UNIVERSITY OF TRENTO

International PhD Program in Biomolecular Sciences

Department of Cellular, Computational and Integrative Biology – CIBIO

XXXIII Cycle

Oncogenic enhancer reprogramming in triple negative breast cancer tumour progression

Supervisor:

Prof. Alessio Zippo

Laboratory of Chromatin Biology and Epigenetics

Department of Cellular, Computational and Integrative Biology, University of Trento

Tutor:

Prof. Alberto Inga

Laboratory of Transcriptional Networks

Department of Cellular, Computational and Integrative Biology, University of Trento

Ph.D. Thesis of

Daniela Michelatti

Academic Year 2019/2020

Declaration of original authorship

I Daniela Michelatti confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged.

Contents

Abstract

Basal breast cancer is a heterogeneous disease whose unfavourable outcome is determined by a high risk of tumour relapse and metastasis formation. The potential of a cancer cell to adapt to foreign environments is favoured by oncogenic cell plasticity, which is supported by epigenetic reprogramming. It was previously demonstrated that MYC acts as an oncogenic reprogramming factor by inducing epigenetic rewiring at enhancers (Poli et al., 2018). This causes the activation of oncogenic pathways and pro-metastatic transcription factors such as *SOX9*, but scant pieces of evidence support a causal link between epigenetic alteration of oncogenic enhancers and cell plasticity.

In the present work, we investigated the establishment of an alternative epigenetic program during tumorigenesis in a basal breast cancer xenograft derived model. We found that tumorigenic cells, primary tumour derived cells and metastasis derived cells showed intrinsically different phenotypic and epigenetic signatures, and that metastatic derived cells were characterized by the acquisition of pro-metastatic features, such as migration and invasion, that may increase their metastatic potential. Specifically, we provided data supporting the notion that changes of the chromatin landscape during tumour progression increased the responsiveness of cancer cells to environmental cues that they may encounter during dissemination and colonization of distant organs. We focused on investigating the role played by putative regulatory elements localized around the *SOX9* locus, whose chromatin accessibility and interaction with the *SOX9* promoter were increased in metastatic cells.

We observed that *SOX9* expression was responsive to the activation of the retinoic acid (ATRA) pathway, and our data suggests that this response may be strengthened by transcriptional memory priming *SOX9* regulatory elements after a first exposure, so that the response is faster and more robust after the second one. *SOX9* transcription modulation and ATRA response were also shown to be linked to the activation of a quiescence program specific of metastatic cells, which we hypothesise may favour cells during the dissemination steps of the metastatic cascade.

List of figures

1. Introduction

1.1. Breast Cancer

Breast carcinoma is among the most frequent invasive cancers in women. It is a group of heterogeneous diseases encompassing different subtypes which are diverse in terms of histology, therapeutic response, dissemination patterns to distant sites and patient outcomes. Implementation of screening, prevention programs and novel treatment strategies decreases breast cancer mortality (Xu et al., 2021). However, it has been estimated that worldwide breast cancer will cause more than 700,000 deaths annually, (Xu et al., 2021).

Breast cancers (BCs) evolve from the epithelium of the mammary gland. The normal breast epithelium comprises a bilayer: the inner layer contains luminal cells required for milk production, whereas the outer layer is composed of myoepithelial cells required for milk ejection (Daniel and Smith, 1999). This apparent simplicity hides a more complex cellular hierarchy, which is revealed during carcinogenesis. Moreover, several histological and molecular subtypes of breast cancer have now been defined by the use of immunological, genetic, and gene-molecular profiling.

Studies based on global gene expression analyses have provided additional insights into this complex scenario, defining five molecular 'intrinsic' subtypes of breast cancer (Luminal A, Luminal B, HER2-enriched, Claudin-low and Basal) and a Normal Breast-like group (Sørlie et al., 2001). Luminal A is a low-grade, hormone-receptor-positive, HER2-negative subtype which is usually characterized by good prognosis and has a 63.3% incidence (Parise and Caggiano, 2017); Luminal B cancer subtype is proliferative and hormone receptor-positive, its prognosis is slightly worse than Luminal A, and its incidence is around 8.9% (Plevritis et al., 2018). Both Luminal A and B BCs are treated by long-term hormone therapy over 5–10 years (Yang and Polley, 2019). The normal-like subtype is similar to luminal A disease: hormone-receptor-positive, HER2 negative, low-proliferating; still, its prognosis is slightly worse than luminal A cancer (Liu et al., 2014).

HER2-enriched breast cancer is hormone receptor-negative and HER2 positive: HER2 positive cancers tend to grow faster than luminal cancers and can have a worse prognosis, but they are often successfully treated with targeted therapies at the HER2 protein; furthermore, it affects 4.2% of BC patients. Finally, the triple-negative breast cancer subtype (TNBC) is hormone receptor- and HER2 negative that constitutes 13% of all breast cancers. It is characterized by the severest prognosis (Harris et al., 2006; Morris et al., 2007) and it is often used as a surrogate for identifying the aggressive basal breast cancer subtype, although they are not biologically synonymous (Tischkowitz et al., 2007); therapy for this BC subtype is managed with chemotherapy only for early-stage disease (Carey et al., 2007).

Triple-negative breast cancer accounts for about 10– 15% of newly diagnosed breast cancers (BC) and is associated with lower survival than other BC subtypes (Agostinetto et al., 2021; Howlader et al., 2019). Furthermore, more than 30% of patients with TNBC eventually develop metastasis and relapses often occur during the first 2–3 years from surgery (Dent et al., 2007). This scenario reflects an intrinsic aggressive behaviour: TNBC is frequently associated with high histological grade and high proliferation index (Ki67) as well as the lack of oncogenic targets, namely hormone receptors and human epidermal growth factor receptor-2 (HER2) (Zhu et al., 2020).

The genetic background of the TNBC primary tumour is well understood. Triple-negative breast cancer is characterized by the low expression of ER/PR/HER2, and mutations in BRCA1/BRCA2 are commonly associated with the basal subtype of breast cancer on a genomic level (Chen et al., 2018; Foulkes et al., 2003). Cancers that do not present BRCA1/BRCA2 mutations but have other causes of homologous recombination deficiency (HRD) are characterized by high tumour grade, lymphocytic infiltrate, pushing margins, ER and HER2-negativity, an association with TP53 mutations, c-Myc amplification, and multiple chromosome abnormalities (Chen et al., 2018). TP53 mutations were found to be the most common somatic aberration, observed in 53.8% of cases, while the TNBC samples also showed frequent mutations in PIK3CA (10.2%) (Atchley et al., 2008). The most common copy number alterations were identified for the PARK2 (Parkinson disease 2) (6%) and RB1 (retinoblastoma gene 1) genes (7.7%) (Shah et al., 2012).

The PI3K/AKT/mTOR pathway is one of the key and active pathways involved in the chemoresistance and survival of TNBC (Khan et al., 2019). Dysregulation of the PI3K/AKT/mTOR pathway causes changes in cell survival, differentiation, and proliferation that are frequently observed during carcinogenesis of TNBC, but also other BC tumour types (Khan et al., 2019; Sporikova et al., 2018).

1.2. Metastasis, seeding and cell plasticity

Metastasis accounts for more than 90% of cancer-related deaths. Unlike primary tumours, which can often be treated using local surgery or radiation, metastasis is a systemic disease. Currently, systemic approaches such as screening, chemotherapy, targeted therapy and immunotherapy, are therefore the basis of metastasis prevention and treatment (Ganesh and Massagué, 2021).

Although cancer cell dissemination from the primary tumour can start early during tumour progression, the vast majority of cells leaving the primary site fail to colonize distant organs and instead capitulate to various stresses (Harper et al., 2016; Hosseini et al., 2016; Hu and Curtis, 2020; Srinivas et al., 2005). In order to form metastases, cancer cells must deal with a series of steps termed the 'metastatic cascade': aggressive cancer cells must leave the primary tumour, disseminate through the bloodstream, and eventually reach distant organs to give rise to one or several metastases. By acting on heterogeneous cancer cell populations, these pressures select for cells with increased fitness to colonize distant organs. Intra-tumour heterogeneity can be a consequence of genetic variation, differences in gene regulation, transitions between cellular states, or environmental perturbations (Gupta et al., 2011; Meacham and Morrison, 2013; Nik-Zainal et al., 2012; Shah et al., 2012).

