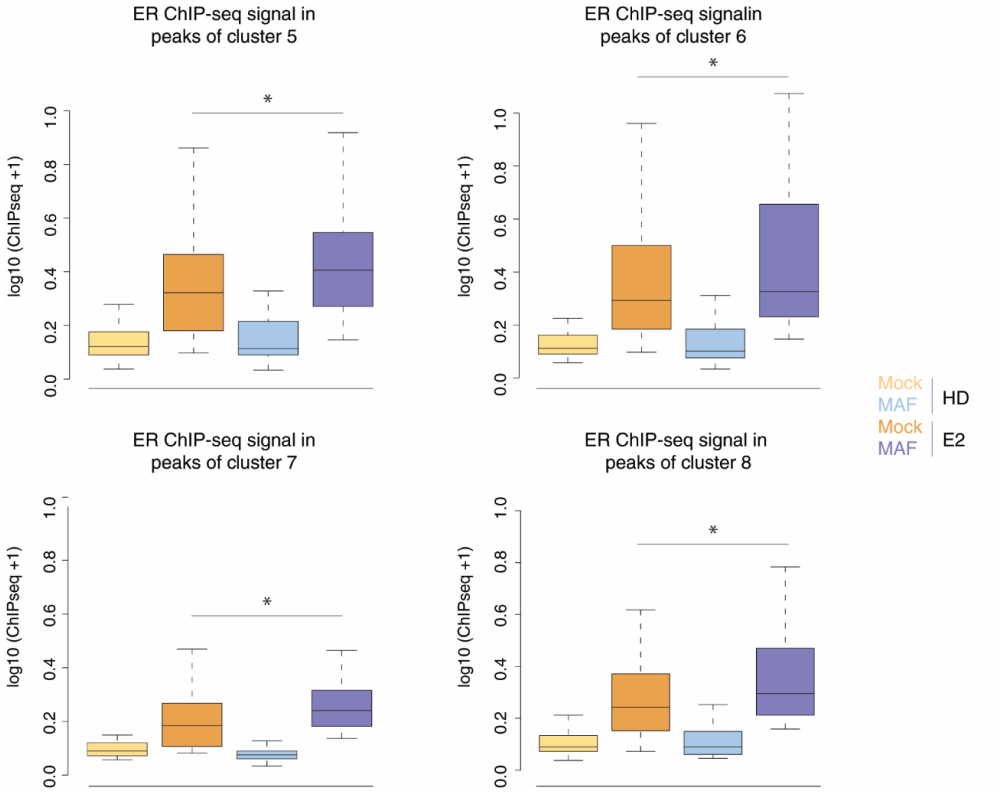
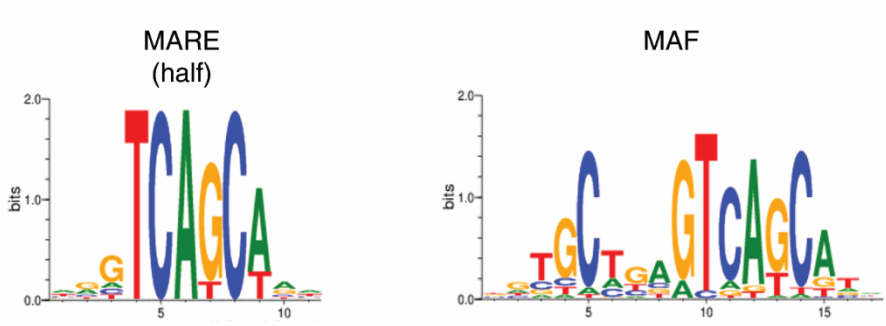


Figure 38. Estrogen receptor (ER) recruitment to chromatin is enriched in peaks associated to ER/MAF-deregulated genes in MAF-overexpressing MCF7 cells after E2 stimulation. (A) ChIP-seq signal intensity for ER binding sites associated to genes that are up-regulated (clusters 5 and 6) or down-regulated (clusters 7 and 8) in the presence of MAF and E2. Distance to transcription start site (TSS) < 50kb. Asterisks show statistical significance (p-value < 10^{-16}, Wilcoxon test, two-sided).

Since ER-DNA interactions in BCa cells were increased in the presence of MAF, we explored whether MAF can assist in tethering ER to the DNA. To address this question, we performed an *in silico* analysis to assess the presence of MAF consensus DNA binding motifs within the regions bound by ER in MAF-overexpressing cells after E2 administration (5261 ER binding sites). Two different position-weight matrices were used for analyzing the presence of MAF binding sites, one corresponding to a half MARE sequence, recognized by all MAF family

proteins, and the other specific for the MAF transcription factor (Figure 39A). Both matrices were obtained from MatBase, the Genomatix software transcription factor database. Indeed, we identified an overrepresentation of MAF binding motifs around the set of ER binding sites found in MAF-overexpressing cells after E2 stimulation (Figure 39B). These results support a role for MAF as an associated protein with the ability to interact with DNA and ER to stabilize the ER complex on the chromatin.

A



B

Matrices	Number of input sequences	Number of matches in input	Expected	Standard Deviation	Overrepresentation	Z-Score
MARE	887	983	554.5	23.55	1.77	18.18
MAF	1343	1621	916.71	30.27	1.77	23.25

Figure 39. MAF binding motifs are enriched within estrogen receptor (ER) binding sites. (A) Position-weight matrices of two described MAF binding motifs obtained from the MatBase database (Genomatix software). MARE, MAF response element (half sites). **(B)** Prediction of MAF binding sites in the vicinity of ER ChIP-seq peaks identified in MAF-overexpressing cells after E2 stimulation generated with the MatInspector program from Genomatix. The number of input sequences with at least one match of the MARE or MAF matrix, the total number of matches in all input sequences, the expected match numbers in an equally sized sample of the genome and the standard deviation, the overrepresentation and the Z-score are shown. The Z-score represents the distance from the population mean in units of the population standard deviation (a Z-score below -2 or above 2 can be considered statistically significant).

ChIP-seq tracks at specific genome regions show that the intensity of ER ChIP-seq peaks in close proximity to *GREB1* is higher upon E2 administration, confirming a successful E2 stimulation (Figure 40). Accordingly, novel targets that were subject to cross-talk between

MAF and E2 signaling, such as *FGF18*, *JAG1*, *PTHLH* and *SOX9*, have proximal ER ChIP-seq peaks.

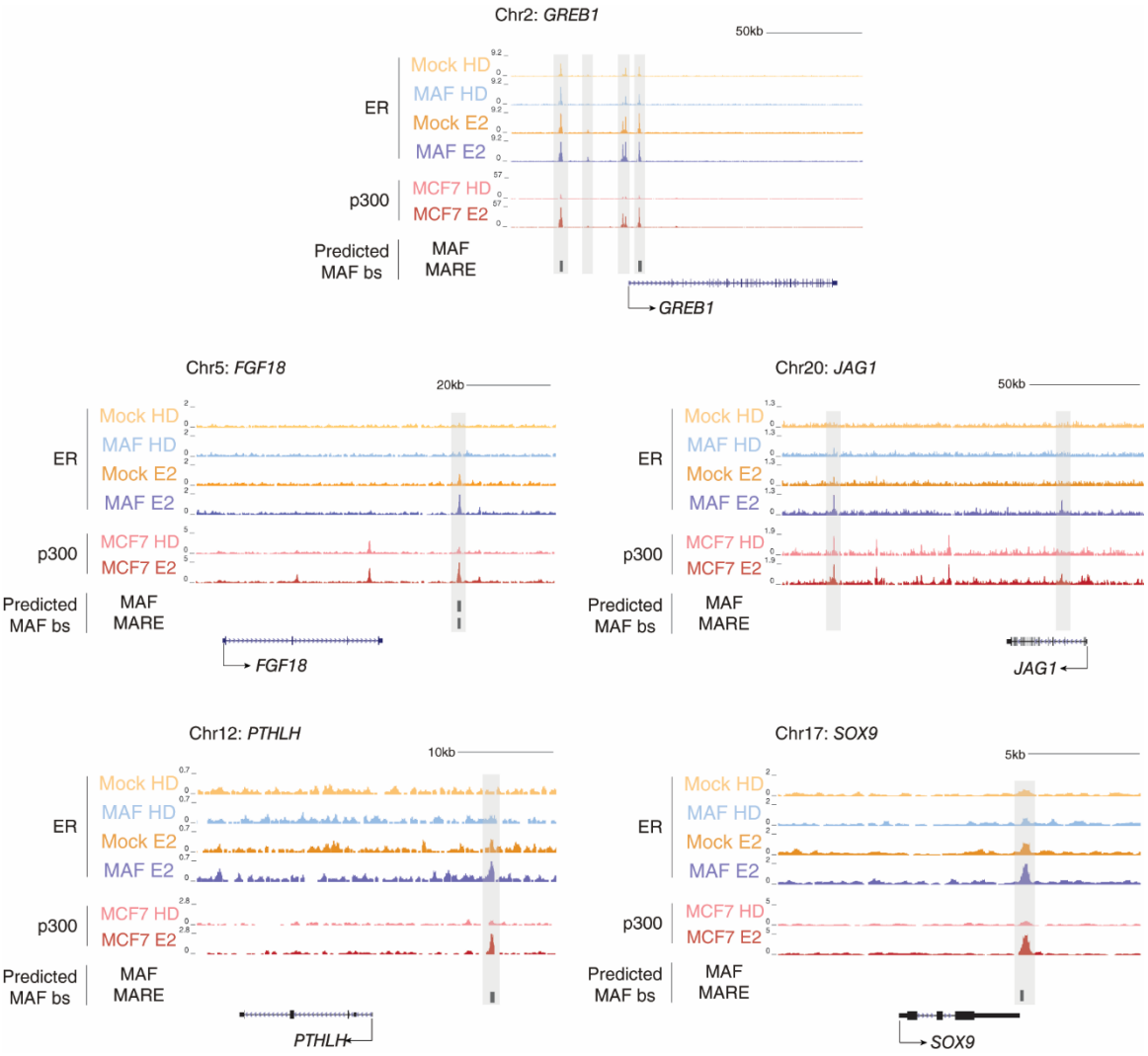


Figure 40. Genome browser screenshots of ChIP-seq profiles at representative target genes. Estrogen receptor (ER) ChIP-seq tracks from control and MAF-overexpressing MCF7 cells before (HD) and after E2 stimulation are shown. p300 ChIP-seq tracks (published data[368]) depict active enhancer regions. Predicted MAF binding sites (using the MAF or MARE matrices) within ER peaks are represented in black.

Although some of the ER ChIP-seq peaks are located at considerable distances from the putative target gene, they all correlate with p300 peaks in MCF7 cells after E2 stimulation. These observations further corroborate that these ER binding sites are located at active enhancer regions that potentially affect target gene expression. Notably, predicted MAF

binding sites are located within the majority of ER ChIP-seq peaks identified in MAF-overexpressing cells after E2 administration associated to the aforementioned MAF and ER targets. Collectively, our results indicate that ER and MAF converge on the same transcriptional regulatory regions and E2 enhances this recruitment, potentially to induce the expression of genes that are known from the literature to confer metastatic properties.

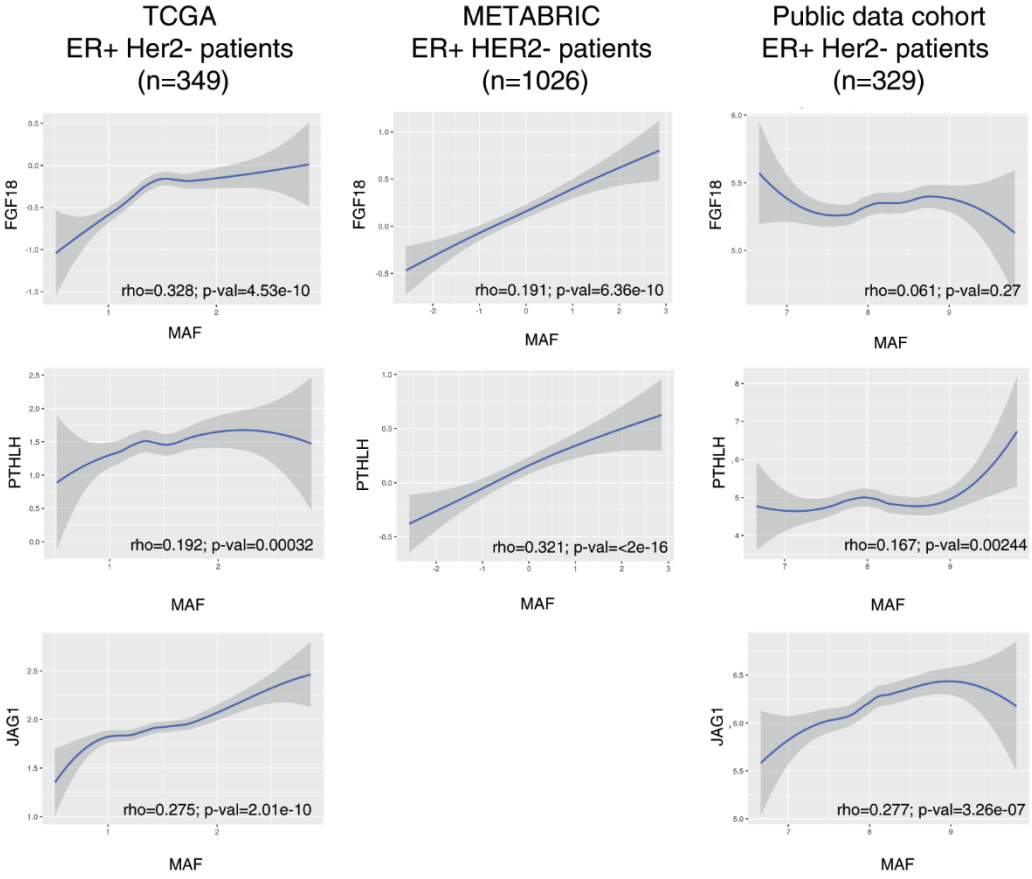


Figure 41. Correlation of MAF expression with representative target genes in ER+ HER2- breast cancer (BCa) patients. Gene expression data was retrieved from a public data cohort containing four micro-array studies (GSE2430, GSE2603, GSE5327 and GSE12276) from The Cancer Genome Atlas (TCGA) and from the Molecular taxonomy of breast cancer international consortium (METABRIC) datasets. The correlation was determined using the Sperman's correlation test.

Next, we validated the association between *MAF* expression and E2 and MAF-coregulated genes in three independent BCa patient sample cohorts. Of note, analysis focused on ER+ HER2- BCa patient samples from The Cancer Genome Atlas (TCGA) cohort dataset revealed a significant correlation between *MAF* and *PTHLH*, *JAG1* and *FGF18* expression (Figure 41). The correlation between *MAF* and both *FGF18* and *PTHLH* was corroborated by analysis of

the Molecular taxonomy of breast cancer international consortium (METABRIC) data from 1026 patients with ER+ HER2- BCa. The METABRIC cohort dataset did not provide *JAG1* expression data and therefore its correlation with *MAF* expression could not be assessed. Finally, analysis of patient sample cohorts from which we had the primary tumor expression profiles (GSE2430, GSE2603, GSE5327 and GSE12276) further confirmed the correlation of *MAF* with *PTHLH* and *JAG1* expression in ER+ HER2- BCa patients. Together, this data show that our findings are not confined to *in vitro* models but are also clinically relevant.

Discussion

DISCUSSION

Optimization of a One-bead-one-compound combinatorial peptide library screening for the identification of high-affinity ligands against metastatic breast cancer cells

After the discovery of the MAF transcription factor as a bone metastasis mediator in ER+ BCa patients, this protein became a very attractive target for the treatment and prevention of bone relapse[303]. However, the lack of enzymatic activity and binding pockets for small molecule design make transcription factors difficult to target[369]. Additionally, their activity depends on association with other nuclear proteins and its structure may be disordered when isolated from their binding partners, which makes even harder the development of chemical inhibitors. Thus, one possibility to target MAF function in BCa cells was to identify proteins under MAF transcriptional control with the potential of being therapeutically targeted. Among MAF-regulated proteins, those located at the plasma membrane represent ideal drug targets, since they play crucial roles such as cell signaling, are involved in interactions with the stroma and are extremely accessible on the cell surface[370].

The OBOC technology allows to rapidly synthesize hundreds of compounds using the "split-and-mix" synthesis method and screen for interactions using living cells[337]. One advantage of this approach is the possibility to introduce non-natural amino acids into the peptide libraries, which are resistant to proteolytic degradation and suitable for *in vivo* applications. Moreover, OBOC screenings do not require prior knowledge of the target molecule, which can be identified later.

A frequent problem when targeting membrane receptors is the necessity of detergents or other reagents that may affect their structure due to the poor solubility of such proteins. Using this methodology, the receptor is displayed in native conditions and peptide hits that would bind the receptor in its natural environment are eliminated.

Considering all the advantages that the OBOC technology offers, our objective was to use a OBOC peptide library screening to discover membrane proteins under MAF transcriptional control that mediate bone metastasis in BCa cells while identifying synthetic peptide binders

to block them. The OBOC combinatorial library screening method had been previously used to optimize the specificity and selectivity of peptide binders for known receptors[340,344,371], but never applied to the cancer metastasis field to identify novel drivers of bone metastasis. For this reason, it was a challenging strategy that entailed a high risk but also a high reward if successful, representing an innovative approach with the potential to have a broad impact on bone metastasis treatment.

Previous work reported successful isolation of high affinity ligands through a COPAS after incubation with OBOC library beads with living cells[342]. COPAS sorting represented a significant improvement in terms of efficacy compared to the labor-intensive and time-consuming manual isolation of positive hits under a microscope. Thus, we attempted to implement this system in our laboratory using integrins and selective integrin-binding peptides, which are molecules typically used in ligand binding assays. Unfortunately, after selecting the best conditions for cell-bead association, we were unable to properly sort positive hits by means of a COPAS. However, we developed a novel and more efficient methodology based on the incubation of 30μm-sized resin beads with fluorescent living cells and subsequent sorting through a FACS instrument.

After technique optimization, two different libraries of 32 and 48 compounds were designed to identify peptide ligands against highly metastatic BCa cells. Large OBOC libraries prepared using the "split and mix" synthesis method result in a variety of peptides consisting on permutations of amino acids with similar or identical masses, which hamper the sequence deconvolution process[351]. Thus, to simplify chromatographic and MS analyses, a multi-objective genetic algorithm-supported approach was used for the design of smaller libraries. Peptide diversity was prioritized rather than library size, thus achieving libraries with maximal number of peptides with maximal mass or sequence diversity[353,352].

Libraries were synthesized by Fmoc SPPS[345] using D-amino acids, which are resistant to proteolysis *in vivo*, and the successful synthesis of all library components was confirmed by MS. When using UPLC-MS technology, peptides are separated according to their hydrophobicity. The detection of very hydrophobic peptides is challenging due to solubility problems, sustained non-covalent interactions with the stationary phase in the liquid chromatography column or insufficient separation prior to MS analysis[353,352]. On the contrary, the detection of very hydrophilic peptides might fail due to insufficient retention on the liquid chromatography column prior to MS analysis. Besides hydrophobicity, polarity and net charge also have an

impact on sequence analysis. In this study, all these properties were considered during the library design to facilitate the detection process. Moreover, a liquid chromatography method was adapted by adjusting the solvent polarity gradient to allow an accurate detection and separation of all peptide components from each library.

Through screening the two OBOC peptide libraries against highly and poorly bone metastatic live populations of BCa cells, six peptide ligands with affinity for highly metastatic BCa cells, defined by MAF overexpression, were identified. Nonetheless, further experiments should be performed to increase the robustness of our observations.

Analysis of the identified peptides revealed that D-Tryptophan (w) and D-Serine (s) at positions 1 and 2, respectively, are crucial for MAF-overexpressing BCa cell binding. Moreover, Glycine (G) and D-Serine (s) in positions 3 and 5, respectively, are preferred for binding to this cell population. On the basis of this information, new highly focused OBOC libraries should be designed incorporating these features to further optimize those peptides that selectively bind highly metastatic BCa cell populations. Focused libraries with the identified crucial amino acids fixed at the appropriate positions while other non-essential positions containing several amino acids could be screened under harsher conditions, such as shorter incubation times, lower bead surface substitution or addition of receptor competitive antagonists to identify derivative peptides with higher affinity and higher specificity in a hit-to-lead selection process.

The interaction of the identified peptide ligands against highly bone metastatic BCa cells should be validated by resynthesis on beads and exposure to both highly and poorly metastatic cell lines. Moreover, to determine whether the efficiency and specificity of the discovered peptides is maintained in an *in vivo* setting, peptides could be conjugated to a fluorescent dye and injected into mouse models to track tumor uptake[372]. Then, the localization of peptide hits could be evaluated in mice bearing MAF-overexpressing xenografts by fluorescence optical imaging. Peptide accumulation in the tumor tissue and demarcation of tumor margins would indicate peptide specificity. If proven specific, fluorescent peptide conjugates would have the potential to be used as probes for *in situ* detection of bone metastatic lesions from BCa patients in the clinics.

Moreover, the most specific and high-affinity ligands for bone metastatic BCa cells could be conjugated to nanoparticles with loading of therapeutic agents for precise-targeting

nanotherapeutics[373]. For instance, previous work showed the successful development of paclitaxel-loaded nanoparticles coupled to peptide ligands against ovarian cancer cells[374]. Paclitaxel-loaded nanoparticles exhibited higher cytotoxicity in *in vitro* studies, higher tumor localization, intracellular uptake, antitumor efficacy and lower systemic toxicity profile *in vivo* when coupled to targeting-peptide ligands compared to nontargeted nanoparticles. Thus, selective peptides against MAF-overexpressing BCa cells have a strong translational potential in tumor-specific chemotherapeutic drug delivery. This class of targeted therapy with reduced systemic toxicity should be delivered after removal of the primary tumor to prevent or treat disseminated bone metastasis.