Genomic instability, genetic variation within a tumour and the high phenotypic variability in malignant cells contribute to the heterogeneity of cancer cell populations (Al-Hajj et al., 2003; Ganesh and Massagué, 2021). Stem-like cancer cells can react to both cell-intrinsic developmental programs and external stromal signals through switching between different phenotypic states. This feature, called phenotypic plasticity, allows cancer cells to adjust to particular microenvironments, overcome the difficulties of the metastatic seeding and resist therapy (Fumagalli et al., 2020; Quintanal-Villalonga et al., 2020; Rios et al., 2019). The potential of cancer cells to establish metastatic relapse may arise from the heterogeneity of cancer cell populations, which might be influenced by the composition of the tumour stroma.

1.2.1. Metastatic relapse in TNBC

For breast cancer, the statistics emphasize the role of metastasis in BC mortality: 5-years survival rates in patients with distant metastasis are only 22% (13% at ten years), compared to 90% for patients with local disease (Nagao et al., 2021). Furthermore, metastatic relapse in BC patients is currently not allowing effective treatment options. Approximately 25% of BC patients will relapse with distant metastasis: the majority of patients, unfortunately, relapse following treatment with curative intent (Yao et al., 2019). TNBC is characterized by a propensity for visceral and brain metastases, absence of bone metastases and typically early relapse (< 3 years).

Understanding how BC cells go through the metastatic cascade and the development of therapies to target metastatic processes continues to be a significant clinical challenge. Age, smoking, ethnicity, menopause, endogenous hormones, the histological status of cells, duration of breast feeding, and the underlying biology of the tumour, such as grade and size of the primary tumour, can increase the probability of metastasis occurrence, as described in Shekar et al., 2020.

It is widely acknowledged that the outbreak of metastatic lesions is a very inefficient process, and it is considered as the rate-limiting step of breast cancer metastasis (Aguirre-Ghiso, 2007; Luzzi et al., 1998; Valastyan and Weinberg, 2011). During dissemination, breast cancer cells are usually difficult to be detected, as they mainly circulate throughout the body as single rare cells, and only a limited number of genes and markers are used to define the phenotypic and molecular alterations that characterize disseminated tumour cells (DTCs) (Giancotti, 2013). Furthermore, during dissemination, cancer cells exhibit resistance to chemotherapy, due to the low proliferation rate (Marches et al., 1999). DTCs can enter a state of dormancy upon arrival to distant sites, through leaving the proliferative cycle for a certain period or by stalling in an equilibrium between proliferation and death. Reawakening of dormant cells is the result of further evolution of surviving DTCs, that accumulate molecular alterations and interact with the tumour microenvironment (Demicheli et al., 2007; Pantel et al., 2008).

While considerable progress has been made to investigate the specific processes at the basis of breast tumour initiation, recent studies have recently been focusing on elucidating the roles of key genes, the underlying molecular mechanisms and signalling pathways involved in the lethal late stages of metastatic dissemination.

1.2.2. Metastasis: from intravasation to macrometastasis

Metastasis is traditionally defined as a linear series of steps, collectively referred to as the invasion-metastasis cascade (**Figure 1**) (Fidler, 2003). Firstly, cancer cells at the primary site of tumour growth detach from one another or adjacent normal cells, digest the underlying basement membrane and penetrate the underlying interstitial matrix (invasion). Subsequently, tumour cells encourage tumour vascularization (neoangiogenesis), exploit its discontinuities to gain access to the vessels (intravasation),

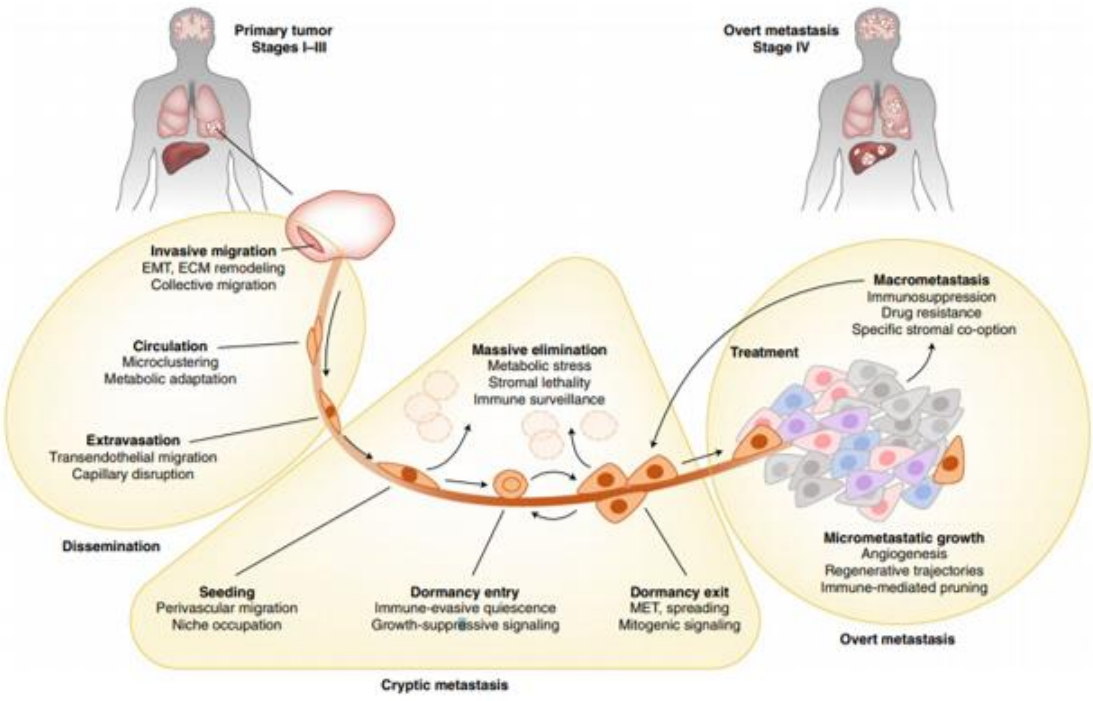


Figure 1 Steps, biological functions and cancer cell vulnerabilities in the metastasis cascade. *The majority of cancer cells leaving a primary tumour are unable to survive and are cleared. Cancer cells that survive and retain the ability to regenerate the tumour during the cryptic phase of metastasis are called metastasis-initiating cells (MICs). MICs launch overt metastatic growth in distant organs, develop along tissue-regenerative trajectories and deploy organ-specific stromal co-option functions. ECM, extracellular matrix; EMT, epithelial–mesenchymal transition; MET, mesenchymal–epithelial transition. (Ganesh et al., Nature Medicine 2021)*

and disseminate through the bloodstream (dissemination). Finally, after depositing at the microcirculatory system of the target organ and infiltrating its stroma (extravasation), cancer cells take advantage of their plasticity to survive and eventually outgrow into macroscopic lesions (colonization).

During local invasion, cancer cells detach from primary tumour and enter the stroma and the adjacent tissue parenchyma. To invade the stroma, carcinoma cells must first violate the basement membrane (BM), in part by separating their epithelial and stromal compartments. Degradation of the extracellular matrix (ECM) macromolecules by matrix metalloproteinases (MMPs) facilitates tumour cell invasion (Gonzalez-Avila et al., 2019; Justilien et al., 2012). MMPs are zinc-dependent endopeptidases implicated in basement membrane degradation, angiogenesis and cell migration, contributing to tumour invasion and metastasis (Balbín et al., 2003; Deryugina and Quigley, 2006). MMPs are differentially regulated by tumour cells, and their functions are controlled by their interactions with cellular and extracellular components of the tumour microenvironment and distant pre-metastatic sites (Lian et al., 2021). Besides playing a structural role, the BM holds a source of growth factors that can be released by carcinoma-secreted proteases (Hu et al., 2008).

Once cancer cells have successfully intravasated into the bloodstream, they can disseminate widely through the venous and arterial circulation; at this stage, tumour cells are referred to as circulating tumour cells (CTCs). A small percentage of CTCs successfully enter distant sites and persist as disseminated tumour cells (DTCs) (Dasgupta et al., 2017). DTCs in the bloodstream must survive a variety of stresses in order to reach distant organ sites (Gkountela et al., 2019; Yu et al., 2013). Once settled in the microcirculation of distant organs, DTCs may initiate intraluminal growth and get in direct contact with the tissue parenchyma through rupturing the walls of surrounding vessels (Al-Mehdi et al., 2000). Otherwise, cancer cells may extravasate to enter the tissue parenchyma, through penetrating the endothelial and pericyte layers that separate vessel lumina from the stroma.

Extravasated cancer cells must survive in the foreign microenvironment they encounter in distant tissues to form micrometastasis, which usually differs greatly from that present in the primary tumour site (Hiratsuka et al., 2008; Kaplan et al., 2005; Kiel

et al., 2005; Psaila and Lyden, 2009). If disseminated cancer cells endure their initial arrival at the microenvironment of a distant tissue and thrive, they still are not guaranteed to proliferate and form large macroscopic metastases. Instead, it seems that the majority of DTCs suffer either slow degradation over time, or persist as microcolonies in a state of long-term quiescence (Chambers et al., 2002; Demicheli et al., 1998; Naumov et al., 2002).

1.2.3. Metastatic cell plasticity and dormancy

Once cancer cells successfully penetrate the blood or lymphatic circulatory system, they can disseminate throughout the body (Ganesh and Massagué, 2021). When circulating through the bloodstream, cancer cells are likely to enter a state where they remain latent, characterized by the stalling in the G0-G1 phases of the cell cycle: this state is termed quiescence. DTCs can remain latent, constantly fluctuating between quiescent and proliferative states until conditions allow these cells to escape clearance by the immune system and grow as a metastatic outbreak (Aguirre-Ghiso et al., 2013; Phan and Croucher, 2020). Quiescence is a reversible state by which under appropriate environmental cues, DTCs may reawaken by re-entering into the cell cycle and eventually give rise to metastatic lesions, thus acting as metastasis initiating cells (MICs) (Yachida et al., 2010). The switch between the proliferative and dormant state depends on stromal signals, cell contacts, extracellular matrix and metabolic cues, and allows MICs survival and outbreak (Celià-Terrassa and Kang, 2018) (Oskarsson et al., 2014).

MICs undergo dynamic phenotypic changes, which might be initiated by hyperactivation of specific signalling pathways (Kocal et al., 2016), a gain of migration capacity, anoikis evasion and growth reinitiating capacity (Er et al., 2018; Ganesh et al., 2020).

Dormancy can be defined as either cellular dormancy, where intrinsic or extrinsic mechanisms drive solitary or small groups of DTCs to enter quiescence; angiogenic dormancy, where the tumour mass is characterized by an equilibrium between cells that proliferate and cells that die because of poor vascularization, and immune-mediated

dormancy, where the immune system keeps a proliferating tumour mass constant via a persistent cytotoxic activity (Sosa et al., 2014). Of importance, these mechanisms are not mutually exclusive but they may coexist, thereby underlying the contribution of multiple players in the onset and maintenance of the dormant state.

Indeed, evidences providing insights into the molecular mechanisms of cellular dormancy and reactivation are accumulating. Crosstalk between cancer cells and their microenvironment plays a crucial role in these mechanisms, which is affected by complex interactions between cancer cells, stromal cells and surrounding extracellular matrix (ECM) components, and host immunity. After an extended period in the bloodstream, DTCs ultimately reach distant sites and face a new composition of ECM. Then, the binding of membrane receptors on DTCs activates various signalling cascades, pushing cell cycle promotion and escape of dormancy. The host immune system initially acts as a tumour suppressor but eventually favours cancer progression and promotes metastatic outgrowth by reactivating DTCs (Cameron et al., 2000; Luzzi et al., 1998; Park and Nam, 2020).

Recent evidence demonstrates that extracellular signal-regulated kinase (ERK) activation plays a crucial role in determining the switch between the proliferative and the dormant state (Aguirre-Ghiso et al., 2003). High levels of the p38 mitogen-activated protein kinase (MAPK) activity during ERK-induced proliferation function as an inhibitory regulator of ERK and prevent cell proliferation by inducing G0-G1 arrest or triggering senescence and apoptosis (Sosa et al., 2011).

Different microenvironments limit or permit cancer progression to metastatic sites, modulating the composition of dormant DTCs reservoirs. Dormancy activation might also happen long before the metastatic seeding: chemotherapy targeting the breast primary tumour was also demonstrated to induce dormancy and facilitating cancer cell dissemination (Karagiannis et al., 2017). Nevertheless, as recently demonstrated, the immune system activity plays a central role in the regulation of the dormant milieu: Correia et al. demonstrated that natural killer (NK) cell population is increased in the dormant reservoir of basal-like breast cancer and that interleukin-15-based immunotherapy sustains dormancy, preventing hepatic metastases. Moreover, exit from dormancy induced by the CXCL12-CXCR4 axis follows a reduction of the NK cell

compartment, allowing metastatic outgrowth (Correia et al., 2021). In multiple myeloma, osteoblasts express GAS6, the ligand for AXL, and prostate cancer cells disseminating to the same endosteal niche bind GAS6 on osteoblasts and thus enter dormancy (Khoo et al., 2019). Interestingly, exposure of primary tumour cells to hypoxia was shown to prime subsequent disseminated tumour cells to enter into a dormant state dependent on NR2F1 expression, indicating that the DTC state might already pre-exist within the primary tumour (Fluegen et al., 2017). Furthermore, in the work of Malladi et al., the WNT pathway was highlighted as a regulator of quiescence, which allows cancer cells to evade innate immunity remaining latent for extended periods. In this context *SOX9* transcription factor was identified in latency competent breast cancer cells as an essential promoter of survival under immune surveillance and metastatic outgrowth, under permissive conditions (Malladi et al., 2016). The *SOX9*-dependent resistance to NK cells was also pointed out in lung cancer metastases during the progressive stages of metastatic dormancy outbreak (Laughney et al., 2020).

Furthermore, recent results point towards a role for retinoic acid pathway activation as a trigger for NR2F1-dependent dormancy induction and maintenance by integrating quiescence epigenetic programmes and survival in DTCs (Sosa et al., 2015). In addition, induction of NR2F1 in an HNSCC PDX line resulted in lung metastasis inhibition through the expression of a dormancy gene signature, which includes the transcription factors SOX9 and RARβ(Khalil et al., 2021).

1.2.3.1. The retinoic acid pathway and quiescence

The concept of cancer cells remodelling their epigenetic programs to adapt to different microenvironments and eventually enter a quiescent state is in line with the notion of cancer fitness (Yeh and Ramaswamy, 2015). For example, embryonic environments can reset cellular programmes: axolotl (Ambystoma mexicanum) oocyte extracts can reprogram breast cancer cells and reverse their tumorigenicity. Oocyte extract-treated tumour-xenografts gene expression profiling shows that tumour reprogramming encompasses cell cycle arrest and acquisition of a quiescent state by induction of the cell-cycle inhibitor p27 and reducing signalling dependent on the transcription factor JUN and the tumour suppressor Rb86 (N. et al., 2018). These quiescent breast cancer cells also displayed involvement of epigenetic reprogramming; this link was previously reported in retinoic acid-induced dormancy: retinoic acid was found to upregulate the master receptor NR2F1 and to induce the expression of pluripotency genes *SOX9*, *RARß*, and *NANOG,* leading to quiescence in head and neck squamous cell carcinoma (HNSCC) cells (Sosa et al., 2015).

All-trans retinoic acid (ATRA) is the active metabolite of vitamin A, and its biological effects are mediated by the activation of retinoic acid receptors (RARs) (Germain et al., 2006; Samarut and Rochette-Egly, 2012). Vitamin A is absorbed in the small intestine and is esterified to retinyl esters, transported in the bloodstream and stored in the liver (Dawson et al., 2000)(**Figure 2**). Hydrolysis of retinyl esters results in retinol: peripheral cells assume retinoids mainly as *all-trans*-retinol (atROL) bound to retinol-binding proteins (RBPs) in the plasma through the transporter STRA6 (Chen et al., 2016). After target organs take up the atROL/RBP complex, atROL is re-esterified into retinyl ester by the lecithin retinol acyltransferase (LRAT) or binds to the cellular retinol-binding protein (CRBP) (Batten et al., 2004). AtROL/CRBP complexes are the first substrate in the metabolic pathway that leads to the production of ATRA: the substrate is oxidised to *all-trans*-retinal (atRAL) by the retinol dehydrogenase (RDH) and subsequently oxidised to

ATRA by various retinaldehyde dehydrogenases (RALDH). Oxidation of atRAL is the rate-limiting step in ATRA production (Rao et al., 2021; Raverdeau and Mills, 2014).

ATRA is shuttled by cellular retinoic acid-binding protein (CRABP) to the nucleus, where different RARs are activated (Bourguet et al., 2000). Alternatively, ATRA can be shuttled by the fatty acid-binding protein (FABP) to bind peroxisome proliferator-activated receptors (PPARs) in the nucleus (Borland et al., 2011).