We identified six peptide hits capable of binding to MAF-overexpressing BCa cell populations. However, the identity of the receptors that recognize these peptides remains to be elucidated. To this end, the six peptide ligands should be resynthesized and conjugated to specific IP beads. Then, to achieve protein-peptide interactions, IP beads would be incubated with membrane extracts from MAF-overexpressing cells and the receptors with high affinity for the discovered peptide ligands would be precipitated, washed and analyzed by MS. Results should be confirmed by Western Blot or IP analyses using specific antibodies against the newly identified membrane protein. Moreover, these specific antibodies could be used in competitive binding assays to test for peptide binder specificity.

Once MAF-transcriptionally controlled membrane proteins are identified, their contribution to the development of bone metastasis in BCa patients should be characterized. To this end, candidates should be clinically validated by examination of their expression levels in publicly available BCa datasets and histological sections of BCa primary and bone metastatic samples. Following clinical validation, mechanistic functional analyses of the candidates in different *in vitro* and *in vivo* BCa models should be conducted. Gain- and loss-of-function experiments may allow to unravel relevant cell-autonomous traits as well as cancer cell-stroma interactions to expand our knowledge on MAF-mediated BCa metastasis to bone.

It is worth to mention that, although we could identify specific ligands for highly bone metastatic cell populations, our approach has some limitations that make high-throughput cell-based screenings impractical.

Firstly, to identify peptides differentially expressed between highly and poorly bone metastatic BCa cells, we fluorescently labeled both cell populations with RFP and screened OBOC

libraries against each of them. Then, by comparing the peptide hits identified in both populations, we selected those that were specific for the highly metastatic. Previous work showed the utility of an On-Bead-Two-Color (OBTC) cell screen for high throughput screening to identify peptides specific for lung cancer cells[375]. This technique is a modification of the OBOC method, where two cell populations are labeled with two different fluorophores, mixed in a 1:1 ratio and incubated with peptide library beads. After incubation, only beads coated with one color cells are manually picked under a fluorescence microscope. This approach is optimal for the selection of peptides that recognize cells that express a target receptor while ignoring almost identical cells that do not express this receptor and was used to successfully identify specific ligands for lung cancer cells compared to normal epithelial cells from the same patient. We believe that this strategy could be implemented in our technology by labeling highly and poorly bone metastatic BCa cells with a different fluorophore. Then, the screening of OBOC libraries based on 30μm-sized resin beads could be performed against both cell populations simultaneously and the FACS system would be used to sort only those beads associated to highly metastatic BCa cells. Thus, labeling our two cell populations of interest with a different fluorophore would save significant time in the sorting process and make our technology more efficient.

Secondly, peptide mass detection and sequence deconvolution were extremely challenging. Previous work had reported efficient deconvolution of hit peptides using MS after sorting the cell-bound beads through the COPAS[342]. However, the authors used a small library of 7 peptides that were tested independently, not as a peptide mixture, which considerably facilitated sequence determination by MS. In our study, 32 and 48 compounds from two OBOC libraries were tested simultaneously, thereby increasing sample complexity. Only the 32-component library showed significant binding, and the MS spectra of library peptides to identify the positive hits was performed by manual isolation and fragmentation of each component. Since it was not possible to automatize the process of isolating and fragmenting peptide hits, mass detection and sequence deconvolution became tedious and time-consuming.

Peptides cleaved from the beads after library synthesis could be successfully identified using the optimized UPLC-MS method. However, cell-bead incubation and subsequent sorting interfered with accurate sequence determination using MS due to a significant increase in background noise. We hypothesize that the buffers used for sorting and for removal of bound cells from isolated beads may generate impurities that hamper sequence deconvolution. Thus,

to prepare the hit peptides for MS analysis, we attempted to either boil the sorted beads, use a highly concentrated NaCl solution to strip off the cells before cleaving the peptide from the beads or directly sort the beads into tubes with nitrocellulose membranes to cleave the peptides without using buffers to remove bound cells. Moreover, several washing steps were performed after removal of bound cells. However, we obtained poor sequencing results due to high background noise. We believe that a systematical exploration of alternative cell dissociation buffer treatments, such as 8 M guanidine hydrochloride, which has been previously used to remove binding cells from OBOC libraries[340], may lead to a reduction of the background noise in MS analyses, thereby facilitating sequence deconvolution of the sorted hits.

Finally, we observed that the number of sorted events correlated with success in sequence deconvolution. This information indicates that the amount of each positive hit needs to reach a threshold to stand out against the background noise for successful sequence identification, which would substantially extend the sorting time when using bigger libraries. Together, these observations make our approach not feasible for higher throughput library screenings and highlight the need of further technique optimization.

In conclusion, we report a novel OBOC technology based on the synthesis of peptides on 30μm-sized resin beads to facilitate automated and efficient sorting through a FACS instrument after incubation with living cells. For the first time, coupling of this technique to MS analysis has allowed the identification of specific peptides against MAF-overexpressing BCa cells with the potential to become therapeutic agents for the treatment and prevention of bone metastasis. Our data highlights the potential of using this methodology to circumvent the amount of time and resources used by conventional drug lead discovery approaches, where knowing the identity of the targeted protein is a prerequisite.

Our methodology combines cell-based assays with automated sorting, which is essential for high-throughput screening of combinatorial libraries. However, further optimization of the peptide sequence deconvolution process is indispensable for the identification of positive hits from big OBOC libraries.

Characterization of the MAF interactome by proximity-dependent biotin identification (BioID)

Cofactor interactions specify the transcriptional activity of transcription factors. Previous work has demonstrated that MAF activity is dependent on cell type and interaction with binding partners that modulate its function. For instance, the synergism between MAF and the nuclear factor for activated T cells (NFAT1) control IL-4 expression in T-cells[376]. In contrast, interaction of MAF with c-Myb results in repression of Aminopeptidase N (CD13/APN) expression during hematopoietic cell development[377]. These observations indicate that MAF can act both as transcriptional activator and repressor and that the association of MAF with specific partners has an important impact in gene expression regulation.

A different strategy to target the activity of transcription factors is to interfere with the function of their cofactors. In this study, we performed an unbiased proteomic screen using BioID to define the MAF interaction landscape in BCa cells. The characterization of the MAF interactome provides insight into the binding partners that modulate MAF activity in this particular context and offers the possibility to increase available therapeutic options for this difficult-to-drug oncogene.

BioID has emerged as a powerful tool for the accurate identification of protein-protein interactions in their natural environment. This technique is based on the fusion of a protein of interest to a mutated biotin ligase that, upon biotin addition, releases reactive biotinyl-AMP from biotin and ATP, which covalently interacts with adjacent amines on proximal proteins[378]. Since biotinylation is a rare modification and biotin can be selectively isolated by streptavidin affinity purification, BioID allows the identification of proteins in close proximity to the bait protein via MS. The main advantage of this method is that it allows for the detection of low affinity and transient interactions in a cellular environment of choice.

In our experiments, we used a second-generation biotin ligase named BioID2, which is 30% smaller than the original BirA R118G*, allows a more-selective targeting of fusion proteins and shows enhanced labeling of proximal proteins[379]. Other methods had been recently developed to upgrade this technology. For instance, BASU, another promiscuous biotin ligase, was recently engineered from *Bacillus subtilis*[380]. This enzyme has higher activity and faster kinetics than BioID, and has allowed the direct study of RNA-protein interaction in living

cells. Moreover, ancestral BirA for proximity-dependent biotin identification (AirID) has been recently developed using an ancestral enzyme reconstruction algorithm[381]. AirID and BioID show high sequence similarity but the former has higher biotinylation activity than the latter. BioID, BASU and AirID require 18-24 h of incubation with biotin for a proper labeling. In contrast, other biotin ligases such as TurboID and miniTurbo have faster labeling kinetics and biotin can be added 10 min or less in cell culture, which allows to detect dynamic biological processes with higher temporal resolution[382]. TurboID has higher biotin affinity than miniTurbo, which, in some cases can cause cellular toxicity[383]. Thus, it could be worth to consider these novel tools for future protein-protein interaction studies.

Our work identified unique and shared interacting partners for MAF S and L isoforms in MCF7 cells. Since both MAF S and L were reported to equally promote bone metastasis in BCa cells[303], we focused on the shared interactors for both isoforms, defining a network of 126 high-confidence MAF proximity interactors mainly involved in transcriptional regulation and chromatin modification. However, our work offers a deep characterization of the MAF interactome and provides a list of isoform-unique binding partners for future studies to expand insight into functional differences of each isoform, which are still unclear.

Previous clinical data showed that bisphosphonates treatment in premenopausal patients with MAF-positive tumors is associated with increased extra-skeletal recurrence, adverse outcomes and poorer overall survival[334]. The association of menopausal status with worse outcome in these patients supports a role for E2, the main ER ligand, in contributing to MAF-positive tumor aggressiveness. Thus, among the 126 high-confidence MAF interactors discovered by BioID proximity labeling in MCF7 cells, ER captured our full attention.

The robustness of the MAF-ER interaction was supported by co-IP and PLA experiments. Moreover, we attempted to map the region of MAF that interacts with ER by eliminating critical protein domains prior to PLA. Our data strongly suggest that the transactivation domain is the point of MAF-ER interaction. Notwithstanding, this approach had some limitations that should be addressed. Deletion of the MAF C-terminal region (MAF ΔC-t), which contains the DNA-binding domain, resulted in nuclear and cytoplasmic MAF localization, thereby suggesting a nuclear retention signal within this domain. Although MAF (ΔC-t) truncated form retained the ability to interact with ER, a nuclear retention signal should be added for a more accurate quantification of PLA signal. Moreover, PLA experiments to map the domain required for interaction with ER used transiently expressed full-length and

truncated MAF plasmids that led to high protein expression levels. Since the PLA approach detects proteins in close proximity, namely direct interactors, indirect interactors or vicinal proteins that do not physically interact with the protein of interest, excessive levels of protein expression can lead to an increase in non-specific PLA signal. To address this issue, truncated MAF sequences should be cloned into viral plasmids for the generation of stable cells lines that exhibit lower protein expression levels. Thus, performing these experiments in cells whose MAF expression is closer to physiological levels would reduce background signal and result in more reliable data.

Since ER has a pivotal role in the onset and progression of BCa and MAF drives bone metastasis in ER+ BCa patients[303], we prioritized the identification of the MAF interaction landscape in MCF7 cells, an ER+ BCa model. Nonetheless, we also explored the MAF interactome in MDA-MB-231 cells, a model for ER- BCa. In this study, we show that there are common proximal proteins identified in both models. However, a large number of interactions found in MCF7 cells are lost in MDA-MB-231 cells, which indicates that part of the MAF interactome is cell line-specific and that the presence of ER may modulate MAF interaction with specific partners. Although we only analyzed the presence of representative components of the major chromatin remodeling complexes found to interact with MAF in MCF7 cells, our work provides a resource of MAF-proximal proteins in MDA-MB-231 cells to explore cell line-specific binding partners that might provide insight into MAF function in an ER- setting.

The molecular interplay of MAF and ER transcription factors in breast cancer cells

The classical ER mechanism of action involves dimerization upon E2 binding, translocation to the nucleus and binding to EREs to regulate the expression of specific genes[30]. However, ER can be recruited to the genome at non-canonical sites through being tethered to the DNA by other transcription factors. For instance, FOXA1, GATA3, PBX1 and AP2γ can assist in tethering ER to the DNA[384]. Moreover, ER binding at assisted-loading sites has been shown to be dependent on AP-1 transcription factors[385,358].

ER has been reported to cooperate with a large collection of cofactors for gene expression regulation[386]. These cofactors associate with the ER complex either to recruit other cofactors, regulate ER protein structure or directly modify chromatin accessibility. Thus, alteration of the expression levels of key ER-associated factors might lead to altered ER transcriptional activity.

In this study, we demonstrate a novel functional association of MAF, a member of the AP-1 family of transcription factors, and ER. Our findings reveal that E2-mediated effects are influenced by MAF overexpression in BCa cells and suggest a model in which MAF recruits chromatin remodelers to increase chromatin accessibility, thus allowing ER recruitment and modulating the expression of MAF-ER coregulated genes.

We report a panel of genes regulated by both MAF and ER transcription factors, some of them known to confer metastatic properties in BCa, among other cancer types. For instance, FGF18 overexpression has been associated with the progression of gastric, ovarian, colorectal and breast cancers by inducing EMT and by promoting neoangiogenesis, thereby modulating the tumor microenvironment[362,387,388]. Moreover, FGF18 enrichment predicts poor survival in gastric and ovarian cancers[389,390]. High levels of JAG1 expression in BCa cells are associated with bone metastasis development through activation of the Notch pathway in osteoblasts, which triggers IL-6 production and subsequent activation of osteoclasts, thus aggravating osteolytic lesions[249]. SOX9, a stem-cell related transcription factor, has been related with the acquisition of metastasis-seeding abilities in cancer cells and is associated with poor survival in BCa patients[391,363]. Finally, PRHrP, encoded by the *PTHLH* gene, promotes osteoclast formation and has a key role in osteolytic bone metastasis[224]. *PTHLH* had previously been reported to be under MAF transcriptional control[303]. Our results expand upon this knowledge and indicate that, indeed, PRHrP expression is regulated by MAF, but is also potentiated by E2 signaling. Additionally, MAF and E2 also cooperate to down-regulate genes such as kynureninase (KYNU) and B-cell translocation gene 1 (BTG1), which have been reported to act as tumor suppressors in BCa[392,393].

The robustness of the association of MAF expression with some the MAF-ER coregulated genes (*FGF18*, *JAG1* and *PTHLH*) is supported by a validation in 3 different BCa cell lines as well as in patient samples from 3 different ER+ BCa datasets, demonstrating the clinical relevance of our findings.

Future studies for the identification of the critical targets by MAF and ER cooperation might provide new targets for therapeutic intervention. Of note, preclinical data proved that neutralizing antibodies against JAG1 show an excellent safety profile and are associated with a significant reduction of bone metastasis burden[252]. Moreover, it has been shown that PRHrP-targeting antibodies inhibit osteolysis caused by metastatic BCa cells and decreases tumor burden in bone[224]. Thus, it is tempting to think that JAG1 or PTHrP antibodies in combination with hormonal therapy could impair the growth of MAF-overexpressing BCa cells.

Regarding the molecular mechanisms of the MAF-ER crosstalk, we provide evidence that MAF facilitates the selective access of ER to specific genomic sites by maintaining an accessible chromatin configuration. Our results are consistent with previous work showing interactions between AP-1 family members with nuclear receptors. For instance, this exact cooperative mechanism has been described between AP-1 and the glucocorticoid receptor (GR)[394]. AP-1 occupancy maintains chromatin accessibility and is a pre-requisite for the attraction of GR to specific regulatory regions. We demonstrate that this mechanism of chromatin priming by AP-1 transcription factors is extended to MAF and ER to mediate the regulation of a specific transcriptional program.

ER typically binds enhancer elements that can be at considerable distances from target genes[367]. We have shown a genome-wide interaction between MAF and ER that occur mainly at regulatory elements located in accessible chromatin regions and is enhanced after E2 stimulation. Moreover, through *in silico* analyses, we identified overrepresentation of MAF consensus DNA binding motifs in ER binding sites, which implies a functional connection at regulatory elements occupied by ER. However, previous work has reported MAF binding sites that considerably differ to consensus sequences and show that MAF can also interact with the DNA via other DNA-binding proteins[315]. These observations highlight the complexity of predicting MAF recognition sites. Thus, investigation of genome-wide MAF binding profiles by ChIP-seq and assessment of the overlap with the ER binding sites observed in MAF-overexpressing cells would allow a more accurate analysis of the genomic regions co-occupied by both MAF and ER transcription factors.

Our findings reveal a prerequisite for chromatin opening by one transcription factor for the subsequent recruitment of other regulatory factors. MAF orchestrates the ER binding profile and therefore the response program mediated by ER upon E2 signaling. Nevertheless, some additional mechanistic studies could contribute to fully dissect the mechanism of MAF and

ER cooperation. Further investigations should attempt to analyze ER binding profiles in cells with overexpression of a MAF truncated form with no DNA binding domain (MAF ΔC-t) to determine whether interference with the binding activity of the transcription factor impedes chromatin opening, thus reducing ER binding. Moreover, to determine whether MAF can assist in tethering ER to the DNA, ChIP-seq analysis could be performed with an ER mutant without DNA binding domain to assess whether ER binding sites described in MAF-overexpressing cells are abolished.

FOXA1 is a well-known determinant of ER function and has been reported to act as a pioneer transcription factor that can directly bind to compacted chromatin without the requirement of additional proteins to facilitate the interaction between ER and the DNA[395]. Similarly, our data points at MAF as a pioneer factor that binds to specific chromatin sites and recruits chromatin remodeling proteins, thereby creating access for ER within those specific regions. We show that chromatin accessibility is broadly increased in the presence of MAF. Thus, ER occupies a small fraction of its DNA binding sites due to restrictive effects of chromatin architecture and MAF-dependent chromatin opening as well as E2 signaling dictates ER binding at specific sites. MAF contributes to create ER-chromatin interactions that potentiate ER activating or repressing functions depending on the recruitment of additional coregulatory proteins.

Targeting the pioneer factor instead of ER represents an opportunity to block ER transcriptional activity. However, transcription factors are difficult therapeutic targets and the identification of regulatory enzymes that influence MAF function may be a more feasible option.

The observation that MAF maintains an accessible chromatin environment indicates that MAF requires additional partners to facilitate interactions between ER and the DNA. Chromatin-remodeling proteins are critical regulators of the activity of transcription factors, and may represent an unexploited option to disrupt oncogenic MAF function. Consistently, our work offers a variety of chromatin remodelers captured by BioID proximity labeling, some of which are involved in cancer and targetable with small-molecule inhibitors. Thus, the exploration of additional nodes in the MAF-proximal protein network may represent an option for novel therapeutic opportunities for MAF-overexpressing tumors.

KDM1A is a chromatin remodeling enzyme that demethylates histone 3 at lysine 4 (H3K4) or lysine 9 (H3K9), acting as a corepressor or a coactivator, depending on the context. KDM1A

overexpression has been found in several types of cancer and increased KDM1A activity is associated with increased metastatic potential[396]. Interestingly, KDM1A has been reported as a critical regulator of ER transcriptional activity. This chromatin-remodeling protein is recruited to a significant fraction of ER target genes to demethylate H3K9, a specific mark for transcriptional repression, and allow ER-mediated transcription[397]. Moreover, it has been shown that inhibition of KDM1A demethylation impairs ER recruitment to regulatory regions of E2-responsive genes, compromising ER transcriptional activity, and has an impact on BCa cell proliferation[398]. Hence, among all chromatin remodelers found to interact with MAF, we prioritize the study of KDM1A for its ability to regulate ER function.

On the basis of these observations, it is tempting to speculate that MAF recruits KDM1A to orchestrate the chromatin accessibility changes that result in a modulation of ER recruitment to chromatin. Thus, we believe that exploring the role of KDM1A in MAF-overexpressing BCa cells may broaden our knowledge on MAF-dependent modulation of the E2 response as well as our understanding of MAF oncogenic activity.

Work is currently underway in our laboratory to evaluate the role of KDM1A on MAF-dependent modulation of the E2 response. Preliminary data show that KDM1A is required for the expression of part of the MAF and ER coregulated genes and that KDM1A depletion or inhibition with ORY-1001, a selective covalent inhibitor, reduces growth of BCa cells in the bone. Interestingly, a phase I clinical trial has shown a good safety profile for ORY-1001 in AML patients and a phase II trial is ongoing[399]. Nevertheless, further experiments to understand the contribution of KDM1A in the increase of ER occupancy in a MAF-overexpressing context are still required. The chromatin landscape as well as ER binding patterns in MAF-overexpressing cells still need to be interrogated in the absence of KDM1A. However, we hypothesize that disrupting the MAF-KDM1A transcriptional activation axis may be a compelling strategy to target MAF-overexpressing tumors.

In summary, in this study we show that MAF modulates the chromatin landscape for ER access and stabilizes the ER complex on the chromatin. These interactions of MAF and ER are functionally significant in BCa cells, which reaffirms the importance of considering MAF status to guide the treatment of BCa patients. Our data emphasizes the complexity of gene expression regulation in the ER pathways but provides tools for exploring novel ways of targeting BCa, particularly in a MAF-overexpressing context.

Conclusions

CONCLUSIONS

The MAF interactome in BCa cells includes components of different chromatin-remodeling complexes such as SWI/SNF, INO80, NurD and CoReEST.

MAF interacts with ER in MCF7 cells.

MAF expression modulates the E2 response in MCF7 cells

Genes involved in metastatic processes, such as *FGF18, JAG1* and *PTHLH* are subject to cross-talk between MAF and E2 signaling.

MAF expression is associated with a global increase in chromatin accessibility.

MAF increases ER recruitment to enhancers.

ER recruitment is increased in DNA regions proximal to MAF and ER-deregulated genes.

MAF binding motifs are enriched within ER binding sites in MAF-overexpressing MCF7 cells.

MAF expression correlates with the expression of MAF and E2-dependent genes in patient sample cohorts of ER+ HER2- BCa.

Materials and Methods

METHODS

Cell culture

MCF7 and MDA-MB-231 human breast cancer cell lines were purchased from the American Type Culture Collection (ATCC) and the BoM2 bone metastatic cell line was derived from MCF7 by three rounds of mouse intracardiac injection and subsequent isolation after bone metastasis establishment.