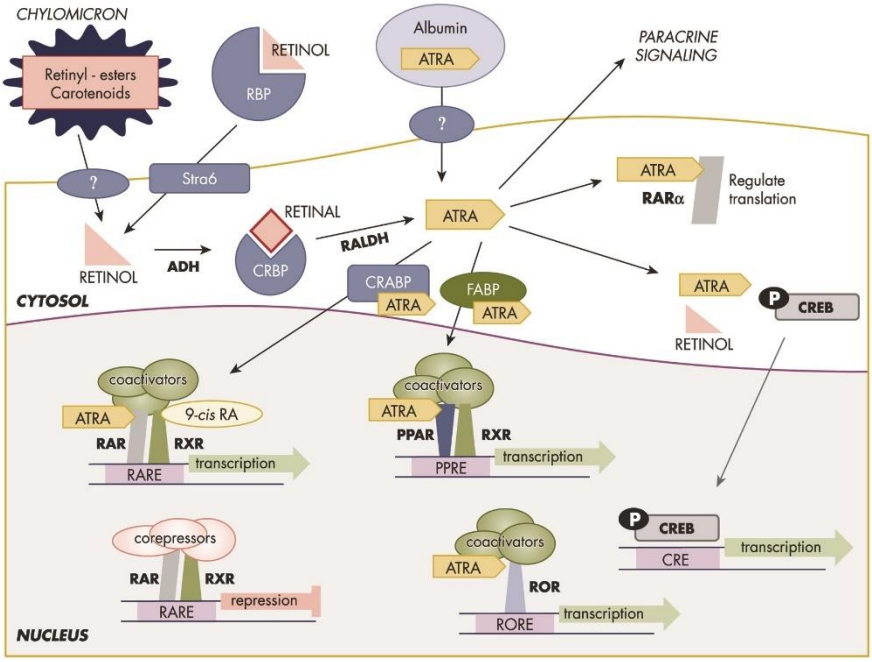


Figure 2 The retinoic acid pathway. *Retinoids reach target cells either as retinol bound to the Retinol binding protein (RBP), or as ATRA, retinyl esters and caroteinoids. Retinol is imported in the cell through the signal receptor and transporter of retinol STRA6, then it is oxidized to all-trans-retinal by the alcohol dehydrogenase (ADH) and bound to the cellular retinol binding protein (CRBP). All-trans-retinal is then oxidized to ATRA by the retinal dehydrogenase (RALDH) and transported by CRABP to the nucleus where different RARs are activated. Alternatively, ATRA can be shuttled by FABP to bind PPARs in the nucleus. In addition to RARs and PPARs, ATRA can also bind to RORs to initiate transcription. (Conoway et al.,* Endocrine Reviews*, 2013)*

ATRA is degraded by CYP26 enzymes, which belong to the cytochrome P450 hydroxylase family. Since CYP26 and RALDH are both regulated by ATRA itself, the metabolism of ATRA forms an auto-regulatory loop; such regulation maintains the endogenous ATRA level within a normal range and allows the organism to respond to exogenous ATRA fluctuations (Stevison et al., 2017).

The retinoid receptors are part of the human nuclear receptor superfamily (Errore. L'origine riferimento non è stata trovata.) and can be divided into four regions. The N terminal domain (NTD) has a ligand-independent transactivation function, the DNA binding domain is a cysteine-rich domain and contains two zinc finger structures, the H

domain contains the nuclear localization sequence, and the ligand (retinoid) binding domain (LBD) has a region required for dimerization with other receptors (Petkovich et

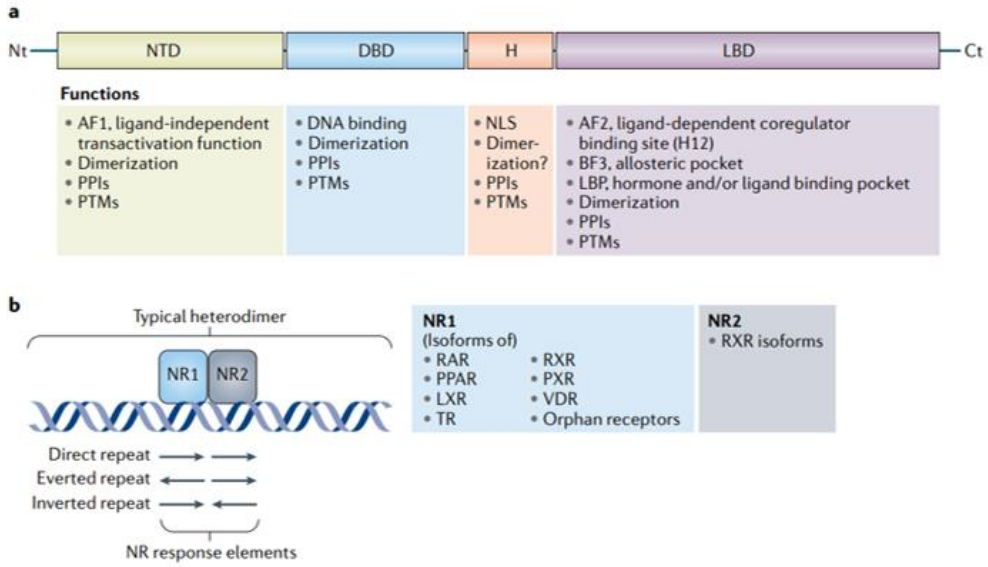


Figure 3 NR functional domain organization and most relevant regulatory functions. *A) Nuclear receptor (NR) domains are depicted following the archetypal structure of a NR, including the N-terminal domain (NTD) with the activation function 1 (AF1), DNA-binding domain (DBD), hinge region (H) and C-terminal ligand-binding domain (LBD) with the activation function 2 (AF2) and ligand-binding pocket (LBP). All NR domains undergo post-translational modifications (PTMs), which could regulate protein stability, and multiple domains are involved in interactions with co-regulator proteins (protein–protein interactions (PPI)). NRs harbour a nuclear localization signal (NLS). B) Typical NR dimer binding to DNA takes place at cognate DNA response elements featuring (imperfect) palindromic sequences or two hexanucleotide half-sites organized as direct, everted or inverted repeats. Typical heterodimers are defined here as the well-characterized heterodimers of which both partners contact DNA, sharing RXR as a common protein partner. LXR: liver X receptor; PPAR: peroxisome proliferator activated receptor; PXR: pregnane X receptor; RAR: retinoic acid receptor; RXR: retinoid X receptor; TR: thyroid receptor; VDR: vitamin D receptor. BF3: binding function 3; Ct: C terminus; Nt: N terminus. (De Bosscher et al., Nature Reviews Endocrinology 2020)*

al., 1987).

Two families of retinoid nuclear receptors have been described, the retinoic acid receptors (RARs) and the retinoid X receptors (RXRs) (Mangelsdorf et al., 1990). The RARs (α, β, and γ) show high affinity for binding to the naturally occurring retinoid ATRA, whereas the RXRs (α, β, and γ) do not bind ATRA. 9-cis-retinoic acid (9cRA) is a naturally occurring, biologically active isomer of ATRA capable of binding and trans-activating both the RXRs and the RARs (Mangelsdorf et al., 1991). In the absence of retinoic acid, RARs are found to be bound to retinoic acid response elements (RAREs) in the form of heterodimers with retinoid receptors (RXRα, β, and γ) and interact with transcriptional corepressors to create a more condensed state of chromatin and inhibit transcription (Cordeiro et al., 2019). When instead ATRA is available, its binding leads to a conformational change of the RAR/RXR heterodimers, causing the detachment of

corepressors and the exposure of binding surfaces to coactivators, which ultimately leads target gene transcription, such as differentiation-related genes (Hashimoto and Shudo, 1991; Yoon et al., 2003).

The development and regeneration of tissues and tumours depend on the balance of different signalling pathways, including the ATRA pathway. Retinoic acid was found to promote or suppress tumour growth in TNBC depending on the cell context-specific balance of activation of transcriptional and non-transcriptional pathways involving RARα (Rossetti et al., 2016). RA was also shown to activate the pro-invasive Src-YAP-Interleukin 6 axis in MDA-MB-468 cells and increase tumour growth and metastasis of MDA-MB-231 cells (Marcato et al., 2015; Mezquita et al., 2018).

It was demonstrated that ATRA stimulate *SOX9* gene expression in breast cancer cell lines, giving evidence for a role in retinoid-mediated growth inhibition (Afonja et al., 2002). Metastasis often develops years after surgical removal of the primary tumour from a small population of disseminated cancer cells that survived as latent entities. It is thus conceivable that retinoic acid inhibition of proliferation may in certain settings not act as a tumour suppressor but instead favour the survival of disseminated cancer cells under immune surveillance and for metastatic outgrowth under permissive conditions. Latency competent cancer cells (LCC) isolated from human lung and breast carcinoma show stem-cell-like characteristics and express SOX2 and SOX9 transcription factors, which allow LCC cells to enter quiescence and evade innate immunity to remain latent for extended periods (Malladi et al., 2016).

Furthermore, RA receptors were recently documented as mediators of higher-order chromatin remodelling: upon ATRA signalling, the RAR/ RXR transcription factor-induced loss of adjacent CTCF binding and changed the higher-order chromatin conformation of the overall locus (Ishihara et al., 2016; Wang et al., 2017).

1.3. Cell plasticity, chromatin rewiring and transcriptional memory

The genome of eukaryotic cells is arranged into chromatin, a viscoelastic polymer comprising DNA, RNA, and associated proteins (**Figure 4**) (Sitbon et al., 2017). Chromatin shows a hierarchical organization, ranging from the basic repeating unit, the nucleosome, to higher-level structures (van Steensel and Belmont, 2017). A nucleosome is constituted by a core particle with ~147 base pairs of double-stranded DNA draped around histone proteins. Linker DNA joins core nucleosomal units. Next, nucleosomes fold into higher-order chromatin structures, where the chromatin filament further loops and compacts DNA (Furlong and Levine, 2018).

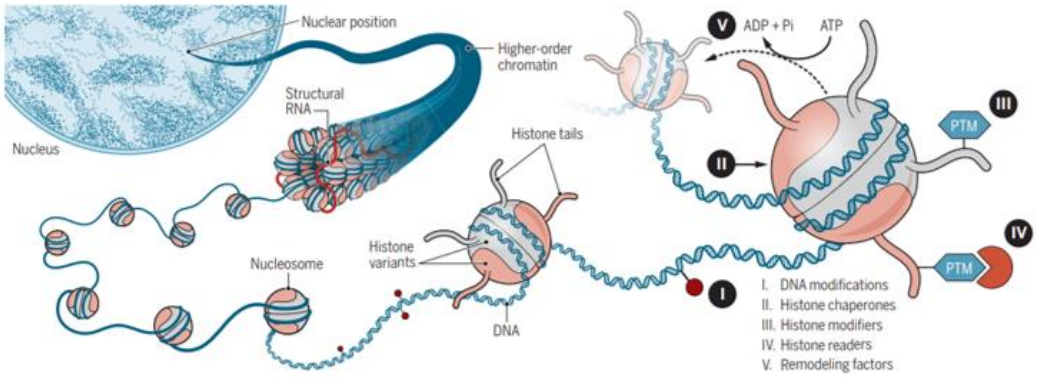


Figure 4 Chromatin organization. *Nucleosomes arrange into higher-order chromatin structures, and noncoding RNA participates in the local organization. Inset (right): the nucleosome is composed of double-stranded DNA draped around an octamer, consisting of an (H3-H4)2 tetramer lined by two H2A-H2B dimers. Variations of the nucleosome by dynamic permutations comprise the choice of histone variants, modifications of DNA bases, and reversible post-translational modifications (PTM) of histone tails, enabling chromatin plasticity. (Yadav et al., Science 2021)*

Modulation at each level of the genome organization enables chromatin-based information to vary to respond to signals for numerous gene regulatory functions (Wingelhofer et al., 2018). This identifies chromatin plasticity as a tool to produce a variety of features for each cell type (Philip et al., 2017). Regulation of cell destiny decisions and cell identity can make use of chromatin, for example, by limiting access to a particular transcription factor or by providing distinct marks on histones that specific proteins can recognize and can interpret in response to signalling (Adam and Fuchs, 2016; Burton and Torres-Padilla, 2014).

1.3.1. Enhancer function in gene expression control

The human genome includes about 20,000 protein-coding genes according to different databases (Farrell et al., 2014; Harrow et al., 2012; O'Leary et al., 2016), which is less than that in *Arabidopsis thaliana* (The Arabidopsis Genome Initiative, 2000). There is a remarkable lack of correspondence between the number of protein-coding genes and organism complexity. The reason for the higher biological complexity in humans is due to the precise and complicated transcription regulation. Precise temporal and spatial gene expressions are achieved by cis-acting regulatory elements such as promoters, enhancers, and insulators. Enhancers play central roles in controlling selective gene transcription (Alexander et al., 2019). Enhancers are defined as fragments of noncoding DNA sequence with multiple transcription factor binding sites, which regulates gene transcription at a long distance (Carter et al., 2002).

Chromosomes organize spatially into topologically associating domains (TADs), where intradomain interactions are preferentially selected to interdomain interactions (Sexton et al., 2012). TAD architecture allows for the reduction of the distance between regulatory elements within its boundaries, facilitating interactions between them (Chen et al., 2020a). Definition of interactions within chromosomal compartments is mediated by the cohesin complex, which associates with CTCF and with active gene regulatory elements to form long-range interactions between its binding sites (Pugacheva et al., 2020).

Enhancers and promoters are physically connected by the Mediator complex: the Mediator complex provides communication between active enhancers and promoters by interacting with proteins that bind to either of these two classes of regulatory DNA elements and stimulates the cooperative assembly of a pre-initiation complex (PIC) and recruitment of RNA Polymerase II (Pol II) for gene activation (Quevedo et al., 2019).

Chromatin patterns commonly used for identifying putative enhancers include the binding of p300, histone H3 acetylated at lysine 27 (H3K27ac) and H3 monomethylated

at K4 (H3K4me1), or mapping chromatin accessibility sites by DNAse-seq or ATAC-seq. Genome profiling has revealed that cell-specific transcription factors and RNA polymerase II (Pol II) are recruited to enhancers (Agrawal et al., 2020); recruitment of transcription machinery components might be favoured by enhancer-promoter physical proximity mediated by chromatin looping (Pennacchio et al., 2013). In addition, many enhancers are bound by RNA polymerase II and actively transcribed, generating noncoding enhancer RNAs (eRNA), widely used to indicate enhancer activity and target gene regulation (Kim et al., 2010).

Enhancers spanning large genomic regions are currently defined as super-enhancers (SEs). Super-enhancers have a median size generally an order of magnitude larger than typical enhancers and show enrichment for factors associated with enhancer activity such as MED1, a component of the Mediator complex and coactivator of transcription (Whyte et al., 2013).

Enhancers share standard genomic and epigenetic features, such as the enrichment for binding sites of both lineage-determining and signal-dependent transcription factors (LDTFs and SDTFs, respectively), which allow for transduction of both internal and external signals (Calo and Wysocka, 2013). Alteration of signalling pathways is a common feature of cancer: de-regulated SDTFs can change the transcriptional program, causing an altered activity of their target enhancers: indeed, oncogenic enhancers are often enriched in binding sites specific for those TFs (Clevers and Nusse, 2012; Fagnocchi et al., 2016). Cell identity is established and maintained by LDTFs (also referred to as 'master TFs'), which bind to super-enhancers to regulate cell-type-specific genes (Hnisz et al., 2013; Whyte et al., 2013). Many LDTFs possess pioneering DNA-binding ability, thus engaging silent chromatin and favouring the recruitment of other TFs and co-factors during cell fate determination (Zaret and Mango, 2016). Accordingly, the aberrant activity of LDTFs is a significant driver of multiple cancers and has been demonstrated to modulate binding of oncogenic TFs and activation of tumour-specific enhancers (Bradner et al., 2017).

Since enhancer activity modulation is highly involved in the maintenance of cell identity and the control of the adaptive capability of cells, the possibility that genetic insults affecting the epigenetic landscape may alter regulatory element activity and thus

subvert cell fate establishment is plausible. Indeed, in T-cell acute lymphoblastic leukaemia (T-ALL), binding motifs for the MYB transcription factor are introduced through a somatic mutation in a specific noncoding site, which creates a super-enhancer upstream of the TAL1 oncogene (Mansour et al., 2014).

Over the last years, million putative enhancers have been annotated in the human genome, generating complex modular and combinatorial regulatory networks, in which not only several enhancers influence the expression of a single target gene, but also single specific enhancers may mediate the synchronous transcriptional bursting of multiple promoters, in response to developmental and environmental cues (Andersson et al., 2014; Fukaya et al., 2016).

1.3.1.1. The role of the chromatin context

While hierarchical levels of chromatin organization ensure the proper packaging of the genomic material into the cell nucleus, they have an equally important role in regulating genome function. Gene expression can be regulated at every level of chromatin organization, from positioning nucleosomes on the DNA to the localization of chromosome territories within the nucleus (Woodworth and Holloway, 2017).