All cell lines were maintained in standard conditions (37ºC, 5% CO_2) in Dulbecco's Modified Eagle Medium (DMEM) (Gibco) supplemented with 10% fetal bovine serum (FBS), glutamine 0.29 mg/mL, penicillin 100 units/mL and streptomycin 0.1 mg/mL (all supplements from Biological Industries). Cells were split every 2 to 3 days and routinely tested to be free of mycoplasma infection.

In experiments to evaluate the effect of E2 stimulation, MCF7 cells were maintained in phenol-red free DMEM (Gibco) supplemented with 5% charcoal-stripped FBS (Capricorn Scientific) for 72h before E2 (10nM, Sigma-Aldrich) or ethanol (vehicle) addition.

Generation of MAF overexpressing cells

For stable MAF overexpression, HA-tagged MAF (short and long isoforms) was PCR amplified using primers containing HpaI restriction sites and cloned into the retroviral plasmid MSCV-neo (See tables 6 and 7). To produce retroviruses, human embryonic kidney 293-T cells (HEK 293T) were seeded at 80% confluence in 150 cm^2 plates and transfected 16 h later with 12 μg of MSCV-Empty or MSCV-MAF, 1.2 μg of VSVG-R, 10.8 μg of GAG-POL plasmids using polyethylenimine (PEI) in a 150 mM sodium chloride (NaCl) solution. The viral supernatant was collected 72 h after transfection, passed through a 0.45 μm cellulose acetate filter (Whatman) and used to infect MCF7 and MDA-MB-231 cells. Specifically, $3x10^5$ cells were seeded into a 6-well plate followed by the addition of viral medium with 8 μg/μL polybrene (Sigma-Aldrich), centrifugation at 2250 rpm for 45 min and overnight (O/N) incubation. Viruses were removed and medium was added for cell recovery. Selection was

performed 24 h after recovery with 1 mg/mL neomycin (G418) (Santa Cruz Biotechnology) for 10 days.

To produce red fluorescent protein (RFP)-containing lentiviruses, HEK293-T cells were seeded at 80% confluence in 150 cm^2 plates and transfected 16 h later with 6 μg of FUW-ubiquitin-SV40-RFP (Addgene plasmid 65448), 6 μg of RRE, 6 μg of RSV and 6 μg VSVG using polyethylenimine (PEI) in a 150 mM NaCl solution. Virus collection and MCF7 and BoM2 infection were performed as explained above. For selection, cells were sorted for RFP-expression.

For expression of full-length and truncated MAF proteins, short, long and truncated MAF sequences were PCR amplified using primers with EcoRI and BamHI restriction sites incorporated and cloned downstream of an N-terminal 2xHA tag of a pCMV5 plasmid (See tables 6 and 7). Truncated MAF sequences included MAF (ΔC-t), encoding MAF amino acids 1 to 283 and lacking the DNA-binding domain, MAF L (ΔN-t 1), encoding MAF amino acids 85 to 404 and lacking part of the transactivation domain, and MAF L (ΔN-t 2), encoding MAF amino acids 120 to 404 and lacking the entire transactivation domain. MCF7 cells were transfected with the generated plasmids using GenJet DNA transfection reagent (SignaGen Laboratories) according to the manufacturer's protocol.

Flow cytometry

For integrin expression analysis, single cell suspensions of MCF7 breast cancer cells were stained with fluorescent conjugated antibodies (See table 8) diluted in PBS for 30 min on ice. Data was analyzed by flow cytometry using a FACSAria (Becton Dickinson) cell sorter with FlowJo software.

Library design based on a Non-dominated Solutions Genetic Algorithm (NSGA-II)

Library design was defined by the number of positions and the number of each amino acid in each position. The selection of each amino acid was based on their physicochemical properties

and a NSGA-II was used to optimize peptide library design. The genetic algorithm transformed each set of possible amino acids into a bit-string, where each bit represented inclusion (1) or exclusion (0) of each amino acid, and searched for the optimal subset of amino acids through defined fitness functions based on total number of peptides for a specific variability and number of peptides with unique mass.

The parameters of NSGA-II were set as follows: individual: single library, population size: 500, representation: bit-string, generations: 100 multiplied by bit-string length or 50 generations without an average relative change in the best fitness function value, crossover rate: 80%, crossover function: scattered, mutation rate: 1%, mutation function: bitflip, selection function: stochastic uniform, elitism: 5%, Pareto fraction: 20%, distance measure function: distance crowding, fitness functions: library size and percentage of peptides without mass overlap.

Solid-phase peptide synthesis

Fluorenylmethoxycarbonyl (Fmoc)-(D)-amino acids (Iris Biotech GmbH) were used to build peptides up on TentaGel M NH2 Monosized Amino TentaGel Microspheres (Rapp Polymere, M30352. Loading: 0.24 mmol/g) preloaded with a 4-hydroxymethylbenzoic acid (HMBA) linker. All reactions were performed manually in polyptopylene syringes with a polyethylene porous disk. TentaGel beads were swollen in water for 24 h. The first amino acid was coupled to the HMBA linker from the beads through ester bond formation in dichloromethane (DCM) (Scharlau) using a four-fold excess of amino acid, 4 equivalents of *N,N′*-Diisopropylcarbodiimide (DIC) (Novabiochem) and 0.4 equivalents of 4-(Dimethylamino)pyridine (DMAP) for 30 min. This reaction was repeated three times and was followed by a washing step to remove byproducts and excess reagents.

After coupling the first amino acid, a small portion was separated to calculate the loading capacity of the resin. Fmoc was removed in 5 mL 20% v/v piperidine/ *N,N*-dimethylformamide (DMF) (Carlo Erba reagents), the supernatant was collected and Fmoc absorbance was measured at 301 nm using UV spectroscopy. Loading was calculated using the equation: $X = A \cdot V \cdot \varepsilon \cdot m \cdot b$, where A stands for Fmoc absorbance, V stands for the solvent volume, ε stands for the molar extinction coefficient of Fmoc at 301 nm (7800 M^{-1} cm^{-1}), m stands for the mass of the resin in g and b stands for the loading in mmol/g.

After loading determination, peptide chain was elongated through peptide bond formation in DMF using three-fold excess of F-moc protected amino acid. 2-(1H-Benzotriazole-1-yl)-1,1,3,3-tetramethylaminium tetrafluoroborate (TBTU) (Novabiochem) was used to activate the carboxyl group of the incoming amino acid and *N,N*-Diisopropylethylamine (DIPEA) (Sigma-Aldrich) was used as a coupling reagent in a 1:2 ratio relative to the amino acid. Following amino acid coupling, washes and Fmoc deprotection with 20% (v/v) piperidine (Scharlau) solution in DMF were performed to allow the reaction with the next amino-acid. Side chains were cleaved by a 95% Trifluoroacetic acid (TFA) (Sigma-Aldrich), 2.5% Triisopropylsilane (TIS) (Sigma-Aldrich), and 2.5% water solution. Coupling completion and Fmoc deprotection were monitored by Kaiser colorimetric test assay by detection of primary amines. The coupling-washing-deprotection cycle was repeated until the desired peptide chain was formed and final washes with water, DMF, DCM and diethyl ether were carried out. After synthesis, peptides were cleaved from the resin by means of ammonia vapors to obtain C-terminal amidated peptides and dissolved in 20% (v/v) acetonitrile (Scharlau)/water. Beads were removed by centrifugation at 13,000 rpm for 30 min and peptide sequence was verified by UPLC-MS, high-resolution mass spectrometry and nano LC-MS/MS to confirm the detection of all the library components.

Peptide libraries were synthesized by SPPS using the "split and mix" method. Peptides from Library 1 followed the sequence x1-x2-x3-x4-x5-y, where the variable positions were occupied by *G, w* in x1, *s, e* in x2, *G, r* in x3, *a, r* in x4 and *s, a* in x5. Peptides from Library 2 followed the sequence x1-x2-p-x4-x5-x6-y, where the variable positions were occupied by *a, w* in x1, *s, r* in x2, *a, e* in x4, *i, e, s* in x5 and *G, w* in x6. After coupling of the first residue *y* in all TentaGel beads, beads were divided into equal parts by adding DMF to the reaction vial and dividing the volume into two different new vials. Each portion was coupled to the following amino acid as previously described, Fmoc was removed and beads were washed and pooled together, mixed and split again into different portions (two or three, depending on the library). This cycle was repeated until synthesis was completed and the desired peptide combinations were obtained. Finally, beads from each library were mixed together.

Ultra-high-performance liquid chromatography-mass spectrometry (UPLC-MS)

The UPLC system used for the detection of cleaved peptides with C-terminal amidation consisted of an Acquity H-class quaternary solvent manager and flow-through-needle sample manager with a photodiode array detector coupled to a Micromass ZQ Mass Spectrometer (electrospray ionization) (Waters). Chromatographic separation was performed on a BEH C18 column (50 mm x 2.1 mm, 1.7 μm) and analyzed using Masslynx 4.1 software. Compounds were separated with an isocratic flow consisting of 0% acetonitrile with 0.07% formic acid (FA) in the first 5 min followed by a linear gradient of 100% H_2O with 0.1% FA to 100% acetonitrile in the following 10 min at a flow rate of 0.6 mL/min.

Detection was performed at 214 nm using positive electrospray ionization mode. The mass/charge ratio (m/z) range was set to 500 to 1000 or 600 to 1200 amu for Library 1 and Library 2, respectively, and MS scan rate was set to 1 sec.

High-resolution mass spectrometry

Samples were diluted 1:2 in acetonitrile with 1% formic acid. A TriVersa NanoMate chip-based automated ion source platform (Advion Biosciences) allowed direct infusion electrospray ionization of samples into an LTQ-FT Ultra mass spectrometer (Thermo Fisher Scientific). Spray voltage was set to 1.7 kV, delivery pressure was set at 0.5 psi in positive mode and the voltage and temperature of the ion transfer tube were maintained at 35 V and 200ºC, respectively. All mass spectra were acquired across the m/z range 100-2000 amu. Elemental compositions from experimental exact monoisotopic mass were acquired with a dedicated algorithm integrated into the Xcalibur software, vs.2.0SR2 (Thermo Fisher Scientific).

Nanoscale liquid chromatography-tandem mass spectrometry (NanoLC-MS/MS)

1 μL of sample was introduced into NanoAcquity columns (trap column: 180 μm Å~ 2 cm C18 Symmetry; analytical column: BEH130 C18 75 mm Å~ 25 cm, 1.7 μm) (Waters). Flow rate was set at 250 nL/min. Mobile phase A and B consisted of H_2O with 0.1% FA and CH3CN with 0.1% FA, respectively. Gradient ranged from 1% B to 35% B over 60 min followed by a gradient of 35% B to 50% B over 5 min and a gradient of 50% to 85% over 2 min. Eluting metabolites were ionized by electrospray ionization with a TriVersa NanoMate chip-based ion source platform (Advion Biosciences) with the voltage set at 1.7 kV and delivery pressure set at 0.5 psi in positive mode. The ion source was coupled to an LTQ-FT Ultra mass spectrometer (Thermo Fisher Scientific) for high-precision mass determinations with the Xcalibur software 2.0, vs. SR2 (Thermo Fisher Scientific). All mass spectra were acquired across the m/z range 350-2000 amu at the operating mode "data-dependent acquisition 1MS1 + 6MS2". Heated capillary voltage was set at 40 V and temperature was maintained at 200ºC.

One-bead-one-compound (OBOC) peptide library cell binding assay

MCF7, MAF-overexpressiong MCF7 and BoM2 breast cancer cells were labeled with red fluorescent protein (RFP). Beads were equilibrated with phosphate-buffered saline (PBS) in a polypropylene tube O/N at 4ºC. Breast cancer cells were detached from the petri dish with a 1:1 mixture of Trypsin (Sigma-Aldrich) and enzyme-free cell dissociation buffer (Gibco), washed, resuspended in DMEM and counted. One million cells were added to polypropylene tubes containing 2.5 mg of library beads in a final volume of 5 mL and the mixture was incubated on a rotator for 4 h at room temperature (RT). After incubation, the cell-bead mixture was either imaged under a Nikon TE200 inverted microscope or prepared for flow cytometry analysis. To cover the necessary amount of beads for subsequent analysis, the cell binding assay was performed twice for each condition.

Sorting of positive hits

Cell-bead mixtures were inserted into a Complex Object Parametric Analyzer and Sorter (COPAS) large particle flow cytometer (Union Biometrica) or a FACSAria cell sorter (Becton Dickinson) and sorted into Amicon Ultra centrifugal filters (Merck Millipore). Sorting threshold and gating were established with Empty TentaGel beads that had never been incubated with cells and a population of RFP-expressing cells. Only those events with blue fluorescence (detection wavelength 450 nm) and red fluorescence (detection wavelength 610 nm) above the defined thresholds, representing the population of beads with strongest association with cells, were isolated. Isolated beads were treated with a 500 μM NaCl solution to remove bound cells, washed three times with H_2O and freeze-dried to eliminate the remaining liquid. Peptides were cleaved from the beads using ammonia vapors O/N and eluted in 50 μL 20% (v/v) acetonitrile/water. The sequence of cleaved peptides was analyzed by UPLC-MS and high-resolution mass spectrometry.

Immunofluorescence

A total of 1x105 cells were seeded onto glass coverslips. The next day, cells were fixed with formalin for 20 min, permeabilized with 0.5% Triton X-100 in PBS for 10 min, blocked with 1% BSA in PBS for 45 min and probed with primary antibodies O/N at 4ºC (See table 8). Primary antibodies were detected using Alexa Fluor 488-conjugated secondary antibodies (Invitrogen). Biotinylated proteins were detected using Alexa Fluor 546-conjugated streptavidin (Invitrogen). DAPI for nuclear staining was contained in the ProLong™ Gold antifade mounting solution (Invitrogen).

Confocal images were taken on the Zeiss LSM780 microscope (Zeiss, Jena, Germany) using a Plan Apochromat 63x/1.4 oil immersion objective, 405- and 488-nm laser excitation at a pixel resolution of 132 nm. Z-stacks were acquired every 500 nm.

Immunoblotting

Trypsinized cell pellets were resuspended in lysis buffer (50 mM Tris/HCl pH 7.4, 1% Triton X-100, 140 mM NaCl, 1 mM EDTA, 1 mM EGTA, 0.1% SDS, 1 mM NaF, 1mM Na_3VO_4) containing protease inhibitor cocktail (Roche), incubated 30 min on ice and centrifuged at 15000 rpm for 10 min at 4ºC. Supernatant was kept as protein extract and concentration was determined by standard Bradford assay (BioRad). 40 μg of protein lysate per sample were mixed with sample buffer (45 mM Tris pH=6.8, 10% glycerol, 1% SDS, 52 mM DTT and 1% bromophenol blue) and heated at 95ºC for 5 min. Proteins were separated by SDS-PAGE and transferred to polyvinylidene difluoride (PVDF) membranes (Millipore). Membranes were blocked in 5% BSA in TBS-Tween (0.1%) for 1 h at RT and incubated at 4ºC O/N with primary antibodies (See table 8). Next, membranes were washed three times with TBS-Tween (0.1%) and incubated 1h at RT with horseradish peroxidase (HRP)-conjugated secondary antibodies or HRP-conjugated streptavidin for biotin detection (See table 8). Finally, membranes were washed three times again and developed with ECL western blotting substrate (Pierce) according to the manufacturer's protocol.

Proximity-dependent biotin identification (BioID)

Short and long MAF isoforms were PCR amplified from pBabe-puro-MAF retroviral vectors using primers containing EcoRI and KpnI (N-terminus tag) or NheI and HpaI (C-terminal tag) restriction enzyme sites (See table 7) and cloned into myc-BioID2-MCS (Addgene plasmid #74223) and MCS-BioID2-MCS (Addgene plasmid #74224).

MCF7 or MDA-MB-231 cells were transfected with control (empty myc-BioID2) and MAF plasmids using GenJet DNA transfection reagent (SignaGen Laboratories) according to the manufacturer's instructions. Five 15 cm plates per condition were grown to 70% confluence prior to treatment with biotin (50 μM) for 24 h. MCF7 samples were performed with 2 biological replicates and MDA-MB-231 cells with 1 biological replicate. Trypsinized cell pellets were washed with ice-cold PBS and lysed in 5 mL of lysis buffer (50 mM Tris-HCl pH=8, 150 mM NaCl, 0.1% SDS, 2 mM $MgCl_2$, 1% Triton X-100, 1 mM EDTA, 1 mM EGTA, 1 mM NaF, 1 mM Na_3VO_4, protease inhibitor cocktail (Roche)) with 1:2000 benzonase (Sigma-

Aldrich) by rotating for 1 h at 4ºC. Samples were sonicated 3x30 s with Bioruptor (Diagenode) and centrifuge at 16,000 g for 30min at 4ºC. Byotinilated proteins were isolated by affinity purification with Dynabeads® MyOne Streptavidin C1 beads (Thermo Fisher Scientific) with rotation for 3 h at 4ºC. Beads where washed once with lysis buffer and three times with 50 mM NH_4HCO_3 and snap-frozen prior to tryptic digestion.

Samples were on bead tryptic digested at 0.1 µg/µL enzyme concentration in 50 mM NH_4HCO_3 at 37ºC O/N. The following morning, additional 1.08 µg trypsin (both replicates 1 and 2) were added and incubated 2 h at 37°C. To stop the digestion formic acid was added to 1% final concentration. Samples were cleaned up though polyLC C18 tips and peptides were eluted with 80% acetonitrile 1% TFA. Next, samples were diluted to 20% acetonitrile 0.1% TFA, loaded into strong cation exchange columns (SCX) and peptides were eluted in 5% NH_4OH 30% methanol. Finally, samples were evaporated to dryness, reconstituted in 50 µL and diluted 1/8 with 3% acetonitrile 1% formic acid aqueous solution for MS analysis.

The nano-LC-MS/MS set up was as follows. Digested peptides were diluted in 3% ACN 1% FA. Sample was loaded to a 300 µm × 5 mm PepMap100, 5 µm, 100 Å, C18 µ-precolumn (Thermo Scientific) at a flow rate of 15 µl/min using a Thermo Scientific Dionex Ultimate 3000 chromatographic system (Thermo Scientific). Peptides were separated using a C18 analytical column NanoEase MZ HSS T3 column (75 µm × 250 mm, 1.8 µm, 100 Å) (Waters) with a 90min run, comprising three consecutive steps with linear gradients from 3 to 35% B in 60min, from 35 to 50% B in 5min, and from 50 % to 85% B in 2min, followed by isocratic elution at 85 % B in 5min and stabilization to initial conditions (A=0.1% FA in water, B=0.1% FA in CH_3CN). The column outlet was directly connected to an Advion TriVersa NanoMate (Advion) fitted on an Orbitrap Fusion Lumos™ Tribrid (Thermo Scientific). The mass spectrometer was operated in a data-dependent acquisition (DDA) mode. Survey MS scans were acquired in the Orbitrap with the resolution (defined at 200 m/z) set to 120,000. The lock mass was user-defined at 445.12 m/z in each Orbitrap scan. The top speed (most intense) ions per scan were fragmented by CID and detected in the linear ion trap. The ion count target value was 400,000 and 10,000 for the survey scan and for the MS/MS scan respectively. Target ions already selected for MS/MS were dynamically excluded for 15 s. Spray voltage in the NanoMate source was set to 1.60 kV. RF Lens were tuned to 30%. Minimal signal required to trigger MS to MS/MS switch was set to 5,000. The spectrometer was working in positive polarity mode and singly charge state precursors were rejected for fragmentation.

A twin database search was performed with two different softwares, Thermo Proteome Discoverer v2.4.1.15 (PD) and MaxQuant v1.6.14.0 (MQ), using Sequest HT and Andromeda search engine nodes for PD and MQ, respectively. The database used in the search was SwissProt Human (released in October 2020) including contaminants and MAF proteins (short and long isoforms). Search was run against targeted and decoy databases to determine the false discovery rate (FDR). Search parameters included trypsin enzyme specificity, allowing for two missed cleavage sites, oxidation in M and acetylation in protein N-terminus as dynamic modifications. Peptide mass tolerance was 10 ppm and the MS/MS tolerance was 0.6 Da. Peptides were filtered at a FDR of 1 % based on the number of hits against the reversed sequence database.

For the quantitative analysis, contaminant identifications were removed and only unique peptides (those that are not shared between different protein groups) were used for the quantitative analysis with SAINTexpress-spc v3.6.1[400]. SAINTexpress compares the prey control spectral counts with the prey test spectral counts for all available replicates. For each available bait and for each available replicate, we took as prey count the maximum count result between PD and MQ. Once we obtained this combined dataset, we ran the SAINTexpress algorithm with BioID2-MAF samples (N- and C-terminal fusion, MAF short and long isoforms) and the corresponding control samples. The algorithm was independently executed for MCF7 (2 biological replicates with 3 technical replicates) and MDA-MB-231 (3 technical replicates) samples. For MCF7 samples, an "-R3" parameter was used in order to select the 3 replicates with the highest spectral counts. High confidence interactors were defined as those with a Bayesian FDR ≤ 0.02 and a fold change ≥ 3. Output SAINTexpress data is provided in Supplementary tables 1 and 2.

Network analysis

Protein-protein interaction data was downloaded from the STRING v11 database[401] with the following settings: meaning of network edges, confidence; active interaction sources, experiments and databases; minimum required interaction score, 0.9. Disconnected nodes in the network were hidden and the Markov Cluster (MCL) algorithm was used with an inflation parameter of 2.2.

Dot plot analysis

SAINTexpress output files of MAF-BioID2 baits (N- and C-terminal fusion, MAF short and long isoforms) or controls were processed through ProHits-viz[402] with the dot plot generator tool (default options) for visualization of selected high-confidence proximity interactors.

GO enrichment analysis for BioID interactors

Statistically enriched GO terms for BioID high-confidence interactors were identified using the standard hypergeometric test; Background population was defined by all MAF interactors that were observed in the BioID proteomics experiment. Significance was defined by the adjusted p-value using the Benjamini and Hochberg (BH) multiple testing correction.

Co-Immunoprecipitation (Co-IP)

Cells were transfected with pCMV5 plasmids for HA-tagged MAF expression (short and long isoforms) or an empty plasmid control using GenJet DNA transfection reagent (SignaGen Laboratories) according to the manufacturer's instructions. 48 h post-transfection cells were harvested in lysis buffer (50 mM Tris-HCl pH=7.4, 100 mM NaCl, 1% Triton X-100, 1 mM EDTA, 1 mM EGTA, 50 mM NaF, 2 mM Na_3VO_4, 10 mM β-glycerophosphate) supplemented with a protease inhibitor cocktail (Roche), incubated 30 min on ice and centrifuged at 15000 rpm for 10 min at 4ºC. Supernatant was kept as protein extract and concentration was determined by standard Bradford assay (BioRad). Input (20 μg of protein lysate) was mixed with 2x sample buffer (45 mM Tris pH=6.8, 10% glycerol, 1% SDS, 52 mM DTT and 1% bromophenol blue) and heated at 95ºC for 5 min prior to immunoblot analysis. 500 μg of the remaining protein lysate were incubated with 25 μL of anti-HA magnetic beads (Pierce) overnight at 4ºC with gentle agitation. The next day, beads were washed 5 times with lysis buffer using a magnetic stand and eluted by boiling with 2x sample buffer. Proteins were separated by SDS-PAGE before immunoblot analysis.

Proximity ligation assay (PLA)

A total of $1x10^5$ cells were seeded onto glass coverslips. The following day, cells were fixed with formalin, permeabilized with 0.5% Triton X-100 in PBS, blocked and probed with primary antibodies O/N at 4ºC (See table 8). PLA experiment was performed with Duolink PLA reagents (Sigma-Aldrich) following the manufacturer's protocol.

For PLA experiments with MAF truncation constructs, cells were transfected 24h prior to fixation with GenJet DNA transfection reagent (SignaGen Laboratories) according to the manufacturer's instructions. Then, PLA experiment was performed with Duolink PLA reagents followed by immunofluorescence using anti-HA primary antibodies (Sigma-Aldrich) and Alexa Fluor 488-conjugated secondary antibodies (Invitrogen) to detect transfected cells.

Confocal images were taken on the Zeiss LSM780 microscope (Zeiss, Jena, Germany) using a Plan Apochromat 63x/1.4 oil immersion objective, 405-, 488-, 543-nm laser excitation at a pixel resolution of 132 nm. Z-stacks were acquired every 500 nm to ensure a count of the PLA puncta over the whole cell volume. The percentage of cells with PLA punctate structures was obtained by counting at least 100 cells in each working condition from 3 independent experiments. Puncta detection was performed using a Fiji tailor made macro[403,404]. Briefly, the macro segmented the nuclei thresholding the DAPI signal and using the Analyze Particles ImageJ plugin. Then, the nuclei mask was applied on the PLA channel and puncta were detected using the Find Maxima plugin (red channel). Finally, the number of dots and nuclei were automatically counted.

For PLA experiments with MAF truncation constructs, the Fiji tailor made macro was modified to select the green nuclei after segmentation by thresholding the DAPI signal in order to ensure the quantification of specific signal from transfected cells. Then, the green nuclei mask was applied on the PLA channel (red) to count the number of dots.

For statistical analysis, PLA signal/nucleus quantifications with a 0.1 added constant were log2 transformed. A linear mixed effects model was fitted with the R package lmerTest[405] using the transformed values as response variable, the different conditions as covariate of interest, the biological replicate as adjusting factor and the technical replicate as random effect to account for the technical replicates variability. Adjustment for multiple testing (single-step correction method) was performed using the R package multcomp[406].

Treatments

Proteolysis Targeting Chimera (PROTAC) ER degrader was prepared in DMSO. Vehicle control (DMSO) or PROTAC 1 μM was added to the cell culture media and different timepoints were tested for the selection of the optimal treatment duration. Final experiments were performed by treating the cells with DMSO or PROTAC 1 μM for 24 h prior to immunoblot or PLA analyses.

RNA-sequencing (RNA-seq)

RNA-seq experiments were performed with two biological replicates of control or MAF-overexpressing MCF7 cells treated with 10nM E2 or vehicle for 6 h. RNA was extracted using PureLink RNA mini kit (Invitrogen) following the manufacturers' instructions. RNA samples were quantified using Qubit Fluorometric Quantification (Thermo Fisher Scientific) and the quality was evaluated using Bioanalyzer (Agilent). Libraries were prepared at the Institute for Research in Biomedicine (IRB Barcelona) and sequencing was performed at the Centre for Genomic Regulation (CRG) Genomics Unit, using 1μg of total RNA and the Illumina HiSeq2500 sequencer.

Bioinformatics analysis:

Reads were aligned to the hg19 genome using STAR 2.7.0e[407], with outFilterMismatchNoverLmax = 0.05 and all others parameters set to default values. Counts per genomic feature were computed with the R[408] package Rsubread[409], function featureCounts. Differential expression between conditions, taking the sequencing round as adjusting factor, was performed using the DESeq2 R package[410]. Comparisons MAF vs Mock, Mock E2 vs Mock, MAF E2 vs MAF, MAF E2 vs Mock E2 and MAF E2 vs Mock (with condition vs control labeling) were all considered for hypothesis testing. The regularized log transformed matrix was used for both selecting candidates and visualization purposes.

Genes with less or equal than an average of five reads were filtered out. We considered 8 different clusters of genes following the next criteria:

Cluster 1 : MAF positive – Genes with both test statistic of MAF vs Mock and MAF E2 vs E2 greater than 2 ; Cluster 2 : MAF negative – Genes with both test statistic of MAF vs Mock and MAF E2 vs Mock E2 smaller than -2; Cluster 3: E2 positive – Genes with both test statistic of Mock E2 vs Mock and MAF E2 vs Mock greater than 2.5; Cluster 4: E2 negative – Genes with both test statistic of Mock E2 vs Mock and MAF E2 vs Mock smaller than -2.5; Cluster 5: E2 and MAF positive – Genes with test statistic of MAF vs Mock, MockE2 vs Mock, MAF E2 vs Mock E2 and MAF E2 vs MAF greater than 1.5, and normalized expression rlog(MAF E2) > rlog(any other condition) + 0.05; Cluster 6: ES positive when MAF – Genes with test statistic of MAF E2 vs MAF > 2 and logfc difference between MAF E2 vs MAF and Mock E2 vs Mock > 0.2, and expression rlog(MAF E2) > rlog(any other condition) + 0.05; Cluster 7: E2 and MAF negative – Genes with test statistic of MAF vs Mock, Mock E2 vs Mock, MAF E2 vs Mock E2 and MAF E2 vs MAF smaller than -1.5, and normalized expression rlog(MAF E2) < rlog(any other condition) - 0.05; Cluster 8: ES negative when MAF – Genes with test statistic of MAF E2 vs MAF < -2 and logfc difference between MAF E2 vs MAF and Mock E2 vs Mock < -0.2, and expression rlog(MAF E2) < rlog(any other condition) - 0.05.

For each cluster of genes, statistically enriched GO terms were identified using the standard hypergeometric test; Significance was defined by the adjusted p-value using the Benjamini and Hochberg (BH) multiple testing correction.

Reverse transcription quantitative polymerase chain reaction (RT-qPCR)

Total RNA from control or MAF-overexpressing MCF7 cells treated with 10nM E2 or vehicle for 6 h was extracted using a PureLink RNA mini kit (Invitrogen) following the manufacturers' instructions. RNA quality and quantity was evaluated using a NanoDrop One spectrophotometer (Thermo Fisher Scientific). cDNA was generated by reverse transcription from 1µg of RNA using the high capacity cDNA reverse transcription kit (Applied Biosystems) and a C1000 Touch Thermal Cycler (Bio-Rad). Real-time PCR reactions were performed using TaqMan universal PCR master mix and specific TaqMan probes (both from Applied Biosystems) (See table 9). Expression values were normalized to the housekeeping gene *GAPDH* using the comparative CT method. Statistical significance was assessed using a two-tailed unpaired Student's t-test. The symbols *, ** and *** indicate significant differences with $p < 0.05$, $p < 0.01$ and $p < 0.001$, respectively.

Assay of transposase accessible chromatin sequencing (ATAC-seq)

ATAC-seq experiment were performed as described in Buenrostro *et al*, 2013[411]. Briefly, 50,000 control or MAF-overexpressing MCF7 cells treated with 10nM E2 or vehicle for 1 h were collected and treated with transposase Tn5 (Nextera DNA library preparation kit, Illumina). DNA was purified using a PureLink quick gel extraction and PCR purification combo kit (Invitrogen). All samples were PCR amplified using NEBNextHigh-Fidelity 2x PCR Master Mix (New England BioLabs) with primers containing a barcode to generate libraries and a C1000 Touch Thermal Cycler (Bio-Rad). Then, DNA was purified again using a PureLink quick gel extraction and PCR purification combo kit. Samples were quantified using Qubit Fluorometric Quantification (Thermo Fisher Scientific) and the quality was evaluated using Bioanalyzer (Agilent). Sequencing was performed at the CRG Genomics Unit, using the Illumina HiSeq2500 sequencer.

Bioinformatics analysis:

Cleaning of adapters was completed using Trimmomatic v0.38[412]. Cleaned reads were then aligned to the hg19 human genome using Bowtie2 v2.2.2[413] with "very sensitive" option and the remainder parameters set to default values. Genrich v0.5[414] was employed to call peaks, separately for samples in Mock, samples in Mock E2, samples in MAF and samples in MAF E2 (options -j -y -r -e chrM). Duplicate sequences were filtered out, sorted and indexed using sambamba, v-0.6.7[415]. The union of called peaks by Genrich (consensus peaks) were annotated with Homer[416] and used for downstream analysis. Counts per consensus peaks were obtained for every sample using the featureCounts R[408] function, package Rsubread[409], ignoring duplicated reads.

For differential analysis of consensus peaks, comparisons between pairwise conditions (Mock, Mock E2, MAF and MAF E2) were assessed using DESeq2 v1.22.1[410], considering the sequencing pool as adjusting variable.

For integration of and RNA-seq and ATAC-seq data, RNA-seq E2-responsive genes (both up- and down-regulated) and MAF-responsive genes (both up- and down-regulated) enrichment in ATAC-seq peaks was evaluated. E2-dependent chromatin opening was defined by peaks (distance to TSS < 50 kb) that had a test-statistic larger than 1 in both Mock E2 vs Mock and MAF E2 vs MAF comparisons, DESeq2 results from ATAC-seq data. MAF-

chromatin opening was defined by peaks (distance to TSS < 50 kb) that had a test-statistic larger than 1 in both MAF vs Mock and MAF E2 vs Mock E2 comparisons, DESeq2 results from ATAC-seq data. The rotation-based approach for enrichment[417] implemented in the R package limma[418] was used to represent the null distribution. The maxmean enrichment statistic proposed in Efron, 2007[419], under restandardization, was considered for competitive testing.

Chromatin immunoprecipitation followed by sequencing (ChIP-seq)

Four 15cm plates per condition were prepared at 70-80% confluency. After treating control or MAF-overexpressing MCF7 cells with 10nM E2 or vehicle for 1 h, cells were crosslinked in 1% formaldehyde in DMEM for 10 min at RT. Then, glycine was added to a final concentration of 0.125 M and cells were incubated for 5 min at RT to stop the fixation. After two washes with ice-cold PBS, cells were harvested by gently scrapping on ice and centrifuged at 3,000 x g for 5 min. Cell pellets were stored at -80ºC until use. Chromatin preparation and ChIP experiments were performed with the ChIP-IT High Sensitivity Kit (Active Motif) according to manufacturer's protocol. ChIPs were performed using 5 μg/ChIP on ER antibodies (See table 8) and control IgG (Abcam). Library preparation was performed at the CRG Genomics Unit using the Illumina HiSeq2500 sequencer.

Bioinformatics analysis:

ChIP-seq samples were mapped against the hg19 human genome assembly using Bowtie with the option –m 1 to discard those reads that could not be uniquely mapped to just one region[420]. MACS was run with the default parameters but with the shift-size adjusted to 100 bp to perform the peak calling against the corresponding control sample[421].

The genome distribution of each set of peaks was calculated by counting the number of peaks fitted on each class of region according to RefSeq annotations[422]. Promoter is the region between 2.5 Kbp upstream and 2.5 Kbp downstream of the transcription start site (TSS). Genic regions correspond to the rest of the gene (the part that is not classified as promoter) and the rest of the genome is considered to be intergenic. Peaks that overlapped with more than one genomic feature were proportionally counted the same number of times.

To identify a list of putative targets of each ChIP-seq peak, we gathered all the genes in the vicinity of that peak at 50 Kbp. As a random control, we repeated the same operation over equivalent lists of peaks in which their position was shifted 1 Mbp. To identify a list of putative enhancers of each differentially expressed gene, we associated the ChIP-seq peaks of one experiment in the vicinity of that gene at 50 Kbp.

The aggregated plots showing the average distribution of ChIP-seq reads of a collection of peaks were generated by counting the number of reads around the summit of each peak and normalizing for the total number of peaks in the particular gene set.

The heatmaps displaying the density of ChIP-seq reads around the summit of each ChIP-seq peak were generated by counting the number of reads in this region for each individual peak and normalizing this value with the total number of mapped reads of the sample. Peaks on each ChIP heatmap were ranked by the logarithm of the average number of reads in the same genomic region. Values corresponding to pairs of comparable heatmaps were normalized together by the maximum value within both experiments.

Boxplots showing the ChIP-seq level distribution for particular ChIP-seq experiments on a set of genomic peaks were calculated by determining the maximum value on this region at this sample, which was assigned afterwards to the corresponding peak, and normalizing for the total number of reads of each experiment.

Each point on the scatterplots of ChIP-seq intensities between MAF E2 and Mock E2 was calculated by determining the maximum value of the sample inside each peak at each condition. Values were normalized by the total number of reads of each ChIP-seq experiment.

The MatInspector program[423] from the Genomatix software was used to identify MAF binding motifs (MARE half and MAF) within ER ChIP-seq peaks identified in the MAF E2 sample. The number of input sequences with at least one match of the MARE (half) or MAF matrix and the number of matches in all input sequences was calculated. The expected match numbers in an equally sized sample of the genome was calculated assuming that the matches were equally distributed in the genome. Motif overrepresentation was computed as the fold factor of match numbers in regions compared to an equally sized sample of the genome. A Z-score of motif overrepresentation was calculated with a continuity correction using the formula $z= (X-E-0.5)/S$, where X is the number of found matches in ER ChIP-seq peaks identified in the MAF E2 sample, E is the expected value and S is the standard deviation.

The UCSC genome browser was used to generate the screenshots of ChIP-seq profiles[424].

Correlation analyses of transcriptomic datasets

A public data cohort containing four micro-array studies (GSE2430, GSE2603, GSE5327 and GSE12276) was processed and merged as detailed in Gawrzak *et al*, 2018[236]. Expression values were adjusted genewise by HER2 status, data set under study, Eklund's metrics[425] and the interaction between data set and Eklund's metrics. If all samples were used, ER status was considered as an adjusting variable as well.

TCGA breast cancer counts data[426] were log2 (+ 0.25 to avoid zero misspecifications) transformed and quantile normalized. Only Stage I, II and III tumors were considered for the analysis. Expression values in log scale were adjusted genewise by HER2 status and tumor stage information. If all samples were used, ER status was also used as an adjusting variable.

Metabric expression matrix[427,428], for Stage I, II and III tumors only, was adjusted genewise by HER2 status, tumor grade, lymph nodes (in log scale), breast surgery status, chemotherapy status, tumor stage and hormone therapy status. ER status was also considered as an adjusting variable when suitable.

The Spearman correlation between MAF and several target genes was estimated using the three corrected expression matrices (public cohort data, TCGA and Metabric). Assessing the statistical significance of the estimated correlation values was performed with the cor.test R function[408].

MATERIALS

Table 6. List of plasmids.

Name	Selection	Supplier	Notes
FUW-ubiquitin-SV40-RFP	Flow citometry	Addgene (#65448)	Plasmid for stable RFP expression. Lentiviral vector
pBabe-puro	Puromycin		Empty pBabe plasmid. Retroviral vector
pBabe-puro-MAF S	Puromycin		Plasmid for stable MAF S expression. Retroviral vector
pBabe-puro-MAF L	Puromycin		Plasmid for stable MAF L expression. Retroviral vector
myc-BioID2-MCS		Addgene (#74223)	Plasmid to fuse a protein of interest to the C-terminus of BioID2
MCS-BioID2-HA		Addgene (#74224)	Plasmid to fuse a protein of interest to the N-terminus of BioID2
myc-BioID2-MAF S			Plasmid for BioID2-MAF S fusion protein expression
MAF S-BioID2-HA			Plasmid for MAF S-BioID2 fusion protein expression
myc-BioID2-MAF L			Plasmid for BioID2-MAF L fusion protein expression
MAF L-BioID2-HA			Plasmid for MAF L-BioID2 fusion protein expression
pCMV5-HA			pCMV5 modified to contain 2xHA tags
pCMV5-HA-MAF S			Plasmid for 2xHA-tagged MAF S expression
pCMV5-HA-MAF L			Plasmid for 2xHA-tagged MAF L expression
pCMV5-HA-MAF (ΔC-t)			Plasmid for 2xHA-tagged truncated MAF expression (Amino acids 1-283. No DNA-binding domain)
pCMV5-HA-MAF L (ΔN-t 1)			Plasmid for 2xHA-tagged truncated MAF L expression (Amino acids 85-404. Without part of the transactivation domain)
pCMV5-HA-MAF L (ΔN-t 2)			Plasmid for 2xHA-tagged truncated MAF L expression (Amino acids 120-404. No transactivation domain)
MSCV-Neo	Neomycin		Empty MSCV plasmid. Retroviral vector
MSCV-Neo-HA-MAF S	Neomycin		Plasmid for stable 2xHA-tagged MAF S expression. Retroviral vector
MSCV-Neo-HA-MAF L	Neomycin		Plasmid for stable 2xHA-tagged MAF L expression. Retroviral vector

Table 7. List of primers.

Name	Primer Sequence (5' to 3')	Notes
Primers to clone MAF into myc-BioID2-MCS plasmid		
EcoRI-MAF	ATATGAATTCATGGCATCAGAACTGGCAAT	Primer for MAF amplification with N-terminal EcoRI restriction site
MAF S-KpnI	TTTTGGTACCTCACATGAAAAACTCGGG	Primer for MAF S amplification with C-terminal KpnI restriction site
MAF L-KpnI	TTTTGGTACCTCATTTTGTGAACACACTGG	Primer for MAF L amplification with C-terminal KpnI restriction site
Primers to clone MAF into MCS-BioID2-HA plasmid		
NheI-MAF	TTTTGCTAGCATGGCATCAGAACTGG	Primer for MAF amplification with N-terminal NheI restriction site
MAF S (No STOP codon)-HpaI	TTTTGTTAACCATGAAAAACTCGGGAGAGG	Primer for MAF S amplification with C-terminal KpnI restriction site (No STOP codon)
MAF L (No STOP codon)-HpaI	TTTTGTTAACTTTTGTGAACACACTGGTAAGTACAC	Primer for MAF L amplification with C-terminal KpnI restriction site (No STOP codon)
Primers to clone MAF into pCMV5 modified to contain 2xHA tags		
EcoRI-MAF	ATATGAATTCATGGCATCAGAACTGGCAAT	Primer for MAF amplification with N-terminal EcoRI restriction site
MAF S-BamHI	TTTTGGATCCTCACATGAAAAACTCGGGA	Primer for MAF S amplification with C-terminal BamHI restriction site
MAF L-BamHI	TTTTGGATCCTCATTTTGTGAACACACTG	Primer for MAF L amplification with C-terminal BamHI restriction site
MAF283-BamHI	TTTTGGATCCTCACTTGCTGACCCCGC	Primer for truncated MAF (ΔC-t) amplification with C-terminal BamHI restriction site at amino acid 283
EcoRI-MAF85	AAAAGAATTCAAGGCGCACCTGGAAGAC	Primer for truncated MAF (NC-t 1) amplification with N-terminal EcoRI restriction site at amino acid 85
EcoRI-MAF120	AAAAGAATTCAACAGCCACCAGCTCCAG	Primer for truncated MAF (NC-t 2) amplification with N-terminal EcoRI restriction site at amino acid 120
Primers to clone 2xHA-tagged MAF into MSCV-Neo plasmid		
HpaI-HA	AAAAGTTAACATGGATTATCCGTATGATGTTCCTG	Primer for HA-tagged MAF amplification with N-terminal HpaI restriction site
MAF S-HpaI	TTTTGTTAACTCACATGAAAAACTCGGGAGAG	Primer for HA-tagged MAF S amplification with C-terminal HpaI restriction site

Table 8. List of antibodies.