At the most fundamental level, the chromatin structure influences gene expression by restricting access of transcription factors and the transcription machinery to the DNA (Klemm et al., 2019). Chromatin exists in two general forms within the nucleus: heterochromatin and euchromatin. Heterochromatin is nucleosome dense, highly condensed and tends to be transcriptionally inactive due to the exclusion of transcriptional regulators from the DNA, while euchromatin shows less nucleosome occupancy, is less compacted, and is more accessible to transcriptional regulators (Allshire and Madhani, 2018; Spielmann et al., 2018). However, the chromatin structure is highly dynamic and can be converted between hetero- and euchromatin states, or a permutation of intermediate states, in response to intracellular or extracellular signals.

Such changes in chromatin state have the potential to expose or obscure gene regulatory elements to the nuclear transcriptional machinery, thus facilitating alterations in gene expression (Jansen et al., 2012; Sartorelli and Puri, 2018).

The activity of enhancers often depends on their ability to physically interact with the transcriptional machinery at promoters through DNA looping (**Figure 5**). Different molecules play a significant role in bridging transcription factors (TFs) and co-activators at enhancers with the pre-initiation complex at promoters, among which Mediator, cohesin, and CCCTC-Binding Factor (CTCF) are the best characterized (Kagey et al., 2010).

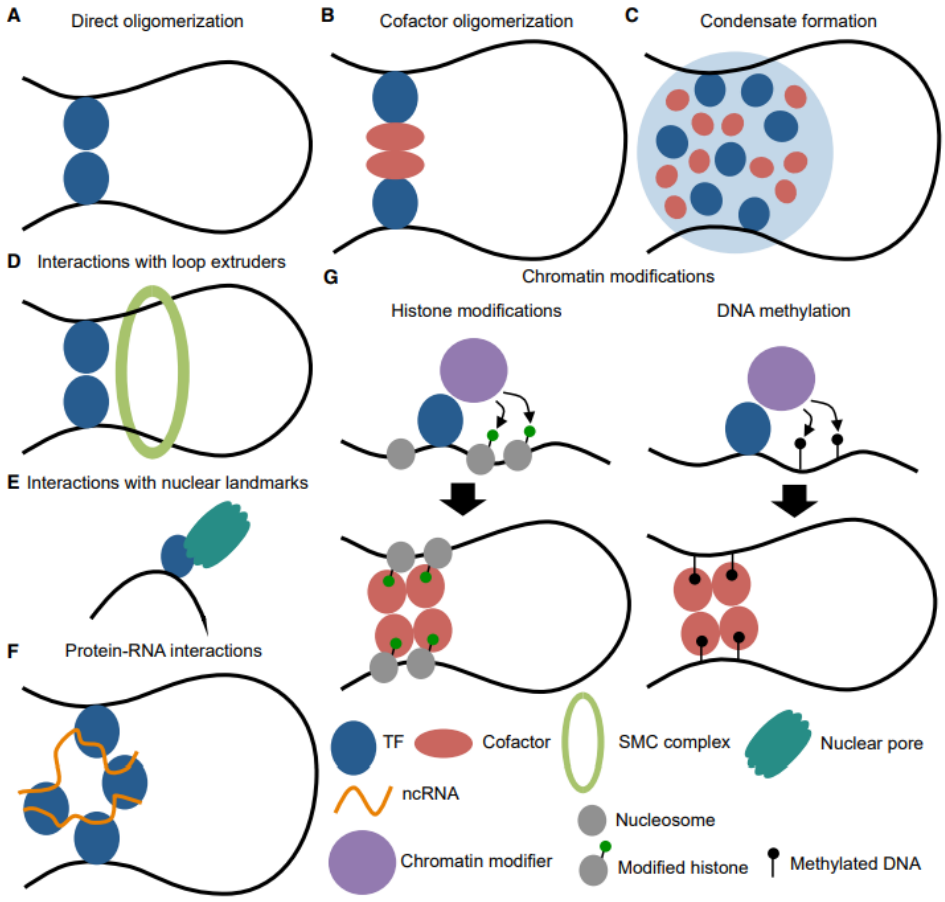


Figure 5 Modality of TF action on 3D genome architecture: *A) Direct oligomerization. B) Cofactor oligomerization. C) Condensate formation. D) Interactions with loop extruders. E) Interactions with nuclear landmarks F) Protein-RNA contacts. G) Chromatin variations, including histone modifications (left panel) and DNA methylation (right panel). (Kim et al., Molecular Cell, 2019)*

Pioneer TFs bind to enhancers and trigger the recruitment of chromatin-remodelling enzymes, resulting in open chromatin and a typical pattern of histone modification on the adjacent nucleosomes, including H3K27ac and H3K4me1 (Long et al., 2016).

TFs are defined by their capability to bind DNA but generally function via interactions with other proteins and RNA. The most instinctive example of TF-driven DNA loops involves the direct oligomerization of TFs, where TFs that bind both promoters and regulatory elements mediate enhancer-promoter interaction through the formation of homodimers (**Figure 5A**)(Lambert et al., 2018; Weintraub et al., 2017). However, TFs can also participate in DNA looping by recruiting cofactor proteins that, in turn, form oligomers (**Figure 5B**) (Wang et al., 2021).

In addition to the solid protein-protein interactions that trigger oligomerization, evidence shows that weak but multivalent interactions among intrinsically disordered regions (IDRs) result in condensates or hubs that exhibit properties of liquid-liquid phase separation (**Figure 5C**) (Banani et al., 2017; Fasciani et al., 2020). Recently, it has been shown that TFs can form condensates together with cofactors in an IDR-dependent manner (Sabari et al., 2018). Furthermore, enhancers enriched for a shared list of TFs binding sites have been detected nearby in ‘‘nuclear microenvironments’’ or ‘‘3D cliques’’ that are enriched for those factors (Petrovic et al., 2019; Sabari et al., 2018).

Interaction between TFs and cohesin and other structural maintenance of chromosomes (SMC) complexes plays a central role in the architectural distribution of the genome. SMC complexes can encircle DNA and actively pull a loop of DNA through the ring until they are blocked or unloaded (**Figure 5D**) (Ganji et al., 2018).

Nuclear genomic localization can also be affected by interaction with nuclear landmarks such as nuclear pore components (Nups) and lamins through direct or indirect interactions with TFs (**Figure 5E**) (Brickner et al., 2019; Fudenberg et al., 2016).

In addition to protein-protein interactions, protein-RNA binding such as the one shown for YY1 and CTCF is involved in structuring the 3D genome (**Figure 5F**) (Kung et al., 2015).

Ultimately, TFs can alter genome architecture through interaction with other TFs or proteins (**Figure 5G**). The recruitment of cofactors that may facilitate DNA methylation, histone tail modifications, or nucleosome positioning can regulate the interaction with other TFs or further cofactors, such as the methyl-CpG binding protein 2 (MeCP2) or the acetyl-lysine binding BRD4; these may induce chromosome reshaping through protein-

563 protein or protein-RNA interactions (O'Malley et al., 2016; Zhu et al., 2018). For
564 example, BRD4 occupancy on distal oestrogen responsive elements (EREs) enriched for
565 H3K27ac is required for recruitment and elongation of RNAPII on EREs and the
566 production of oestrogen-receptor α-dependent enhancer RNAs in ER+ breast cancer
567 cells (Nagarajan et al., 2014).

568 It has been shown that the 3D organization of the genome can regulate TF activity by
569 supporting specificity in coupling enhancers and the genes they target. Nevertheless, at
570 least in some cases, physical proximity precedes enhancer-promoter activation (Rhie et
571 al., 2019).

1.3.2. Cell plasticity and transcriptional memory

Recent efforts have focused on elucidating the relationship between the time-space changes in 3D chromatin architecture, transcriptional regulation and human development and diseases (Vinckier et al., 2020; Zhou et al., 2019). Within the primary tumour mass, tumour cells are often characterized by a developmentally plastic state and self-renewal capacity (Friedmann-Morvinski et al., 2012; Gupta et al., 2011). Transcriptional dysregulation of genes involved in controlling stem cell functions and stem cell signalling in cancer cells promotes dedifferentiation, inducing the development of cancer cells with stem cell-like features, such as self-renewal (Feng et al., 2021). Deregulation of transcription can follow the alteration of structural protein activity: the loss of insulated neighbourhoods or TADs boundaries through impairment of CTCF binding was shown to play a role in carcinogenesis (Hnisz et al., 2018). Epigenetic modulation of gene expression contributes to the dynamic responsiveness of cellular processes that sustains cellular identity and phenotypic fluidity (Robson et al., 2017).