Antigen	ID	Supplier	Source	Application	Dilution
CD29-APC (Integrin β1)	559883	BD Pharmingen	Mouse	Flow citometry	1:1000
CD49f-PE/Cy7 (Integrin α6)	313621	Biolegend	Rat	Flow citometry	1:1000
MAF	130(5)	Inbiomotion	Rabbit	WB	1:50
HA-Tag	H3663	Sigma-Aldrich	Mouse	WB	1:1000
				IF / PLA	1:500
ERα	ab16660	Abcam	Rabbit	WB	1:1000
				IF / PLA	1:500
MYC-Tag	2272	cell signaling	Rabbit	WB	1:1000
ARID1A	HPA005456	Sigma-Aldrich	Rabbit	WB	1:1000
KDM1A/LSD1	ab17721	Abcam	Rabbit	WB	1:1000
MTA1	sc-17773	Santa Cruz	Mouse	WB	1:200
NCoR1	sc-515934	Santa Cruz	Mouse	WB	1:200
NCoA3	sc-5305	Santa Cruz	Mouse	WB	1:200
α-Tubulin	T6199	Sigma Aldrich	Mouse	WB	1:1000
ERα	sc-543	Santa Cruz	Rabbit	ChIP	
IgG	ab172730	Abcam	Rabbit	ChIP	
Rabbit IgG HRP-conjugated	NA934	GE Healthcare	Donkey	WB	1:5000
Mouse IgG HRP-conjugated	31452	ThermoFisher Scientific	Rabbit	WB	1:5000
Streptavidin HRP-conjugated	ab7403	Abcam		WB	1:1000
Streptavidin Alexa Fluor 546-conjugated	S11225	Invitrogen		IF	1:500
Rabbit IgG Alexa Fluor 488-conjugated	A-11008	Invitrogen	Goat	IF	1:500
Mouse IgG Alexa Fluor 488-conjugated	A-11001	Invitrogen	Goat	IF	1:500

Table 9. List of TaqMan probes.

Gene Symbol	ID	Supplier
GAPDH	Hs02786624_g1	Applied Biosystems
MAF	Hs00193519_m1	Applied Biosystems
FGF18	Hs00818572_m1	Applied Biosystems
JAG1	Hs00164982_m1	Applied Biosystems
PTHLH	Hs00174969_m1	Applied Biosystems
SOX9	Hs00165814_m1	Applied Biosystems

Supplementary material

SUPPLEMENTARY MATERIAL

Supplementary Table 1.

Bibliography

BIBLIOGRAPHY

1. Yalaza, M., İnan, A. & Bozer, M. Male Breast Cancer. *J. breast Heal.* **12**, 1–8 (2016).

2. Bray, F. *et al.* Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA. Cancer J. Clin.* **68**, 394–424 (2018).

3. Berry, D. A. *et al.* Effect of Screening and Adjuvant Therapy on Mortality from Breast Cancer. *N. Engl. J. Med.* **353**, 1784–1792 (2005).

4. Hennighausen, L. & Robinson, G. W. Information networks in the mammary gland. *Nat. Rev. Mol. Cell Biol.* **6**, 715–725 (2005).

5. Williams, J. M. & Daniel, C. W. Mammary ductal elongation: Differentiation of myoepithelium and basal lamina during branching morphogenesis. *Dev. Biol.* **97**, 274–290 (1983).

6. Shackleton, M. *et al.* Generation of a functional mammary gland from a single stem cell. *Nature* **439**, 84–88 (2006).

7. Stingl, J. *et al.* Purification and unique properties of mammary epithelial stem cells. *Nature* **439**, 993–997 (2006).

8. Dontu, G. & Ince, T. A. Of mice and women: a comparative tissue biology perspective of breast stem cells and differentiation. *J. Mammary Gland Biol. Neoplasia* **20**, 51–62 (2015).

9. Gregor, M. F. *et al.* The Role of Adipocyte XBP1 in Metabolic Regulation during Lactation. *Cell Rep.* **3**, 1430–1439 (2013).

10. Hugo, E. R., Borcherding, D. C., Gersin, K. S., Loftus, J. & Ben-Jonathan, N. Prolactin Release by Adipose Explants, Primary Adipocytes, and LS14 Adipocytes. *J. Clin. Endocrinol. Metab.* **93**, 4006–4012 (2008).

11. Feuermann, Y., Mabjeesh, S. J. & Shamay, A. Mammary Fat Can Adjust Prolactin Effect on Mammary Epithelial Cells via Leptin and Estrogen. *Int. J. Endocrinol.* **2009**, 427260 (2009).

12. Walden, P. D., Ruan, W., Feldman, M. & Kleinberg, D. L. Evidence That the Mammary Fat Pad Mediates the Action of Growth Hormone in Mammary Gland Development. *Endocrinology* **139**, 659–662 (1998).

13. Ehrhart, E. J., Segarini, P., Tsang, M. L., Carroll, A. G. & Barcellos-Hoff, M. H. Latent transforming growth factor beta1 activation in situ: quantitative and functional evidence after low-dose gamma-irradiation. *FASEB J.* **11**, 991–1002 (1997).

14. Rahimi, N., Saulnier, R., Nakamura, T., Park, M. & Elliott, B. Role of Hepatocyte Growth Factor in Breast Cancer: A Novel Mitogenic Factor Secreted by Adipocytes. *DNA Cell Biol.* **13**, 1189–1197 (1994).

15. Lin, Y. & Li, Q. Expression and function of leptin and its receptor in mouse mammary gland. *Sci. China Ser. C Life Sci.* **50**, 669–675 (2007).

16. Hovey, R. C., Goldhar, A. S., Baffi, J. & Vonderhaar, B. K. Transcriptional Regulation of Vascular Endothelial Growth Factor Expression in Epithelial and Stromal Cells during Mouse Mammary Gland Development. *Mol. Endocrinol.* **15**, 819–831 (2001).

17. Inman, J. L., Robertson, C., Mott, J. D. & Bissell, M. J. Mammary gland development: cell fate specification, stem cells and the microenvironment. *Development* **142**, 1028 LP – 1042 (2015).

18. Gouon-Evans, V., Rothenberg, M. E. & Pollard, J. W. Postnatal mammary gland development requires macrophages and eosinophils. *Development* **127**, 2269 LP – 2282 (2000).

19. O'Brien, J., Martinson, H., Durand-Rougely, C. & Schedin, P. Macrophages are crucial for epithelial cell death and adipocyte repopulation during mammary gland involution. *Development* **139**, 269 LP – 275 (2012).

20. Russo, J. & Russo, I. H. Development of the human breast. *Maturitas* **49**, 2–15 (2004).

21. Brisken, C. & O'Malley, B. Hormone action in the mammary gland. *Cold Spring Harb. Perspect. Biol.* **2**, a003178–a003178 (2010).

22. Wagner, K. U. *et al.* Oxytocin and milk removal are required for post-partum mammary-gland development. *Genes Funct.* **1**, 233–244 (1997).

23. Kleinberg, D. L. Role of IGF-I in normal mammary development. *Breast Cancer Res.*

Treat. **47**, 201–208 (1998).

24. Cannata, D. *et al.* Elevated Circulating IGF-I Promotes Mammary Gland Development and Proliferation. *Endocrinology* **151**, 5751–5761 (2010).

25. Robinson, G. W. & Hennighausen, L. Inhibins and activins regulate mammary epithelial cell differentiation through mesenchymal-epithelial interactions. *Development* **124**, 2701–2708 (1997).

26. Martin, A.-M. & Weber, B. L. Genetic and Hormonal Risk Factors in Breast Cancer. *JNCI J. Natl. Cancer Inst.* **92**, 1126–1135 (2000).

27. Simpson, E. R. & Davis, S. R. Minireview: aromatase and the regulation of estrogen biosynthesis--some new perspectives. *Endocrinology* **142**, 4589–4594 (2001).

28. Mueller, S. O., Clark, J. A., Myers, P. H. & Korach, K. S. Mammary Gland Development in Adult Mice Requires Epithelial and Stromal Estrogen Receptor α. *Endocrinology* **143**, 2357–2365 (2002).

29. Förster, C. *et al.* Involvement of estrogen receptor β in terminal differentiation of mammary gland epithelium. *Proc. Natl. Acad. Sci.* **99**, 15578 LP – 15583 (2002).

30. Hewitt, S. C. & Korach, K. S. Estrogen Receptors: New Directions in the New Millennium. *Endocr. Rev.* **39**, 664–675 (2018).

31. Collaborative Group on Hormonal Factors in Breast Cancer. Menarche, menopause, and breast cancer risk: individual participant meta-analysis, including 118 964 women with breast cancer from 117 epidemiological studies. *Lancet. Oncol.* **13**, 1141–1151 (2012).

32. Mørch, L. S. *et al.* Contemporary Hormonal Contraception and the Risk of Breast Cancer. *N. Engl. J. Med.* **377**, 2228–2239 (2017).

33. Chlebowski, R. T. *et al.* Breast Cancer After Use of Estrogen Plus Progestin and Estrogen Alone: Analyses of Data From 2 Women's Health Initiative Randomized Clinical Trials. *JAMA Oncol.* **1**, 296–305 (2015).

34. Yue, W. *et al.* Genotoxic metabolites of estradiol in breast: potential mechanism of estradiol induced carcinogenesis. *J. Steroid Biochem. Mol. Biol.* **86**, 477–486 (2003).

35. Onland-Moret, N. C. *et al.* Urinary endogenous sex hormone levels and the risk of postmenopausal breast cancer. *Br. J. Cancer* **88**, 1394–1399 (2003).

36. Yager, J. D. & Davidson, N. E. Estrogen Carcinogenesis in Breast Cancer. *N. Engl. J. Med.* **354**, 270–282 (2006).

37. Pasqualini, J. R. *et al.* Concentrations of estrone, estradiol, and estrone sulfate and evaluation of sulfatase and aromatase activities in pre- and postmenopausal breast cancer patients. *J. Clin. Endocrinol. Metab.* **81**, 1460–1464 (1996).

38. Giuliano, A. E. *et al.* Breast Cancer—Major changes in the American Joint Committee on Cancer eighth edition cancer staging manual. *CA. Cancer J. Clin.* **67**, 290–303 (2017).

39. Rivenbark, A. G., O'Connor, S. M. & Coleman, W. B. Molecular and Cellular Heterogeneity in Breast Cancer: Challenges for Personalized Medicine. *Am. J. Pathol.* **183**, 1113–1124 (2013).

40. Burstein, H. J., Polyak, K., Wong, J. S., Lester, S. C. & Kaelin, C. M. Ductal Carcinoma in Situ of the Breast. *N. Engl. J. Med.* **350**, 1430–1441 (2004).

41. Erbas, B., Provenzano, E., Armes, J. & Gertig, D. The natural history of ductal carcinoma in situ of the breast: a review. *Breast Cancer Res. Treat.* **97**, 135–144 (2006).

42. Simpson, P. T., Reis-Filho, J. S., Gale, T. & Lakhani, S. R. Molecular evolution of breast cancer. *J. Pathol.* **205**, 248–254 (2005).

43. Sinn, H.-P. & Kreipe, H. A Brief Overview of the WHO Classification of Breast Tumors, 4th Edition, Focusing on Issues and Updates from the 3rd Edition. *Breast Care* **8**, 149–154 (2013).

44. Hammond, M. E. H. *et al.* American Society of Clinical Oncology/College Of American Pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer. *J. Clin. Oncol.* **28**, 2784–2795 (2010).

45. Bustreo, S. *et al.* Optimal Ki67 cut-off for luminal breast cancer prognostic evaluation: a large case series study with a long-term follow-up. *Breast Cancer Res. Treat.* **157**, 363–371 (2016).

46. Wolff, A. C. *et al.* Human Epidermal Growth Factor Receptor 2 Testing in Breast Cancer: American Society of Clinical Oncology/College of American Pathologists

Clinical Practice Guideline Focused Update. *J. Clin. Oncol.* **36**, 2105–2122 (2018).

47. Parise, C. A. & Caggiano, V. Breast Cancer Survival Defined by the ER/PR/HER2 Subtypes and a Surrogate Classification according to Tumor Grade and Immunohistochemical Biomarkers. *J. Cancer Epidemiol.* **2014**, 469251 (2014).

48. Iqbal, N. & Iqbal, N. Human Epidermal Growth Factor Receptor 2 (HER2) in Cancers: Overexpression and Therapeutic Implications. *Mol. Biol. Int.* **2014**, 852748 (2014).

49. Costa, R. L. B. & Gradishar, W. J. Triple-Negative Breast Cancer: Current Practice and Future Directions. *J. Oncol. Pract.* **13**, 301–303 (2017).

50. Giuliano, A. E., Edge, S. B. & Hortobagyi, G. N. Eighth Edition of the AJCC Cancer Staging Manual: Breast Cancer. *Ann. Surg. Oncol.* **25**, 1783–1785 (2018).

51. Perou, C. M. *et al.* Molecular portraits of human breast tumours. *Nature* **406**, 747–752 (2000).

52. Sørlie, T. *et al.* Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc. Natl. Acad. Sci.* **98**, 10869 LP – 10874 (2001).

53. Sørlie, T. *et al.* Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc. Natl. Acad. Sci.* **100**, 8418 LP – 8423 (2003).

54. Yersal, O. & Barutca, S. Biological subtypes of breast cancer: Prognostic and therapeutic implications. *World J. Clin. Oncol.* **5**, 412–424 (2014).

55. Sabatier, R. *et al.* Claudin-low breast cancers: clinical, pathological, molecular and prognostic characterization. *Mol. Cancer* **13**, 228 (2014).

56. Prat, A. *et al.* Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. (2017). doi:10.1186/bcr2635

57. Herschkowitz, J. I. *et al.* Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. *Genome Biol.* **8**, R76 (2007).

58. Prat, A. *et al.* Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. *Breast Cancer Res.* **12**, R68 (2010).

59. Peppercorn, J., Perou, C. M. & Carey, L. A. Molecular Subtypes in Breast Cancer

Evaluation and Management: Divide and Conquer. *Cancer Invest.* **26**, 1–10 (2008).

60. Prat, A. *et al.* Prognostic Value of Intrinsic Subtypes in Hormone Receptor–Positive Metastatic Breast Cancer Treated With Letrozole With or Without Lapatinib. *JAMA Oncol.* **2**, 1287–1294 (2016).

61. Cejalvo, J. M. *et al.* Intrinsic Subtypes and Gene Expression Profiles in Primary and Metastatic Breast Cancer. *Cancer Res.* **77**, 2213 LP – 2221 (2017).

62. Weigelt, B., Baehner, F. L. & Reis-Filho, J. S. The contribution of gene expression profiling to breast cancer classification, prognostication and prediction: a retrospective of the last decade. *J. Pathol.* **220**, 263–280 (2010).

63. Whitworth, P. *et al.* Chemosensitivity and Endocrine Sensitivity in Clinical Luminal Breast Cancer Patients in the Prospective Neoadjuvant Breast Registry Symphony Trial (NBRST) Predicted by Molecular Subtyping. *Ann. Surg. Oncol.* **24**, 669–675 (2017).

64. Parker, J. S. *et al.* Supervised risk predictor of breast cancer based on intrinsic subtypes. *J. Clin. Oncol.* **27**, 1160–1167 (2009).

65. Waks, A. G. & Winer, E. P. Breast Cancer Treatment: A Review. *JAMA* **321**, 288–300 (2019).

66. Matsen, C. B. & Neumayer, L. A. Breast cancer: a review for the general surgeon. *JAMA Surg.* **148**, 971–979 (2013).

67. Early Breast Cancer Trialists' Collaborative Group (EBCTCG). Effect of radiotherapy after breast-conserving surgery on 10-year recurrence and 15-year breast cancer death: meta-analysis of individual patient data for 10 801 women in 17 randomised trials. *Lancet* **378**, 1707–1716 (2011).

68. Sato, K. *et al.* Catheter-based delineation of lumpectomy cavity for accurate target definition in partial-breast irradiation with multicatheter interstitial brachytherapy. *J. Contemp. Brachytherapy* **11**, 108–115 (2019).

69. Cortazar, P. *et al.* Pathological complete response and long-term clinical benefit in breast cancer: the CTNeoBC pooled analysis. *Lancet* **384**, 164–172 (2014).

70. Murphy, B. L., Day, C. N., Hoskin, T. L., Habermann, E. B. & Boughey, J. C. Neoadjuvant Chemotherapy Use in Breast Cancer is Greatest in Excellent Responders:

Triple-Negative and HER2+ Subtypes. *Ann. Surg. Oncol.* **25**, 2241–2248 (2018).

71. Andreopoulou, E. & Sparano, J. A. Chemotherapy in Patients with Anthracycline- and Taxane-Pretreated Metastatic Breast Cancer: An Overview. *Curr. Breast Cancer Rep.* **5**, 42–50 (2013).

72. Korkmaz, A., Topal, T. & Oter, S. Pathophysiological aspects of cyclophosphamide and ifosfamide induced hemorrhagic cystitis; implication of reactive oxygen and nitrogen species as well as PARP activation. *Cell Biol. Toxicol.* **23**, 303–312 (2007).

73. Oun, R., Moussa, Y. E. & Wheate, N. J. The side effects of platinum-based chemotherapy drugs: a review for chemists. *Dalton Trans.* **47**, 6645–6653 (2018).

74. Torti, F. M. *et al.* Cardiotoxicity of epirubicin and doxorubicin: assessment by endomyocardial biopsy. *Cancer Res.* **46**, 3722–3727 (1986).

75. Schiff, P. B., Fant, J. & Horwitz, S. B. Promotion of microtubule assembly in vitro by taxol. *Nature* **277**, 665–667 (1979).

76. Zasadil, L. M. *et al.* Cytotoxicity of paclitaxel in breast cancer is due to chromosome missegregation on multipolar spindles. *Sci. Transl. Med.* **6**, 229ra43 (2014).

77. Ringel, I. & Horwitz, S. B. Studies with RP 56976 (taxotere): a semisynthetic analogue of taxol. *J. Natl. Cancer Inst.* **83**, 288–291 (1991).

78. Jones, S. E. *et al.* Randomized phase III study of docetaxel compared with paclitaxel in metastatic breast cancer. *J. Clin. Oncol.* **23**, 5542–5551 (2005).

79. Latchman, J., Guastella, A. & Tofthagen, C. 5-Fluorouracil toxicity and dihydropyrimidine dehydrogenase enzyme: implications for practice. *Clin. J. Oncol. Nurs.* **18**, 581–585 (2014).

80. Howard, S. C., McCormick, J., Pui, C.-H., Buddington, R. K. & Harvey, R. D. Preventing and Managing Toxicities of High-Dose Methotrexate. *Oncologist* **21**, 1471–1482 (2016).

81. Kumler, I., Brunner, N., Stenvang, J., Balslev, E. & Nielsen, D. L. A systematic review on topoisomerase 1 inhibition in the treatment of metastatic breast cancer. *Breast Cancer Res. Treat.* **138**, 347–358 (2013).

82. Burstein, H. J. *et al.* Adjuvant Endocrine Therapy for Women With Hormone Receptor–Positive Breast Cancer: ASCO Clinical Practice Guideline Focused Update. *J. Clin. Oncol.* **37**, 423–438 (2018).

83. (EBCTCG), E. B. C. T. C. G. Relevance of breast cancer hormone receptors and other factors to the efficacy of adjuvant tamoxifen: patient-level meta-analysis of randomised trials. *Lancet* **378**, 771–784 (2011).

84. Dowsett, M. *et al.* Integration of Clinical Variables for the Prediction of Late Distant Recurrence in Patients With Estrogen Receptor–Positive Breast Cancer Treated With 5 Years of Endocrine Therapy: CTS5. *J. Clin. Oncol.* **36**, 1941–1948 (2018).

85. Swaby, R. F., Sharma, C. G. N. & Jordan, V. C. SERMs for the treatment and prevention of breast cancer. *Rev. Endocr. Metab. Disord.* **8**, 229–239 (2007).

86. Early Breast Cancer Trialists' Collaborative Group (EBCTCG). Aromatase inhibitors versus tamoxifen in early breast cancer: patient-level meta-analysis of the randomised trials. *Lancet* **386**, 1341–1352 (2015).

87. Robertson, J. F. R. & Blamey, R. W. The use of gonadotrophin-releasing hormone (GnRH) agonists in early and advanced breast cancer in pre- and perimenopausal women. *Eur. J. Cancer* **39**, 861–869 (2003).

88. Francis, P. A. *et al.* Tailoring Adjuvant Endocrine Therapy for Premenopausal Breast Cancer. *N. Engl. J. Med.* **379**, 122–137 (2018).

89. Mehta, R. S. *et al.* Overall Survival with Fulvestrant plus Anastrozole in Metastatic Breast Cancer. *N. Engl. J. Med.* **380**, 1226–1234 (2019).

90. Weir, H. M. *et al.* AZD9496: An oral estrogen receptor inhibitor that blocks the growth of ER-positive and ESR1 mutant breast tumours in preclinical models. *Cancer Res.* canres.2357.2015 (2016). doi:10.1158/0008-5472.CAN-15-2357

91. Tria, G. S. *et al.* Discovery of LSZ102, a Potent, Orally Bioavailable Selective Estrogen Receptor Degrader (SERD) for the Treatment of Estrogen Receptor Positive Breast Cancer. *J. Med. Chem.* **61**, 2837–2864 (2018).

92. Joseph, J. D. *et al.* The selective estrogen receptor downregulator GDC-0810 is efficacious in diverse models of ER+ breast cancer. *Elife* **5**, e15828 (2016).

93. Piccart-Gebhart, M. J. *et al.* Trastuzumab after adjuvant chemotherapy in HER2-positive breast cancer. *N. Engl. J. Med.* **353**, 1659–1672 (2005).

94. Slamon, D. *et al.* Adjuvant Trastuzumab in HER2-Positive Breast Cancer. *N. Engl. J. Med.* **365**, 1273–1283 (2011).

95. von Minckwitz, G. *et al.* Adjuvant Pertuzumab and Trastuzumab in Early HER2-Positive Breast Cancer. *N. Engl. J. Med.* **377**, 122–131 (2017).

96. Geyer, C. E. *et al.* Lapatinib plus Capecitabine for HER2-Positive Advanced Breast Cancer. *N. Engl. J. Med.* **355**, 2733–2743 (2006).

97. Martin, M. *et al.* Neratinib after trastuzumab-based adjuvant therapy in HER2-positive breast cancer (ExteNET): 5-year analysis of a randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet. Oncol.* **18**, 1688–1700 (2017).

98. Finn, R. S. *et al.* Palbociclib and Letrozole in Advanced Breast Cancer. *N. Engl. J. Med.* **375**, 1925–1936 (2016).

99. Tripathy, D. *et al.* Ribociclib plus endocrine therapy for premenopausal women with hormone-receptor-positive, advanced breast cancer (MONALEESA-7): a randomised phase 3 trial. *Lancet. Oncol.* **19**, 904–915 (2018).

100. Johnston, S. R. D. *et al.* Abemaciclib Combined With Endocrine Therapy for the Adjuvant Treatment of HR+, HER2−, Node-Positive, High-Risk, Early Breast Cancer (monarchE). *J. Clin. Oncol.* JCO.20.02514 (2020). doi:10.1200/JCO.20.02514

101. Robson, M. *et al.* Olaparib for Metastatic Breast Cancer in Patients with a Germline BRCA Mutation. *N. Engl. J. Med.* **377**, 523–533 (2017).

102. Litton, J. K. *et al.* Talazoparib in Patients with Advanced Breast Cancer and a Germline BRCA Mutation. *N. Engl. J. Med.* **379**, 753–763 (2018).

103. André, F. *et al.* Alpelisib for PIK3CA-Mutated, Hormone Receptor–Positive Advanced Breast Cancer. *N. Engl. J. Med.* **380**, 1929–1940 (2019).

104. Baselga, J. *et al.* Everolimus in Postmenopausal Hormone-Receptor–Positive Advanced Breast Cancer. *N. Engl. J. Med.* **366**, 520–529 (2011).

105. Bachelot, T. *et al.* Randomized Phase II Trial of Everolimus in Combination With

Tamoxifen in Patients With Hormone Receptor–Positive, Human Epidermal Growth Factor Receptor 2–Negative Metastatic Breast Cancer With Prior Exposure to Aromatase Inhibitors: A GINECO Study. *J. Clin. Oncol.* **30**, 2718–2724 (2012).

106. Schmid, P. *et al.* Atezolizumab and Nab-Paclitaxel in Advanced Triple-Negative Breast Cancer. *N. Engl. J. Med.* **379**, 2108–2121 (2018).

107. Hanahan, D. & Weinberg, R. A. Hallmarks of Cancer: The Next Generation. *Cell* **144**, 646–674 (2011).

108. Massagué, J. & Obenauf, A. C. Metastatic colonization by circulating tumour cells. *Nature* (2016). doi:10.1038/nature17038

109. Micalizzi, D. S., Maheswaran, S. & Haber, D. A. A conduit to metastasis: circulating tumor cell biology. *Genes Dev.* **31**, 1827–1840 (2017).

110. Nagrath, S. *et al.* Isolation of rare circulating tumour cells in cancer patients by microchip technology. *Nature* **450**, 1235–1239 (2007).

111. Braun, S. *et al.* A Pooled Analysis of Bone Marrow Micrometastasis in Breast Cancer. *N. Engl. J. Med.* **353**, 793–802 (2005).

112. Kessenbrock, K., Plaks, V. & Werb, Z. Matrix Metalloproteinases: Regulators of the Tumor Microenvironment. *Cell* **141**, 52–67 (2010).

113. Wolf, K. *et al.* Multi-step pericellular proteolysis controls the transition from individual to collective cancer cell invasion. *Nat. Cell Biol.* **9**, 893–904 (2007).

114. Yilmaz, M. & Christofori, G. EMT, the cytoskeleton, and cancer cell invasion. *Cancer Metastasis Rev.* **28**, 15–33 (2009).

115. Giampieri, S. *et al.* Localized and reversible TGFbeta signalling switches breast cancer cells from cohesive to single cell motility. *Nat. Cell Biol.* **11**, 1287–1296 (2009).

116. Roh-Johnson, M. *et al.* Macrophage contact induces RhoA GTPase signaling to trigger tumor cell intravasation. *Oncogene* **33**, 4203–4212 (2014).

117. Magnon, C. *et al.* Autonomic nerve development contributes to prostate cancer progression. *Science* **341**, 1236361 (2013).

118. Friedl, P. & Gilmour, D. Collective cell migration in morphogenesis, regeneration and

cancer. *Nat. Rev. Mol. Cell Biol.* **10**, 445–457 (2009).

119. Condeelis, J. & Pollard, J. W. Macrophages: obligate partners for tumor cell migration, invasion, and metastasis. *Cell* **124**, 263–266 (2006).

120. Labernadie, A. *et al.* A mechanically active heterotypic E-cadherin/N-cadherin adhesion enables fibroblasts to drive cancer cell invasion. *Nat. Cell Biol.* **19**, 224–237 (2017).

121. Thiery, J. P., Acloque, H., Huang, R. Y. J. & Nieto, M. A. Epithelial-Mesenchymal Transitions in Development and Disease. *Cell* **139**, 871–890 (2009).

122. Dongre, A. & Weinberg, R. A. New insights into the mechanisms of epithelial–mesenchymal transition and implications for cancer. *Nat. Rev. Mol. Cell Biol.* **20**, 69–84 (2019).

123. Shamir, E. R. *et al.* Twist1-induced dissemination preserves epithelial identity and requires E-cadherin. *J. Cell Biol.* **204**, 839–856 (2014).

124. Fischer, K. R. *et al.* Epithelial-to-mesenchymal transition is not required for lung metastasis but contributes to chemoresistance. *Nature* **527**, 472–476 (2015).

125. Zheng, X. *et al.* Epithelial-to-mesenchymal transition is dispensable for metastasis but induces chemoresistance in pancreatic cancer. *Nature* **527**, 525–530 (2015).

126. Labelle, M., Begum, S. & Hynes, R. O. Direct signaling between platelets and cancer cells induces an epithelial-mesenchymal-like transition and promotes metastasis. *Cancer Cell* **20**, 576–590 (2011).

127. Franco, A. T., Corken, A. & Ware, J. Platelets at the interface of thrombosis, inflammation, and cancer. *Blood* **126**, 582–588 (2015).

128. Santos, M. F. *et al.* Comparative analysis of innate immune system function in metastatic breast, colorectal, and prostate cancer patients with circulating tumor cells. *Exp. Mol. Pathol.* **96**, 367–374 (2014).

129. Mohme, M., Riethdorf, S. & Pantel, K. Circulating and disseminated tumour cells — mechanisms of immune surveillance and escape. *Nat. Rev. Clin. Oncol.* **14**, 155–167 (2017).

130. Le Gal, K. *et al.* Antioxidants can increase melanoma metastasis in mice. *Sci. Transl. Med.*

7, 308re8 LP-308re8 (2015).

131. Piskounova, E. *et al.* Oxidative stress inhibits distant metastasis by human melanoma cells. *Nature* **527**, 186–191 (2015).

132. Zheng, Y. *et al.* Expression of β-globin by cancer cells promotes cell survival during blood-borne dissemination. *Nat. Commun.* **8**, 14344 (2017).

133. Aceto, N. *et al.* Circulating tumor cell clusters are oligoclonal precursors of breast cancer metastasis. *Cell* **158**, 1110–1122 (2014).

134. Kienast, Y. *et al.* Real-time imaging reveals the single steps of brain metastasis formation. *Nat. Med.* **16**, 116–122 (2010).

135. Läubli, H. & Borsig, L. Selectins promote tumor metastasis. *Semin. Cancer Biol.* **20**, 169–177 (2010).

136. Nguyen, D. X., Bos, P. D. & Massagué, J. Metastasis: from dissemination to organ-specific colonization. *Nat. Rev. Cancer* **9**, 274–84 (2009).

137. Al-Mehdi, A. B. *et al.* Intravascular origin of metastasis from the proliferation of endothelium-attached tumor cells: a new model for metastasis. *Nat. Med.* **6**, 100–102 (2000).

138. Tominaga, N. *et al.* Brain metastatic cancer cells release microRNA-181c-containing extracellular vesicles capable of destructing blood–brain barrier. *Nat. Commun.* **6**, 6716 (2015).

139. Zhou, W. *et al.* Cancer-Secreted miR-105 Destroys Vascular Endothelial Barriers to Promote Metastasis. *Cancer Cell* **25**, 501–515 (2014).

140. Burek, M. *et al.* Hypoxia-Induced MicroRNA-212/132 Alter Blood-Brain Barrier Integrity Through Inhibition of Tight Junction-Associated Proteins in Human and Mouse Brain Microvascular Endothelial Cells. *Transl. Stroke Res.* **10**, 672–683 (2019).

141. Weis, S., Cui, J., Barnes, L. & Cheresh, D. Endothelial barrier disruption by VEGF-mediated Src activity potentiates tumor cell extravasation and metastasis . *J. Cell Biol.* **167**, 223–229 (2004).

142. Tichet, M. *et al.* Tumour-derived SPARC drives vascular permeability and extravasation

through endothelial VCAM1 signalling to promote metastasis. *Nat. Commun.* **6**, 6993 (2015).

143. Padua, D. *et al.* TGFβ Primes Breast Tumors for Lung Metastasis Seeding through Angiopoietin-like 4. *Cell* **133**, 66–77 (2008).

144. Rahn, J. J. *et al.* MUC1 mediates transendothelial migration in vitro by ligating endothelial cell ICAM-1. *Clin. Exp. Metastasis* **22**, 475–483 (2005).

145. Draffin, J. E., McFarlane, S., Hill, A., Johnston, P. G. & Waugh, D. J. J. CD44 potentiates the adherence of metastatic prostate and breast cancer cells to bone marrow endothelial cells. *Cancer Res.* **64**, 5702–5711 (2004).

146. Zen, K. *et al.* CD44v4 is a major E-selectin ligand that mediates breast cancer cell transendothelial migration. *PLoS One* **3**, e1826 (2008).

147. Reymond, N., d'Água, B. B. & Ridley, A. J. Crossing the endothelial barrier during metastasis. *Nat. Rev. Cancer* **13**, 858–870 (2013).

148. Schumacher, D., Strilic, B., Sivaraj, K. K., Wettschureck, N. & Offermanns, S. Platelet-Derived Nucleotides Promote Tumor-Cell Transendothelial Migration and Metastasis via P2Y2 Receptor. *Cancer Cell* **24**, 130–137 (2013).

149. Qian, B.-Z. *et al.* CCL2 recruits inflammatory monocytes to facilitate breast-tumour metastasis. *Nature* **475**, 222–225 (2011).

150. Shiozawa, Y. *et al.* Human prostate cancer metastases target the hematopoietic stem cell niche to establish footholds in mouse bone marrow. *J. Clin. Invest.* **121**, 1298–1312 (2011).

151. Valiente, M. *et al.* Serpins promote cancer cell survival and vascular co-option in brain metastasis. *Cell* **156**, 1002–1016 (2014).

152. Wang, D., Sun, H., Wei, J., Cen, B. & DuBois, R. N. CXCL1 Is Critical for Premetastatic Niche Formation and Metastasis in Colorectal Cancer. *Cancer Res.* **77**, 3655–3665 (2017).

153. Urosevic, J. *et al.* ERK1/2 Signaling Induces Upregulation of ANGPT2 and CXCR4 to Mediate Liver Metastasis in Colon Cancer. *Cancer Res.* **80**, 4668 LP – 4680 (2020).

154. Peinado, H. *et al.* Melanoma exosomes educate bone marrow progenitor cells toward a

pro-metastatic phenotype through MET. *Nat. Med.* **18**, 883–891 (2012).

155. Costa-Silva, B. *et al.* Pancreatic cancer exosomes initiate pre-metastatic niche formation in the liver. *Nat. Cell Biol.* **17**, 816–826 (2015).

156. Malanchi, I. *et al.* Interactions between cancer stem cells and their niche govern metastatic colonization. *Nature* **481**, 85–89 (2012).

157. Oskarsson, T. *et al.* Breast cancer cells produce tenascin C as a metastatic niche component to colonize the lungs. *Nat. Med.* **17**, 867–874 (2011).

158. Gomis, R. R. & Gawrzak, S. Tumor cell dormancy. *Mol. Oncol.* **11**, 62–78 (2017).

159. Havas, K. M. *et al.* Metabolic shifts in residual breast cancer drive tumor recurrence. *J. Clin. Invest.* **127**, 2091–2105 (2017).

160. Conejo-Garcia, J. R. *et al.* Tumor-infiltrating dendritic cell precursors recruited by a beta-defensin contribute to vasculogenesis under the influence of Vegf-A. *Nat. Med.* **10**, 950–958 (2004).

161. Ghajar, C. M. *et al.* The perivascular niche regulates breast tumour dormancy. *Nat. Cell Biol.* **15**, 807–817 (2013).

162. Tsai, J. H., Donaher, J. L., Murphy, D. A., Chau, S. & Yang, J. Spatiotemporal Regulation of Epithelial-Mesenchymal Transition Is Essential for Squamous Cell Carcinoma Metastasis. *Cancer Cell* **22**, 725–736 (2012).

163. Massagué, J. & Obenauf, A. C. Metastatic colonization by circulating tumour cells. (2016). doi:10.1038/nature17038

164. Vanharanta, S. & Massague, J. Origins of metastatic traits. *Cancer Cell* **24**, 410–421 (2013).

165. Chatterjee, A., Rodger, E. J. & Eccles, M. R. Epigenetic drivers of tumourigenesis and cancer metastasis. *Semin. Cancer Biol.* **51**, 149–159 (2018).

166. Valastyan, S. & Weinberg, R. A. Tumor metastasis: molecular insights and evolving paradigms. *Cell* **147**, 275–292 (2011).

167. Hendrix, M. J. C. *et al.* Reprogramming metastatic tumour cells with embryonic microenvironments. *Nat. Rev. Cancer* **7**, 246–255 (2007).

168. Bockhorn, J. *et al.* Differentiation and loss of malignant character of spontaneous pulmonary metastases in patient-derived breast cancer models. *Cancer Res.* **74**, 7406–7417 (2014).

169. Patel, S. A., Rodrigues, P., Wesolowski, L. & Vanharanta, S. Genomic control of metastasis. *Br. J. Cancer* (2020). doi:10.1038/s41416-020-01127-6

170. Vanharanta, S. *et al.* Epigenetic expansion of VHL-HIF signal output drives multiorgan metastasis in renal cancer. *Nat. Med.* **19**, 50–56 (2013).

171. Quina, A. S., Buschbeck, M. & Di Croce, L. Chromatin structure and epigenetics. *Biochem. Pharmacol.* **72**, 1563–1569 (2006).

172. Cheng, Y. *et al.* Targeting epigenetic regulators for cancer therapy: mechanisms and advances in clinical trials. *Signal Transduct. Target. Ther.* **4**, 62 (2019).

173. Fang, F. *et al.* Breast cancer methylomes establish an epigenomic foundation for metastasis. *Sci. Transl. Med.* **3**, 75ra25-75ra25 (2011).

174. Carmona, F. J. *et al.* Epigenetic disruption of cadherin-11 in human cancer metastasis. *J. Pathol.* **228**, 230–240 (2012).

175. Byles, V. *et al.* SIRT1 induces EMT by cooperating with EMT transcription factors and enhances prostate cancer cell migration and metastasis. *Oncogene* **31**, 4619–4629 (2012).

176. Perillo, B. *et al.* DNA oxidation as triggered by H3K9me2 demethylation drives estrogen-induced gene expression. *Science* **319**, 202–206 (2008).

177. Metzger, E. *et al.* LSD1 demethylates repressive histone marks to promote androgen-receptor-dependent transcription. *Nature* **437**, 436–439 (2005).

178. Fang, Y., Liao, G. & Yu, B. LSD1/KDM1A inhibitors in clinical trials: advances and prospects. *J. Hematol. Oncol.* **12**, 129 (2019).

179. Chen, T. & Dent, S. Y. R. Chromatin modifiers and remodellers: regulators of cellular differentiation. *Nat. Rev. Genet.* **15**, 93–106 (2014).

180. Clapier, C. R., Iwasa, J., Cairns, B. R. & Peterson, C. L. Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. *Nat. Rev. Mol. Cell Biol.* **18**, 407–422 (2017).

181. Nair, S. S. & Kumar, R. Chromatin remodeling in cancer: a gateway to regulate gene transcription. *Mol. Oncol.* **6**, 611–619 (2012).

182. Xu, G. *et al.* ARID1A determines luminal identity and therapeutic response in estrogen-receptor-positive breast cancer. *Nat. Genet.* **52**, 198–207 (2020).

183. Tam, W. L. & Weinberg, R. A. The epigenetics of epithelial-mesenchymal plasticity in cancer. *Nat. Med.* **19**, 1438–1449 (2013).

184. Ganesh, K. *et al.* L1CAM defines the regenerative origin of metastasis-initiating cells in colorectal cancer. *Nat. Cancer* **1**, 28–45 (2020).

185. Esteller, M. Epigenetics in Cancer. *N. Engl. J. Med.* **358**, 1148–1159 (2008).

186. Lee, W. H. *et al.* Cytidine methylation of regulatory sequences near the pi-class glutathione S-transferase gene accompanies human prostatic carcinogenesis. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 11733–11737 (1994).

187. Cairns, P. *et al.* Molecular detection of prostate cancer in urine by GSTP1 hypermethylation. *Clin. Cancer Res.* **7**, 2727–2730 (2001).

188. Esteller, M. Cancer epigenomics: DNA methylomes and histone-modification maps. *Nat. Rev. Genet.* **8**, 286–298 (2007).

189. Esteller, M. *et al.* Inactivation of the DNA-Repair Gene MGMT and the Clinical Response of Gliomas to Alkylating Agents. *N. Engl. J. Med.* **343**, 1350–1354 (2000).

190. Hegi, M. E. *et al.* MGMT gene silencing and benefit from temozolomide in glioblastoma. *N. Engl. J. Med.* **352**, 997–1003 (2005).

191. Jones, P. A., Issa, J.-P. J. & Baylin, S. Targeting the cancer epigenome for therapy. *Nat. Rev. Genet.* **17**, 630–641 (2016).

192. Berdasco, M. & Esteller, M. Clinical epigenetics: seizing opportunities for translation. *Nat. Rev. Genet.* **20**, 109–127 (2019).

193. Obenauf, A. C. & Massagué, J. Surviving at a Distance: Organ-Specific Metastasis. *Trends in cancer* **1**, 76–91 (2015).

194. Paget, S. THE DISTRIBUTION OF SECONDARY GROWTHS IN CANCER OF THE BREAST. *Lancet* **133**, 571–573 (1889).

195. Kim, M.-Y. *et al.* Tumor self-seeding by circulating cancer cells. *Cell* **139**, 1315–1326 (2009).

196. Denève, E. *et al.* Capture of Viable Circulating Tumor Cells in the Liver of Colorectal Cancer Patients. *Clin. Chem.* **59**, 1384–1392 (2013).

197. Chen, W., Hoffmann, A. D., Liu, H. & Liu, X. Organotropism: new insights into molecular mechanisms of breast cancer metastasis. *npj Precis. Oncol.* **2**, 4 (2018).

198. Kennecke, H. *et al.* Metastatic Behavior of Breast Cancer Subtypes. *J. Clin. Oncol.* **28**, 3271–3277 (2010).

199. Savci-Heijink, C. D., Halfwerk, H., Koster, J. & van de Vijver, M. J. A novel gene expression signature for bone metastasis in breast carcinomas. *Breast Cancer Res. Treat.* **156**, 249–259 (2016).

200. Yerushalmi, R. *et al.* Patterns of relapse in breast cancer: changes over time. *Breast Cancer Res. Treat.* **120**, 753–759 (2010).

201. Yates, L. R. *et al.* Genomic Evolution of Breast Cancer Metastasis and Relapse. *Cancer Cell* **32**, 169-184.e7 (2017).

202. Minn, A. J. *et al.* Genes that mediate breast cancer metastasis to lung. *Nature* **436**, 518–524 (2005).

203. Morales, M. *et al.* RARRES3 suppresses breast cancer lung metastasis by regulating adhesion and differentiation. *EMBO Mol. Med.* **6**, 865–881 (2014).

204. Smid, M. *et al.* Genes Associated With Breast Cancer Metastatic to Bone. *J. Clin. Oncol.* **24**, 2261–2267 (2006).

205. Kang, Y. *et al.* A multigenic program mediating breast cancer metastasis to bone. *Cancer Cell* **3**, 537–549 (2003).

206. Roodman, G. D. Mechanisms of bone metastasis. *N. Engl. J. Med.* **350**, 1655–1664 (2004).

207. Macedo, F. *et al.* Bone Metastases: An Overview. *Oncol. Rev.* **11**, 321 (2017).

208. Coleman, R. E. Metastatic bone disease: clinical features, pathophysiology and treatment strategies. *Cancer Treat. Rev.* **27**, 165–176 (2001).

209. Fornetti, J., Welm, A. L. & Stewart, S. A. Understanding the Bone in Cancer Metastasis. *J. Bone Miner. Res.* **33**, 2099–2113 (2018).

210. Hadjidakis, D. J. & Androulakis, I. I. Bone remodeling. *Ann. N. Y. Acad. Sci.* **1092**, 385–396 (2006).

211. Iaquinta, M. R. *et al.* Adult Stem Cells for Bone Regeneration and Repair . *Frontiers in Cell and Developmental Biology* **7**, 268 (2019).