Moreover, it is increasingly evident that cells can retain a memory of the initial transcriptional response through epigenetic mechanisms, which "prime" the genome so that it can respond more quickly to subsequent exposure to the signal (Kerr and Corbett, 2010; Phillips et al., 2019). Epigenetic transcriptional memory was first reported in inflammatory context: in yeast, epigenetic memory was shown to be initiated by the binding of transcription factors to promoters and regulatory elements, which repress gene transcription after exposure to stimuli such as IFN-γ (Light et al., 2013). IFNγ was also shown to prime HeLa cells changing local chromatin structure at primed genes and inducing a transcriptional memory that propagates in cells for longer than two weeks and allows a stronger and faster activation of the interferon response upon re-exposure to IFNγ (Siwek et al., 2020). In addition, the transcription factor FOS was shown to interact with JUN and to cooperate with STAT3 in the establishment of inflammatory memory through bookmarking of memory domains (Larsen et al., 2021). Furthermore,

598 in HEK293F, TNFα stimulation drives the establishment of a TET-dependent
599 transcriptional memory (Zhao et al., 2020).

600 Evidence is accumulating to demonstrate that other processes, besides
601 inflammation, can drive transcriptional memory establishment: monitoring the
602 response to wounds, it was shown that stem cells are able to bear memories of their
603 original niche, migration, encounters with inflammation, and adaptation to the new fate
604 and tasks (Uy et al., 2021).

605 The mechanisms of transcriptional regulation have also been suggested to be
606 implicated in supporting metastatic seeding through the integrated stress response
607 signalling activation (Nagelkerke et al., 2013). It is thus conceivable that DTCs that have
608 been primed with certain stimuli and are thus able to respond more efficiently to insults
609 they may encounter during their challenging journey towards distant sites are more
610 capable to adapt to the microenvironments they encounter and have thus more chances
611 to survive the dissemination process.

1.3.3. Chromatin rewiring and enhancer reprogramming drive cancer

Genome-Wide Association Studies (GWAS) have indicated that a large portion of genetic variants that predispose to cancer are located in non-coding cis-regulatory elements (Campbell et al., 2020). Among these, alterations of enhancers play a crucial role in cancer biology: oncogenic enhancer reprogramming can be defined as cancer-related alterations that, independently of their origin, cause aberrant oncogenic activity, leading to de-regulated transcriptional programs, which foster tumour progression and metastasis dissemination (Bi et al., 2020; Fagnocchi et al., 2018; Huang et al., 2021).

Enhancer reprogramming contribution to cancer development often derives from the alteration of regular gene expression; both *cis*- and *trans*-factors evade the regulation that maintains cell-type specificity and enforce an alternative, de-regulated enhancer activity in cancer cells (Errore. L'origine riferimento non è stata trovata.)(Lin et al., 2020).

Cis-acting aberrations include single-nucleotide polymorphisms (SNPs), small insertions or deletions (INDELs), and more significant structural variants such as focal amplifications, large deletions, inversions, and translocation of existing enhancers (Khurana et al., 2016). SNPs that occur in the body of existing enhancers can disrupt TF-binding sites, directly inactivating enhancers and leading to transcriptional down-regulation of the original target gene, therefore, favouring tumour onset (Kandaswamy et al., 2016; Mansour et al., 2014; Oldridge et al., 2015).

Trans-acting chromatin-modifying proteins (CMPs) are the ultimate effectors of cis-acting regulatory elements: they modulate DNA-histone interaction, change chromatin conformation, and increase or decrease the binding of functional DNA-regulating protein complexes. Consequently, their mutation and misexpression/function have been widely associated with both solid and haematological malignancies, representing them as significant mediators of cancer disruption and progression. Nonetheless, they

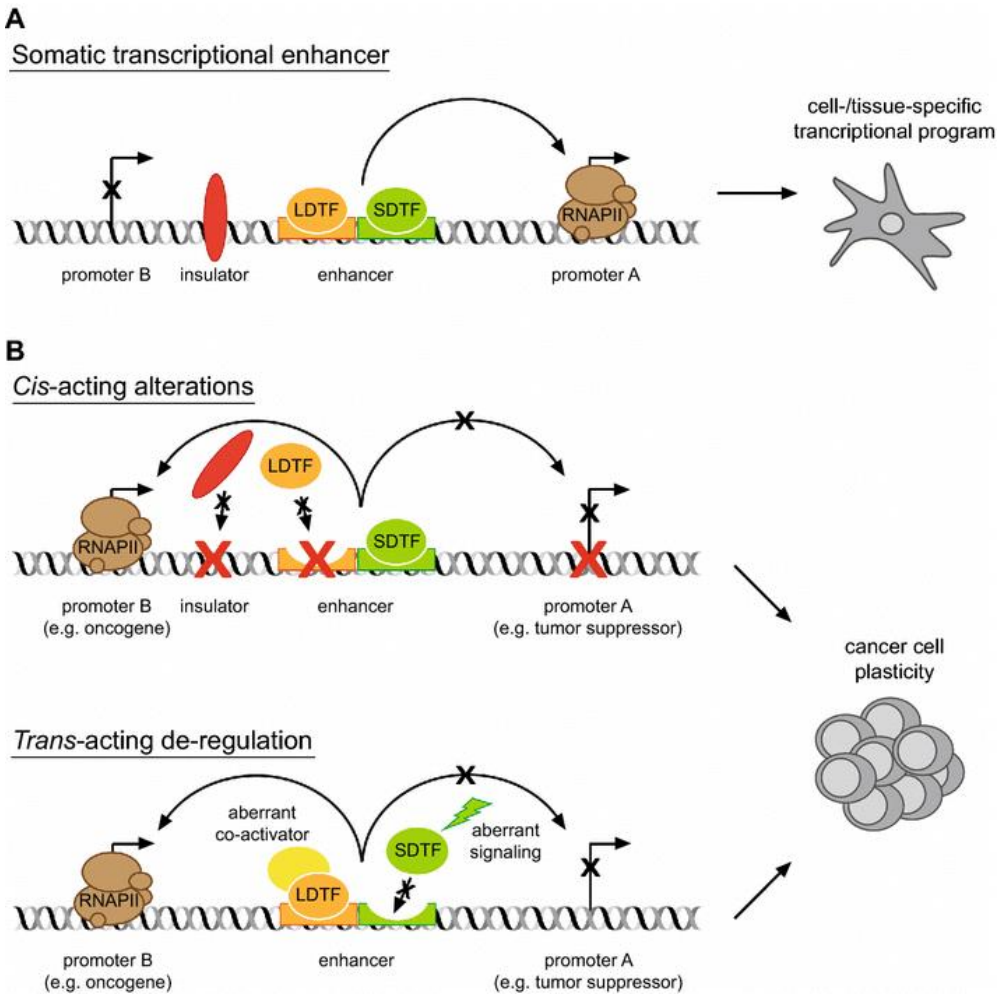


Figure 6 Aberrant activity of transcriptional enhancers favours cancer cell plasticity. *A) Transcriptional enhancers control cell- and tissue-specific transcriptional programs, integrating lineage-determining and signal-dependent transcription factors. B) Aberrant functionality of both cis- and trans-acting factors impose disrupted transcriptional landscapes, which might induce reprogramming and favour cancer cell plasticity. Black crosses: loss of promoter, enhancer, or trans-factors activity; red crosses: point mutations in cis-regulatory elements. Arched arrows: functional interaction between enhancers and promoters. LDTF: lineage-determining transcription factors; SDTF: signal-dependent transcription factors; RNAPII: RNA polymerase II (Fagnocchi et al., Cellular and Molecular Life Sciences 2018)*

typically exert their tumorigenic function combined with mutations of other well-defined oncogenes or tumour suppressor genes (Morgan and Shilatifard, 2015).

DNA methylation is mediated by DNA methyltransferases (DNMTs). In physiological conditions, CpG islands throughout the genome are methylated, keeping these regions silenced, while the promoters of active genes are hypomethylated. Moreover, hypo-methylation followed by oncogenic TFs binding and methylation plasticity at enhancers predict metastatic progression and patient mortality, respectively (Fleischer et al., 2017; Vidal et al., 2017).