212. Clarke, B. Normal bone anatomy and physiology. *Clin. J. Am. Soc. Nephrol.* **3 Suppl 3**, S131–S139 (2008).

213. Florencio-Silva, R., Sasso, G. R. da S., Sasso-Cerri, E., Simoes, M. J. & Cerri, P. S. Biology of Bone Tissue: Structure, Function, and Factors That Influence Bone Cells. *Biomed Res. Int.* **2015**, 421746 (2015).

214. Tomkinson, A., Reeve, J., Shaw, R. W. & Noble, B. S. The death of osteocytes via apoptosis accompanies estrogen withdrawal in human bone. *J. Clin. Endocrinol. Metab.* **82**, 3128–3135 (1997).

215. Parfitt, A. M. The cellular basis of bone remodeling: The quantum concept reexamined in light of recent advances in the cell biology of bone. *Calcif. Tissue Int.* **36**, S37–S45 (1984).

216. Weilbaecher, K. N., Guise, T. A. & McCauley, L. K. Cancer to bone: a fatal attraction. *Nat. Rev. Cancer* **11**, 411–425 (2011).

217. Zaidi, M. Skeletal remodeling in health and disease. *Nat. Med.* **13**, 791–801 (2007).

218. Riggs, B. L. The mechanisms of estrogen regulation of bone resorption. *J. Clin. Invest.* **106**, 1203–1204 (2000).

219. Kelly, T. *et al.* Expression of heparanase by primary breast tumors promotes bone resorption in the absence of detectable bone metastases. *Cancer Res.* **65**, 5778–5784 (2005).

220. Liao, J., Schneider, A., Datta, N. S. & McCauley, L. K. Extracellular calcium as a candidate mediator of prostate cancer skeletal metastasis. *Cancer Res.* **66**, 9065–9073 (2006).

221. Saidak, Z. *et al.* Extracellular calcium promotes the migration of breast cancer cells through the activation of the calcium sensing receptor. *Exp. Cell Res.* **315**, 2072–2080 (2009).

222. McAllister, S. S. *et al.* Systemic endocrine instigation of indolent tumor growth requires osteopontin. *Cell* **133**, 994–1005 (2008).

223. Lynch, C. C. *et al.* MMP-7 promotes prostate cancer-induced osteolysis via the solubilization of RANKL. *Cancer Cell* **7**, 485–496 (2005).

224. Guise, T. A. *et al.* Evidence for a causal role of parathyroid hormone-related protein in the pathogenesis of human breast cancer-mediated osteolysis. *J. Clin. Invest.* **98**, 1544–1549 (1996).

225. Jones, D. H. *et al.* Regulation of cancer cell migration and bone metastasis by RANKL. *Nature* **440**, 692–696 (2006).

226. Russell, H. V, Hicks, J., Okcu, M. F. & Nuchtern, J. G. CXCR4 expression in neuroblastoma primary tumors is associated with clinical presentation of bone and bone marrow metastases. *J. Pediatr. Surg.* **39**, 1506–1511 (2004).

227. Muller, A. *et al.* Involvement of chemokine receptors in breast cancer metastasis. *Nature* **410**, 50–56 (2001).

228. Domanska, U. M. *et al.* CXCR4 inhibition with AMD3100 sensitizes prostate cancer to docetaxel chemotherapy. *Neoplasia* **14**, 709–718 (2012).

229. Sung, V., Stubbs, J. T. 3rd, Fisher, L., Aaron, A. D. & Thompson, E. W. Bone sialoprotein supports breast cancer cell adhesion proliferation and migration through differential usage of the alpha(v)beta3 and alpha(v)beta5 integrins. *J. Cell. Physiol.* **176**, 482–494 (1998).

230. Grzesik, W. J. & Robey, P. G. Bone matrix RGD glycoproteins: immunolocalization and interaction with human primary osteoblastic bone cells in vitro. *J. Bone Miner. Res.* **9**, 487–496 (1994).

231. Sanz-Rodriguez, F., Hidalgo, A. & Teixido, J. Chemokine stromal cell-derived factor-1alpha modulates VLA-4 integrin-mediated multiple myeloma cell adhesion to CS-1/fibronectin and VCAM-1. *Blood* **97**, 346–351 (2001).

232. Salvador, F., Llorente, A. & Gomis, R. R. From latency to overt bone metastasis in breast cancer: potential for treatment and prevention. *J. Pathol.* **249**, 6–18 (2019).

233. Johnson, R. W. *et al.* Induction of LIFR confers a dormancy phenotype in breast cancer cells disseminated to the bone marrow. *Nat. Cell Biol.* **18**, 1078–1089 (2016).

234. Bragado, P. *et al.* TGF-beta2 dictates disseminated tumour cell fate in target organs through TGF-beta-RIII and p38alpha/beta signalling. *Nat. Cell Biol.* **15**, 1351–1361 (2013).

235. Gao, H. *et al.* The BMP inhibitor Coco reactivates breast cancer cells at lung metastatic sites. *Cell* **150**, 764–779 (2012).

236. Gawrzak, S. *et al.* MSK1 regulates luminal cell differentiation and metastatic dormancy in ER+breast cancer. *Nat. Cell Biol.* (2018). doi:10.1038/s41556-017-0021-z

237. Malladi, S. *et al.* Metastatic Latency and Immune Evasion through Autocrine Inhibition of WNT. *Cell* **165**, 45–60 (2016).

238. Shiozawa, Y. *et al.* GAS6/AXL axis regulates prostate cancer invasion, proliferation, and survival in the bone marrow niche. *Neoplasia* **12**, 116–127 (2010).

239. Balz, L. M. *et al.* The interplay of HER2/HER3/PI3K and EGFR/HER2/PLC-gamma1 signalling in breast cancer cell migration and dissemination. *J. Pathol.* **227**, 234–244 (2012).

240. Schewe, D. M. & Aguirre-Ghiso, J. A. ATF6alpha-Rheb-mTOR signaling promotes survival of dormant tumor cells in vivo. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 10519–10524 (2008).

241. Lu, X. *et al.* VCAM-1 promotes osteolytic expansion of indolent bone micrometastasis of breast cancer by engaging alpha4beta1-positive osteoclast progenitors. *Cancer Cell* **20**, 701–714 (2011).

242. Lu, X. *et al.* ADAMTS1 and MMP1 proteolytically engage EGF-like ligands in an osteolytic signaling cascade for bone metastasis. *Genes Dev.* **23**, 1882–1894 (2009).

243. Coleman, R. E. Prevention and treatment of bone metastases. *Nat. Rev. Clin. Oncol.* **9**, 76–78 (2012).

244. Sterling, J. A. *et al.* The hedgehog signaling molecule Gli2 induces parathyroid hormone-related peptide expression and osteolysis in metastatic human breast cancer cells. *Cancer Res.* **66**, 7548–7553 (2006).

245. Pratap, J. *et al.* Ectopic runx2 expression in mammary epithelial cells disrupts formation of normal acini structure: implications for breast cancer progression. *Cancer Res.* **69**, 6807–6814 (2009).

246. Yin, J. J. *et al.* TGF-beta signaling blockade inhibits PTHrP secretion by breast cancer cells and bone metastases development. *J. Clin. Invest.* **103**, 197–206 (1999).

247. Sanders, J. L. *et al.* Extracellular calcium-sensing receptor expression and its potential role in regulating parathyroid hormone-related peptide secretion in human breast cancer cell lines. *Endocrinology* **141**, 4357–4364 (2000).

248. Yano, S. *et al.* Calcium-sensing receptor activation stimulates parathyroid hormone-related protein secretion in prostate cancer cells: role of epidermal growth factor receptor transactivation. *Bone* **35**, 664–672 (2004).

249. Sethi, N., Dai, X., Winter, C. G. & Kang, Y. Tumor-derived JAGGED1 promotes osteolytic bone metastasis of breast cancer by engaging notch signaling in bone cells. *Cancer Cell* **19**, 192–205 (2011).

250. Hiraga, T., Kizaka-Kondoh, S., Hirota, K., Hiraoka, M. & Yoneda, T. Hypoxia and hypoxia-inducible factor-1 expression enhance osteolytic bone metastases of breast cancer. *Cancer Res.* **67**, 4157–4163 (2007).

251. Dunn, L. K. *et al.* Hypoxia and TGF-beta drive breast cancer bone metastases through parallel signaling pathways in tumor cells and the bone microenvironment. *PLoS One* **4**, e6896 (2009).

252. Zheng, H. *et al.* Therapeutic Antibody Targeting Tumor- and Osteoblastic Niche-Derived Jagged1 Sensitizes Bone Metastasis to Chemotherapy. *Cancer Cell* **32**, 731-747.e6 (2017).

253. Boucharaba, A. *et al.* Platelet-derived lysophosphatidic acid supports the progression of osteolytic bone metastases in breast cancer. *J. Clin. Invest.* **114**, 1714–1725 (2004).

254. Yi, B., Williams, P. J., Niewolna, M., Wang, Y. & Yoneda, T. Tumor-derived platelet-

derived growth factor-BB plays a critical role in osteosclerotic bone metastasis in an animal model of human breast cancer. *Cancer Res.* **62**, 917–923 (2002).

255. Achbarou, A. *et al.* Urokinase overproduction results in increased skeletal metastasis by prostate cancer cells in vivo. *Cancer Res.* **54**, 2372–2377 (1994).

256. Nelson, J. B. *et al.* Identification of endothelin-1 in the pathophysiology of metastatic adenocarcinoma of the prostate. *Nat. Med.* **1**, 944–949 (1995).

257. Yin, J. J. *et al.* A causal role for endothelin-1 in the pathogenesis of osteoblastic bone metastases. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 10954–10959 (2003).

258. Iwamura, M., Hellman, J., Cockett, A. T., Lilja, H. & Gershagen, S. Alteration of the hormonal bioactivity of parathyroid hormone-related protein (PTHrP) as a result of limited proteolysis by prostate-specific antigen. *Urology* **48**, 317–325 (1996).

259. Coleman, R. E. Clinical features of metastatic bone disease and risk of skeletal morbidity. *Clin. Cancer Res.* **12**, 6243s-6249s (2006).

260. De Felice, F., Piccioli, A., Musio, D. & Tombolini, V. The role of radiation therapy in bone metastases management. *Oncotarget; Vol 8, No 15* (2017).

261. D'Oronzo, S., Coleman, R., Brown, J. & Silvestris, F. Metastatic bone disease: Pathogenesis and therapeutic options: Up-date on bone metastasis management. *J. bone Oncol.* **15**, 4 (2019).

262. Anselmetti, G. C. Osteoplasty: Percutaneous Bone Cement Injection beyond the Spine. *Semin. Intervent. Radiol.* **27**, 199–208 (2010).

263. Kalyan, S., Chandrasekaran, V., Quabius, E. S., Lindhorst, T. K. & Kabelitz, D. Neutrophil uptake of nitrogen-bisphosphonates leads to the suppression of human peripheral blood γδ T cells. *Cell. Mol. Life Sci.* **71**, 2335–2346 (2014).

264. Cabillic, F. *et al.* Aminobisphosphonate-pretreated dendritic cells trigger successful Vγ9Vδ2 T cell amplification for immunotherapy in advanced cancer patients. *Cancer Immunol. Immunother.* **59**, 1611–1619 (2010).

265. Junankar, S. *et al.* Real-Time Intravital Imaging Establishes Tumor-Associated Macrophages as the Extraskeletal Target of Bisphosphonate Action in Cancer. *Cancer Discov.* **5**, 35 LP – 42 (2015).

266. Dedes, P. G. *et al.* Expression of matrix macromolecules and functional properties of breast cancer cells are modulated by the bisphosphonate zoledronic acid. *Biochim. Biophys. Acta - Gen. Subj.* **1820**, 1926–1939 (2012).

267. Gralow, J. R. *et al.* Phase III Randomized Trial of Bisphosphonates as Adjuvant Therapy in Breast Cancer: S0307. *JNCI J. Natl. Cancer Inst.* **112**, 698–707 (2020).

268. Gnant, M. *et al.* Adjuvant endocrine therapy plus zoledronic acid in premenopausal women with early-stage breast cancer: 62-month follow-up from the ABCSG-12 randomised trial. *Lancet. Oncol.* **12**, 631–641 (2011).

269. Gnant, M. *et al.* Overall survival with adjuvant zoledronic acid in patients with premenopausal breast cancer with complete endocrine blockade: Long-term results from ABCSG-12. *J. Clin. Oncol.* **29**, 520 (2011).

270. Coleman, R. *et al.* Adjuvant zoledronic acid in patients with early breast cancer: final efficacy analysis of the AZURE (BIG 01/04) randomised open-label phase 3 trial. *Lancet Oncol.* **15**, 997–1006 (2014).

271. Fizazi, K. *et al.* Denosumab versus zoledronic acid for treatment of bone metastases in men with castration-resistant prostate cancer: a randomised, double-blind study. *Lancet (London, England)* **377**, 813–822 (2011).

272. Stopeck, A. T. *et al.* Denosumab compared with zoledronic acid for the treatment of bone metastases in patients with advanced breast cancer: a randomized, double-blind study. *J. Clin. Oncol.* **28**, 5132–5139 (2010).

273. Jensen, A. B. *et al.* The cathepsin K inhibitor odanacatib suppresses bone resorption in women with breast cancer and established bone metastases: results of a 4-week, double-blind, randomized, controlled trial. *Clin. Breast Cancer* **10**, 452–458 (2010).

274. Bone, H. G. *et al.* Odanacatib for the treatment of postmenopausal osteoporosis: development history and design and participant characteristics of LOFT, the Long-Term Odanacatib Fracture Trial. *Osteoporos. Int.* **26**, 699–712 (2015).

275. Rucci, N., Susa, M. & Teti, A. Inhibition of protein kinase c-Src as a therapeutic approach for cancer and bone metastases. *Anticancer. Agents Med. Chem.* **8**, 342–349 (2008).

276. Campone, M. *et al.* Phase II study of single-agent bosutinib, a Src/Abl tyrosine kinase inhibitor, in patients with locally advanced or metastatic breast cancer pretreated with chemotherapy. *Ann. Oncol.* **23**, 610–617 (2012).

277. Antonarakis, E. S. *et al.* A phase 2 study of KX2-391, an oral inhibitor of Src kinase and tubulin polymerization, in men with bone-metastatic castration-resistant prostate cancer. *Cancer Chemother. Pharmacol.* **71**, 883–892 (2013).

278. Schott, A. F. *et al.* Phase II studies of two different schedules of dasatinib in bone metastasis predominant metastatic breast cancer: SWOG S0622. *Breast Cancer Res. Treat.* **159**, 87–95 (2016).

279. Broom, R. J. *et al.* Everolimus and Zoledronic Acid in Patients With Renal Cell Carcinoma With Bone Metastases: A Randomized First-Line Phase II Trial. *Clin. Genitourin. Cancer* **13**, 50–58 (2015).

280. Yu, Y. *et al.* Everolimus and zoledronic acid—a potential synergistic treatment for lung adenocarcinoma bone metastasis. *Acta Biochim. Biophys. Sin. (Shanghai).* **46**, 792–801 (2014).

281. Gnant, M. *et al.* Effect of Everolimus on Bone Marker Levels and Progressive Disease in Bone in BOLERO-2. *JNCI J. Natl. Cancer Inst.* **105**, 654–663 (2013).

282. Durie, B. G. M. *et al.* Bortezomib with lenalidomide and dexamethasone versus lenalidomide and dexamethasone alone in patients with newly diagnosed myeloma without intent for immediate autologous stem-cell transplant (SWOG S0777): a randomised, open-label, phase 3 trial. *Lancet* **389**, 519–527 (2017).

283. Dimopoulos, M. A. *et al.* Carfilzomib and dexamethasone versus bortezomib and dexamethasone for patients with relapsed or refractory multiple myeloma (ENDEAVOR): a randomised, phase 3, open-label, multicentre study. *Lancet Oncol.* **17**, 27–38 (2016).

284. Pennisi, A. *et al.* Consequences of Daily Administered Parathyroid Hormone on Myeloma Growth, Bone Disease, and Molecular Profiling of Whole Myelomatous Bone. *PLoS One* **5**, e15233 (2010).

285. Kang, M.-G. *et al.* Serum Parathyroid Hormone Is a New Potential Risk Factor in

Multiple Myeloma. *Biomed Res. Int.* **2014**, 804182 (2014).

286. Recker, R. R. *et al.* A Randomized, Double-Blind Phase 2 Clinical Trial of Blosozumab, a Sclerostin Antibody, in Postmenopausal Women with Low Bone Mineral Density. *J. Bone Miner. Res.* **30**, 216–224 (2015).

287. Cosman, F. *et al.* Romosozumab Treatment in Postmenopausal Women with Osteoporosis. *N. Engl. J. Med.* **375**, 1532–1543 (2016).

288. Iyer, S. P. *et al.* A Phase IB multicentre dose-determination study of BHQ880 in combination with anti-myeloma therapy and zoledronic acid in patients with relapsed or refractory multiple myeloma and prior skeletal-related events. *Br. J. Haematol.* **167**, 366–375 (2014).

289. Munshi, N. C. *et al.* Early Evidence of Anabolic Bone Activity of BHQ880, a Fully Human Anti-DKK1 Neutralizing Antibody: Results of a Phase 2 Study in Previously Untreated Patients with Smoldering Multiple Myeloma At Risk for Progression. *Blood* **120**, 331 (2012).

290. Abdulkadyrov, K. M. *et al.* Sotatercept in patients with osteolytic lesions of multiple myeloma. *Br. J. Haematol.* **165**, 814–823 (2014).

291. Qiao, L. *et al.* Endothelin-A receptor antagonists in prostate cancer treatment-a meta-analysis. *Int. J. Clin. Exp. Med.* **8**, 3465–3473 (2015).

292. Smith, M. *et al.* Phase III Study of Cabozantinib in Previously Treated Metastatic Castration-Resistant Prostate Cancer: COMET-1. *J. Clin. Oncol.* **34**, 3005–3013 (2016).

293. Choueiri, T. K. *et al.* Cabozantinib versus everolimus in advanced renal cell carcinoma (METEOR): final results from a randomised, open-label, phase 3 trial. *Lancet Oncol.* **17**, 917–927 (2016).

294. Parker, C. *et al.* Overall Survival Benefit of Radium-223 Chloride (Alpharadin™) in the Treatment of Patients with Symptomatic Bone Metastases in Castration-resistant Prostate Cancer (CRPC): a Phase III Randomized Trial (ALSYMPCA). *Eur. J. Cancer* **47**, 3 (2011).

295. Lipton, A. *et al.* Changes in Bone Turnover Marker Levels and Clinical Outcomes in Patients with Advanced Cancer and Bone Metastases Treated with Bone Antiresorptive

Agents. *Clin. Cancer Res.* **22**, 5713–5721 (2016).

296. Coleman, R. E. *et al.* Predictive value of bone resorption and formation markers in cancer patients with bone metastases receiving the bisphosphonate zoledronic acid. *J. Clin. Oncol.* **23**, 4925–4935 (2005).

297. Brown, J. *et al.* Associations Between Serum Bone Biomarkers in Early Breast Cancer and Development of Bone Metastasis: Results From the AZURE (BIG01/04) Trial. *J. Natl. Cancer Inst.* **110**, 871–879 (2018).

298. Westbrook, J. A. *et al.* Identification and validation of DOCK4 as a potential biomarker for risk of bone metastasis development in patients with early breast cancer. *J. Pathol.* **247**, 381–391 (2019).

299. Rafiei, S., Tiedemann, K., Tabaries, S., Siegel, P. M. & Komarova, S. V. Peroxiredoxin 4: a novel secreted mediator of cancer induced osteoclastogenesis. *Cancer Lett.* **361**, 262–270 (2015).

300. Li, Y. *et al.* A mandatory role of nuclear PAK4-LIFR axis in breast-to-bone metastasis of ERalpha-positive breast cancer cells. *Oncogene* **38**, 808–821 (2019).

301. Coleman, R. E. *et al.* Benefits and risks of adjuvant treatment with zoledronic acid in stage II/III breast cancer. 10 years follow-up of the AZURE randomized clinical trial (BIG 01/04). *J. Bone Oncol.* **13**, 123–135 (2018).

302. Westbrook, J. A. *et al.* CAPG and GIPC1: Breast Cancer Biomarkers for Bone Metastasis Development and Treatment. *Journal of the National Cancer Institute* **108**, (2016).

303. Pavlovic, M. *et al.* Enhanced MAF Oncogene Expression and Breast Cancer Bone Metastasis. *J. Natl. Cancer Inst.* **107**, djv256 (2015).

304. Coleman, R. *et al.* Effect of MAF amplification on treatment outcomes with adjuvant zoledronic acid in early breast cancer : a secondary analysis of the international , open-label , randomised , controlled , phase 3 AZURE (BIG 01 / 04) trial. 1543–1552 (2017). doi:10.1016/S1470-2045(17)30603-4

305. Eychene, A., Rocques, N. & Pouponnot, C. A new MAFia in cancer. *Nat. Rev. Cancer* **8**, 683–693 (2008).

306. Kataoka, K. Multiple mechanisms and functions of maf transcription factors in the

regulation of tissue-specific genes. *J. Biochem.* **141**, 775–781 (2007).

307. Newman, J. R. S. & Keating, A. E. Comprehensive identification of human bZIP interactions with coiled-coil arrays. *Science* **300**, 2097–2101 (2003).

308. Kerppola, T. K. & Curran, T. Maf and Nrl can bind to AP-1 sites and form heterodimers with Fos and Jun. *Oncogene* **9**, 675–684 (1994).

309. Kataoka, K., Fujiwara, K. T., Noda, M. & Nishizawa, M. MafB, a new Maf family transcription activator that can associate with Maf and Fos but not with Jun. *Mol. Cell. Biol.* **14**, 7581–7591 (1994).

310. Rocques, N. *et al.* GSK-3-Mediated Phosphorylation Enhances Maf-Transforming Activity. *Mol. Cell* **28**, 584–597 (2007).

311. Tillmanns, S. *et al.* SUMO modification regulates MafB-driven macrophage differentiation by enabling Myb-dependent transcriptional repression. *Mol. Cell. Biol.* **27**, 5554–5564 (2007).

312. Blank, V. Small Maf proteins in mammalian gene control: mere dimerization partners or dynamic transcriptional regulators? *J. Mol. Biol.* **376**, 913–925 (2008).

313. Kataoka, K., Noda, M. & Nishizawa, M. Transactivation activity of Maf nuclear oncoprotein is modulated by Jun, Fos and small Maf proteins. *Oncogene* **12**, 53–62 (1996).

314. Motohashi, H., Katsuoka, F., Shavit, J. A., Engel, J. D. & Yamamoto, M. Positive or negative MARE-dependent transcriptional regulation is determined by the abundance of small Maf proteins. *Cell* **103**, 865–875 (2000).

315. Yang, Y. & Cvekl, A. Large Maf Transcription Factors: Cousins of AP-1 Proteins and Important Regulators of Cellular Differentiation. *Einstein J. Biol. Med.* **23**, 2–11 (2007).

316. Kannan, M. B., Solovieva, V. & Blank, V. The small MAF transcription factors MAFF, MAFG and MAFK: current knowledge and perspectives. *Biochim. Biophys. Acta* **1823**, 1841–1846 (2012).

317. Nishizawa, M., Kataoka, K., Goto, N., Fujiwara, K. T. & Kawai, S. v-maf, a viral oncogene that encodes a 'leucine zipper' motif. *Proc. Natl. Acad. Sci. U. S. A.* **86**, 7711–7715 (1989).

318. Kawai, S. *et al.* Isolation of the avian transforming retrovirus, AS42, carrying the v-maf oncogene and initial characterization of its gene product. *Virology* **188**, 778–784 (1992).

319. Morito, N. *et al.* Overexpression of c-Maf contributes to T-cell lymphoma in both mice and human. *Cancer Res.* **66**, 812–819 (2006).

320. Murakami, Y. I. *et al.* c-Maf expression in angioimmunoblastic T-cell lymphoma. *Am. J. Surg. Pathol.* **31**, 1695–1702 (2007).

321. Hurt, E. M. *et al.* Overexpression of c-maf is a frequent oncogenic event in multiple myeloma that promotes proliferation and pathological interactions with bone marrow stroma. *Cancer Cell* **5**, 191–199 (2004).

322. Chng, W. J., Glebov, O., Bergsagel, P. L. & Kuehl, W. M. Genetic events in the pathogenesis of multiple myeloma. *Best Pract. Res. Clin. Haematol.* **20**, 571–596 (2007).

323. Boersma-Vreugdenhil, G. R. *et al.* The recurrent translocation t(14;20)(q32;q12) in multiple myeloma results in aberrant expression of MAFB: a molecular and genetic analysis of the chromosomal breakpoint. *Br. J. Haematol.* **126**, 355–363 (2004).

324. Nishizawa, M., Kataoka, K. & Vogt, P. K. MafA has strong cell transforming ability but is a weak transactivator. *Oncogene* **22**, 7882–7890 (2003).

325. Kataoka, K., Nishizawa, M. & Kawai, S. Structure-function analysis of the maf oncogene product, a member of the b-Zip protein family. *J. Virol.* **67**, 2133–2141 (1993).

326. Pouponnot, C. *et al.* Cell context reveals a dual role for Maf in oncogenesis. *Oncogene* **25**, 1299–1310 (2006).

327. Mattioli, M. *et al.* Gene expression profiling of plasma cell dyscrasias reveals molecular patterns associated with distinct IGH translocations in multiple myeloma. *Oncogene* **24**, 2461–2473 (2005).

328. Zhan, F. *et al.* The molecular classification of multiple myeloma. *Blood* **108**, 2020–2028 (2006).

329. Suzuki, A. *et al.* ARK5 is transcriptionally regulated by the Large-MAF family and mediates IGF-1-induced cell invasion in multiple myeloma: ARK5 as a new molecular determinant of malignant multiple myeloma. *Oncogene* **24**, 6936–6944 (2005).

330. Takagaki, K. *et al.* Parathyroid hormone-related protein expression, in combination with nodal status, predicts bone metastasis and prognosis of breast cancer patients. *Exp. Ther. Med.* **3**, 963–968 (2012).

331. Henderson, M. A. *et al.* Parathyroid Hormone–Related Protein Localization in Breast Cancers Predict Improved Prognosis. *Cancer Res.* **66**, 2250 LP – 2256 (2006).

332. Early Breast Cancer Trialists' Collaborative Group (EBCTCG). Adjuvant bisphosphonate treatment in early breast cancer: meta-analyses of individual patient data from randomised trials. *Lancet* **386**, 1353–1361 (2015).

333. Dondossola, E. *et al.* Intravital microscopy of osteolytic progression and therapy response of cancer lesions in the bone. *Sci. Transl. Med.* **10**, eaao5726 (2018).

334. Coleman, R. *et al.* Effect of MAF amplification on treatment outcomes with adjuvant zoledronic acid in early breast cancer: a secondary analysis of the international, open-label, randomised, controlled, phase 3 AZURE (BIG 01/04) trial. *Lancet Oncol.* **18**, 1543–1552 (2017).

335. Zhang, C. *et al.* MafA is a key regulator of glucose-stimulated insulin secretion. *Mol. Cell. Biol.* **25**, 4969–4976 (2005).

336. Li, M. A., Alls, J. D., Avancini, R. M., Koo, K. & Godt, D. The large Maf factor Traffic Jam controls gonad morphogenesis in Drosophila. *Nat. Cell Biol.* **5**, 994–1000 (2003).

337. Lam, K. S., Liu, R., Miyamoto, S., Lehman, A. L. & Tuscano, J. M. Applications of One-Bead One-Compound Combinatorial Libraries and Chemical Microarrays in Signal Transduction Research. *Acc. Chem. Res.* **36**, 370–377 (2003).

338. Meldal, M., Svendsen, I., Breddam, K. & Auzanneau, F. I. Portion-mixing peptide libraries of quenched fluorogenic substrates for complete subsite mapping of endoprotease specificity. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 3314–3318 (1994).

339. Aina, O. H., Marik, J., Liu, R., Lau, D. H. & Lam, K. S. Identification of novel targeting peptides for human ovarian cancer cells using "' one-bead one-compound '" combinatorial libraries. 806–813 (2005).

340. Yang, F. *et al.* Discovery of specific ligands for oral squamous carcinoma to develop anti-cancer drug loaded precise targeting nanotherapeutics. *J. Calif. Dent. Assoc.* **40**, 939–

943 (2012).

341. Park, S. I. *et al.* The use of one-bead one-compound combinatorial library method to identify peptide ligands for α4β1 integrin receptor in non-Hodgkin's lymphoma. *Lett. Pept. Sci.* **8**, 171–178 (2001).

342. Cho, C.-F., Behnam Azad, B., Luyt, L. G. & Lewis, J. D. High-throughput screening of one-bead-one-compound peptide libraries using intact cells. *ACS Comb. Sci.* **15**, 393–400 (2013).

343. Lam, K. S. *et al.* A new type of synthetic peptide library for identifying ligand-binding activity. *Nature* **354**, 82–84 (1991).

344. Yao, N. *et al.* Discovery of targeting ligands for breast cancer cells using the one-bead one-compound combinatorial method. *J. Med. Chem.* **52**, 126–133 (2009).

345. Chan, W. & White, P. D. *Fmoc Solid-Phase Peptide Synthesis: A Practical Approach. New York: Oxford University Press* **222**, (2000).

346. Ruoslahti, E. RGD AND OTHER RECOGNITION SEQUENCES FOR INTEGRINS. *Annu. Rev. Cell Dev. Biol.* **12**, 697–715 (1996).

347. Nomizu, M. *et al.* Structure-activity study of a laminin α1 chain active peptide segment Ile-Lys-Val-Ala-Val (IKVAV). *FEBS Lett.* **365**, 227–231 (1995).

348. Agius, E., Sagot, Y., Duprat, A. M. & Cochard, P. Antibodies directed against the β1-integrin subunit and peptides containing the IKVAV sequence of laminin perturb neurite outgrowth of peripheral neurons on immature spinal cord substrata. *Neuroscience* **71**, 773–786 (1996).

349. Murayama, O., Nishida, H. & Sekiguchi, K. Novel Peptide Ligands for Integrin α6β1 Selected from a Phage Display Library1. *J. Biochem.* **120**, 445–451 (1996).

350. Zhao, P.-L., Nachbar, R. B., Bolognese, J. A. & Chapman, K. Two New Criteria for Choosing Sample Size in Combinatorial Chemistry. *J. Med. Chem.* **39**, 350–352 (1996).

351. Guixer, B. *et al.* Chemically synthesized peptide libraries as a new source of BBB shuttles. Use of mass spectrometry for peptide identification. *J. Pept. Sci.* **22**, 577–591 (2016).

352. Kalafatovic, D., Mauša, G., Todorovski, T. & Giralt, E. Algorithm-supported, mass and sequence diversity-oriented random peptide library design. *J. Cheminform.* **11**, 25 (2019).

353. Kalafatovic, D., Mausa, G., Rešetar Maslov, D. & Giralt, E. Bottom-Up Design Approach for OBOC Peptide Libraries. *Molecules* **25**, 3316 (2020).

354. Bogan, A. A. & Thorn, K. S. Anatomy of hot spots in protein interfaces11Edited by J. Wells. *J. Mol. Biol.* **280**, 1–9 (1998).

355. Chen, Q., Dowhan, D. H., Liang, D., Moore, D. D. & Overbeek, P. A. CREB-binding protein/p300 co-activation of crystallin gene expression. *J. Biol. Chem.* **277**, 24081–24089 (2002).

356. Harbeck, N. *et al.* Breast cancer. *Nat. Rev. Dis. Prim.* **5**, 66 (2019).

357. Yaşar, P., Ayaz, G., User, S. D., Güpür, G. & Muyan, M. Molecular mechanism of estrogen-estrogen receptor signaling. *Reprod. Med. Biol.* **16**, 4–20 (2016).

358. Dahlman-Wright, K. *et al.* Interplay between AP-1 and estrogen receptor α in regulating gene expression and proliferation networks in breast cancer cells. *Carcinogenesis* **33**, 1684–1691 (2012).

359. Heldring, N. *et al.* Multiple sequence-specific DNA-binding proteins mediate estrogen receptor signaling through a tethering pathway. *Mol. Endocrinol.* **25**, 564–574 (2011).

360. Imbratta, C., Hussein, H., Andris, F. & Verdeil, G. c-MAF, a Swiss Army Knife for Tolerance in Lymphocytes . *Frontiers in Immunology* **11**, 206 (2020).

361. Rangel, N., Villegas, V. E. & Rondón-Lagos, M. Profiling of gene expression regulated by 17β-estradiol and tamoxifen in estrogen receptor-positive and estrogen receptor-negative human breast cancer cell lines. *Breast cancer (Dove Med. Press.* **9**, 537–550 (2017).

362. Song, N. *et al.* FGF18 Enhances Migration and the Epithelial-Mesenchymal Transition in Breast Cancer by Regulating Akt/GSK3β/B-Catenin Signaling. *Cell. Physiol. Biochem.* **49**, 1060–1073 (2018).

363. Xue, Y. *et al.* SOX9/FXYD3/Src Axis Is Critical for ER⁺ Breast Cancer Stem Cell Function. *Mol. Cancer Res.* **17**, 238 LP – 249 (2019).

364. Le Dily, F. *et al.* Distinct structural transitions of chromatin topological domains

correlate with coordinated hormone-induced gene regulation. *Genes Dev.* **28**, 2151–2162 (2014).

365. Creyghton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci.* **107**, 21931 LP – 21936 (2010).

366. Rhie, S. K. *et al.* Identification of activated enhancers and linked transcription factors in breast, prostate, and kidney tumors by tracing enhancer networks using epigenetic traits. *Epigenetics Chromatin* **9**, 50 (2016).

367. Carroll, J. S. *et al.* Genome-wide analysis of estrogen receptor binding sites. *Nat. Genet.* **38**, 1289–1297 (2006).

368. Zwart, W. *et al.* Oestrogen receptor-co-factor-chromatin specificity in the transcriptional regulation of breast cancer. *EMBO J.* **30**, 4764–4776 (2011).

369. Chen, A. & Koehler, A. N. Transcription Factor Inhibition: Lessons Learned and Emerging Targets. *Trends Mol. Med.* **26**, 508–518 (2020).

370. Yin, H. & Flynn, A. D. Drugging Membrane Protein Interactions. *Annu. Rev. Biomed. Eng.* **18**, 51–76 (2016).

371. Matharage, J. M., Minna, J. D., Brekken, R. A. & Udugamasooriya, D. G. Unbiased Selection of Peptide-Peptoid Hybrids Specific for Lung Cancer Compared to Normal Lung Epithelial Cells. *ACS Chem. Biol.* **10**, 2891–2899 (2015).

372. Yang, F. *et al.* One-bead one-compound combinatorial library derived targeting ligands for detection and treatment of oral squamous cancer. *Oncotarget; Vol 10, No 52* (2019).

373. Liu, R., Li, X., Xiao, W. & Lam, K. S. Tumor-targeting peptides from combinatorial libraries. *Adv. Drug Deliv. Rev.* **110–111**, 13–37 (2017).

374. Xiao, K. *et al.* "OA02" Peptide Facilitates the Precise Targeting of Paclitaxel-Loaded Micellar Nanoparticles to Ovarian Cancer In Vivo *Cancer Res.* **72**, 2100 LP – 2110 (2012).

375. Udugamasooriya, D. G. & Kodadek, T. On-Bead Two-Color (OBTC) Cell Screen for Direct Identification of Highly Selective Cell Surface Receptor Ligands. *Curr. Protoc. Chem. Biol.* **4**, 35–48 (2012).

376. Ho, I.-C., Hodge, M. R., Rooney, J. W. & Glimcher, L. H. The Proto-Oncogene c-maf Is Responsible for Tissue-Specific Expression of Interleukin-4. *Cell* **85**, 973–983 (1996).

377. Hedge, S. P., Kumar, A., Kurschner, C. & Shapiro, L. H. c-Maf interacts with c-Myb to regulate transcription of an early myeloid gene during differentiation. *Mol. Cell. Biol.* **18**, 2729–2737 (1998).

378. Roux, K. J., Kim, D. I. & Burke, B. BioID: A screen for protein-protein interactions. *Curr. Protoc. Protein Sci.* (2013). doi:10.1002/0471140864.ps1923s74

379. In, D., Jensen, S. C., Noble, K. A., Kc, B. & Roux, K. H. An improved smaller biotin ligase for BioID proximity labeling. (2016). doi:10.1091/mbc.E15-12-0844

380. Ramanathan, M. *et al.* RNA–protein interaction detection in living cells. *Nat. Methods* **15**, 207–212 (2018).

381. Kido, K. *et al.* AirID, a novel proximity biotinylation enzyme, for analysis of protein-protein interactions. *Elife* **9**, (2020).

382. Branon, T. C. *et al.* Efficient proximity labeling in living cells and organisms with TurboID. *Nat. Biotechnol.* **36**, 880–887 (2018).

383. Cho, K. F. *et al.* Proximity labeling in mammalian cells with TurboID and split-TurboID. *Nat. Protoc.* **15**, 3971–3999 (2020).

384. Carroll, J. S. EJE PRIZE 2016: Mechanisms of oestrogen receptor (ER) gene regulation in breast cancer. *Eur. J. Endocrinol.* **175**, R41–R49

385. Miranda, T. B. *et al.* Reprogramming the chromatin landscape: interplay of the estrogen and glucocorticoid receptors at the genomic level. *Cancer Res.* **73**, 5130–5139 (2013).

386. Green, K. A. & Carroll, J. S. Oestrogen-receptor-mediated transcription and the influence of co-factors and chromatin state. *Nat. Rev. Cancer* **7**, 713–722 (2007).

387. Yu, Z., Lou, L. & Zhao, Y. Fibroblast growth factor 18 promotes the growth, migration and invasion of MDA‐MB‐231 cells. *Oncol Rep* **40**, 704–714 (2018).

388. Sonvilla, G. *et al.* FGF18 in colorectal tumour cells: autocrine and paracrine effects. *Carcinogenesis* **29**, 15–24 (2008).

389. Zhang, J. *et al.* FGF18, a prominent player in FGF signaling, promotes gastric tumorigenesis through autocrine manner and is negatively regulated by miR-590-5p. *Oncogene* **38**, 33–46 (2019).

390. El-Gendi, S., Abdelzaher, E., Mostafa, M. F. & Sheasha, G. A. FGF18 as a potential biomarker in serous and mucinous ovarian tumors. *Tumor Biol.* **37**, 3173–3183 (2016).

391. Guo, W. *et al.* Slug and Sox9 Cooperatively Determine the Mammary Stem Cell State. *Cell* **148**, 1015–1028 (2012).

392. Liu, Y. *et al.* A novel role of kynureninase in the growth control of breast cancer cells and its relationships with breast cancer. *J. Cell. Mol. Med.* **23**, 6700–6707 (2019).

393. Zhu, R. *et al.* BTG1 inhibits breast cancer cell growth through induction of cell cycle arrest and apoptosis. *Oncol Rep* **30**, 2137–2144 (2013).

394. Biddie, S. C. *et al.* Transcription Factor AP1 Potentiates Chromatin Accessibility and Glucocorticoid Receptor Binding. *Mol. Cell* **43**, 145–155 (2011).

395. Hurtado, A., Holmes, K. A., Ross-Innes, C. S., Schmidt, D. & Carroll, J. S. FOXA1 is a key determinant of estrogen receptor function and endocrine response. *Nat. Genet.* **43**, 27–33 (2011).

396. Kozub, M. M., Carr, R. M., Lomberk, G. L. & Fernandez-Zapico, M. E. LSD1, a double-edged sword, confers dynamic chromatin regulation but commonly promotes aberrant cell growth. *F1000Research* **6**, 2016 (2017).

397. Garcia-Bassets, I. *et al.* Histone Methylation-Dependent Mechanisms Impose Ligand Dependency for Gene Activation by Nuclear Receptors. *Cell* **128**, 505–518 (2007).

398. Pollock, J. A., Larrea, M. D., Jasper, J. S., McDonnell, D. P. & McCafferty, D. G. Lysine-Specific Histone Demethylase 1 Inhibitors Control Breast Cancer Proliferation in ERα-Dependent and -Independent Manners. *ACS Chem. Biol.* **7**, 1221–1231 (2012).

399. Salamero, O. *et al.* First-in-Human Phase I Study of Iadademstat (ORY-1001): A First-in-Class Lysine-Specific Histone Demethylase 1A Inhibitor, in Relapsed or Refractory Acute Myeloid Leukemia. *J. Clin. Oncol.* **38**, 4260–4273 (2020).

400. Teo, G. *et al.* SAINTexpress: Improvements and additional features in Significance Analysis of INTeractome software. *J. Proteomics* **100**, 37–43 (2014).

401. Szklarczyk, D. *et al.* STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**, D607–D613 (2019).

402. Knight, J. D. R. *et al.* ProHits-viz: a suite of web tools for visualizing interaction proteomics data. *Nat. Methods* **14**, 645–646 (2017).

403. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671–675 (2012).

404. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).

405. Kuznetsova, A., Brockhoff, P. B. & Christensen, R. H. B. lmerTest Package: Tests in Linear Mixed Effects Models. *J. Stat. Software; Vol 1, Issue 13* (2017).

406. Hothorn, T., Bretz, F. & Westfall, P. Simultaneous inference in general parametric models. *Biom. J.* **50**, 346–363 (2008).

407. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

408. R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at: https://www.r-project.org/.

409. Liao, Y., Smyth, G. K. & Shi, W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res.* **41**, e108–e108 (2013).

410. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

411. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).

412. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).

413. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

414. J.M., G. Genrich: Detecting Sites of Genomic Enrichment. (2019). Available at: https://github.com/jsh58/Genrich.

415. Tarasov, A., Vilella, A. J., Cuppen, E., Nijman, I. J. & Prins, P. Sambamba: fast processing of NGS alignment formats. *Bioinformatics* **31**, 2032–2034 (2015).

416. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).

417. Wu, D. *et al.* ROAST: rotation gene set tests for complex microarray experiments. *Bioinformatics* **26**, 2176–2182 (2010).

418. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).

419. Efron, B. & Tibshirani, R. On testing the significance of sets of genes. *Ann. Appl. Stat.* **1**, 107–129 (2007).

420. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).

421. Zhang, Y. *et al.* Model-based Analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).

422. O'Leary, N. A. *et al.* Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **44**, D733–D745 (2016).

423. Cartharius, K. *et al.* MatInspector and beyond: promoter analysis based on transcription factor binding sites. *Bioinformatics* **21**, 2933–2942 (2005).

424. Kent, W. J. *et al.* The human genome browser at UCSC. *Genome Res.* **12**, 996–1006 (2002).

425. Eklund, A. C. & Szallasi, Z. Correction of technical bias in clinical microarray data improves concordance with known biological information. *Genome Biol.* **9**, R26 (2008).

426. Koboldt, D. C. *et al.* Comprehensive molecular portraits of human breast tumours.

Nature **490**, 61–70 (2012).

427. Curtis, C. *et al.* The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* **486**, 346–352 (2012).

428. Pereira, B. *et al.* The somatic mutation profiles of 2,433 breast cancers refine their genomic and transcriptomic landscapes. *Nat. Commun.* **7**, 11479 (2016).