The H3K4me1 histone mark is deposited by MLL3/4 methyltransferases, part of the COMPASS complex, and characterizes transcriptional enhancers (Bae and Lesch, 2020). The H3K27ac histone marks identify active enhancers, and it is deposited by p300/CBP (Sze and Shilatifard, 2016). The H3K27 demethylase UTX is also a subunit of COMPASS and favours enhancer activation by removing H3K27me3, the repressive histone modification introduced by Polycomb group proteins (PcG) which identify 'poised enhancers'. Given their leading role in shaping transcription, deregulation of all the above-mentioned chromatin modulating proteins is recurrent in many cancer types.

ATP-dependent chromatin remodelling complexes (CRCs) are responsible for maintaining and altering chromatin structure by moving, ejecting, or restructuring the nucleosome. Four distinct families of CRCs are known in eukaryotes: SWI/SNF, ISWI, CHD, and INO80. Those CRCs families differ for their functional activity, protein domains, and subunits (St Pierre and Kadoch, 2017). Importantly, SWI/SNF is frequently implicated in malignant transformation, with more than 20% of human cancers carrying mutations in components of this CRC, and can act both as a tumour suppressor and oncogene (Centore et al., 2020).

As mentioned previously, the activity of enhancers depends on their ability to physically interact with the transcriptional machinery at promoters. Therefore, alteration of the functionality of the structural factors mediating enhancer-promoter looping is associated with cancer disruption (Losada, 2014; Soutourina, 2018).

Despite the substantial progress in understanding the genetic and epigenetic changes driving tumour initiation (Charlton et al., 2020; Pomerantz et al., 2020), the mechanisms driving cancer progression and metastasis formation are largely unknown. Since metastasis is the primary cause of death among cancer patients, defining those traits that characterize metastatic capacity represents an unmet medical necessity.

1.4. The SOX family of transcription factors

The SOX family of transcription factors consists of more than 20 members that mediate DNA binding by the HMG domain (High-Mobility Group box) and have regulatory functions in development, cell-fate decision, and differentiation. These proteins have been classified into eight groups based on HMG domain sequence, protein structure and evolutionary relationships (Errore. L'origine riferimento non è stata trovata.). In humans, members of each of these groups are SOXA: SRY; SOXB: SOX1, SOX2, SOX3, SOX14 and SOX21; SOXC: SOX4, SOX11 and SOX12; SOXD: SOX5, SOX6 and SOX13; SOXE: SOX8, SOX9 and SOX10; SOXF: SOX7, SOX17 and SOX18; SOXG: SOX15 and SOX20; and SOXH: SOX30 (Lefebvre et al., 2007).

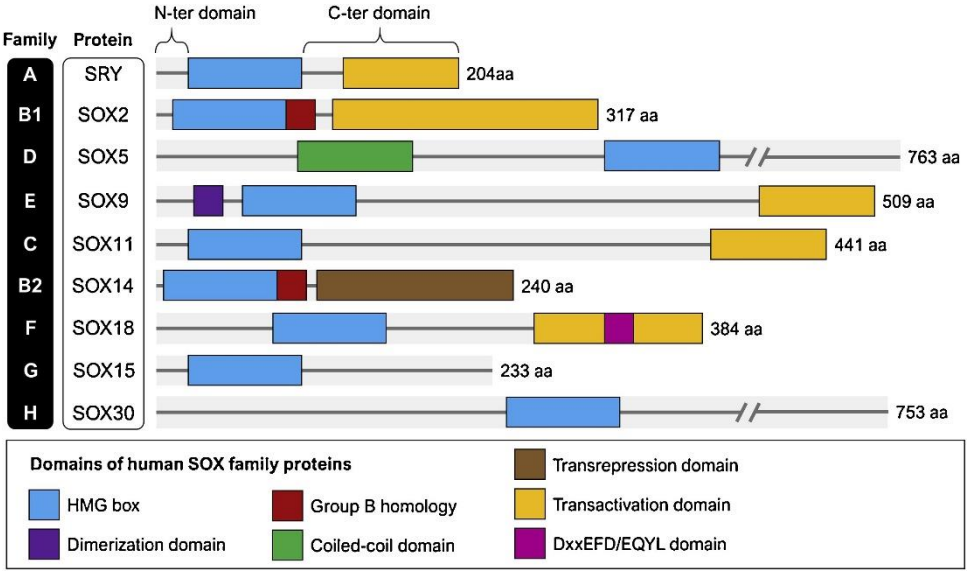


Figure 7 Domain structures of the human SOX protein family. *The human SOX family of transcription factors is subdivided into eight groups (SoxA through SoxH). The highly conserved and characteristic HMG box is specified alongside other functional domains, including the transactivation domain. Groups and representative protein members are indicated to the left. N-terminal (N-ter) and C-terminal (C-ter) domains of SRY are depicted at the top. The sizes in amino acids (aa) of the various SOX proteins are shown to the right. Domains of SOX family proteins are shown in the box. (Grimm et al., Seminars in Cancer Biology 2020)*

SOX2, SOX4, SOX5, SOX8, SOX9, and SOX18 have been shown to be involved in the promotion of different cancer types and to be associated with poor prognosis, while the up-regulation of SOX11 and SOX30 has been demonstrated to be favourable for the outcome in other cancer types (Chen et al., 2020b; Han et al., 2015; Lu et al., 2017; Ma et al., 2020; Miao et al., 2018; Wang et al., 2019; Yao et al., 2015; Zou et al., 2018). Among the SOX family members, SOX2, SOX4, and SOX5 have been shown to be involved

in tumorigenesis, e.g. SOX2 is markedly up-regulated in chemotherapy-resistant cells (Mukherjee et al., 2017). In addition, the SoxF family (SOX7, SOX17, SOX18) has been demonstrated to play an essential role in angio- and lymphangiogenesis, with SOX18 being a potential target for anti-angiogenic therapy (Grimm et al., 2020).

SOX proteins are characterized by the evolutionarily conserved HMG box, which is a 79-amino-acid DNA-binding motif (Stros et al., 2007). The HMG box is responsible for the mediation of the DNA-binding capacity of SOX proteins on a common consensus site, (A/T)(A/T)CAA(A/T)G (Grimm et al., 2020). The HMG box comprises two independent nuclear localization signals (NLSs) and one leucine-rich nuclear export signal (NES), which control the nucleus-cytosol shuttling of SOX proteins; subcellular distribution of SOX proteins has been shown to be distinct during development (Malki et al., 2010).

1.4.1. *SOX9* as a readout of oncogenic enhancer activity

It was previously demonstrated that the TF MYC acts as an oncogenic factor by inducing epigenetic reprogramming in TNBC (Poli et al., 2018). This causes the activation of oncogenic pathways and the expression of pro-metastatic transcription factors such as *SOX9* (Aldaz et al., 2020; Poli et al., 2018). Based on the recent findings and on the established role of *SOX9* in TNBC, in this thesis, *SOX9* is used as a readout of oncogenic enhancers activity.

SOX family transcription factors are well-established regulators of cell fate decisions during development. Recent literature documents that these transcription factors play different roles in adult tissue homeostasis and regeneration. Remarkably, SOX factors overexpression combined with other synergistic factors reprograms differentiated cells into somatic or pluripotent stem cells (Malik et al., 2019). Dysregulation of SOX factors has been further implicated in diseases, including cancer (Sarkar and Hochedlinger, 2013).

1.4.2. *SOX9* in TNBC

Recent evidence demonstrates that a subset of SOX proteins regulates critical aspects of breast cancer biology, including cancer stemness and multiple signalling pathways leading to altered cell proliferation, survival, tumour development, Epithelial-Mesenchymal Transition, cell migration and metastasis, as well as other tumour-associated characteristics (Christin et al., 2020; Wang et al., 2021a).

Many recent studies have shown that multiple members of the SOX transcription factor family are overexpressed and activated in TNBC or basal tumours, and emerging data provide evidence that this family of proteins play an essential role in tumour development and progression (Adam et al., 2015; Jana et al., 2020; Mehta et al., 2019; Panda et al., 2021; Petrovic et al., 2019).

SOX9 is oncogenic and has been shown to be essential for lineage commitment, differentiation and EMT during embryonic development, as well as being crucial for oncogenesis through regulation of cancer stem cell population in breast tumours (Chakravarty et al., 2011).

In the context of breast cancer, SOX9 nuclear expression was found to be significantly enriched in basal-like tumours compared to ER+ and HER2+ breast cancers, and its overexpression is correlating with lower distant metastasis-free survival in TNBC (**Figure 8** and **Figure 9**) (Guan et al., 2019; Pomp et al., 2015; Tang et al., 2020; Tariq et al., 2020).

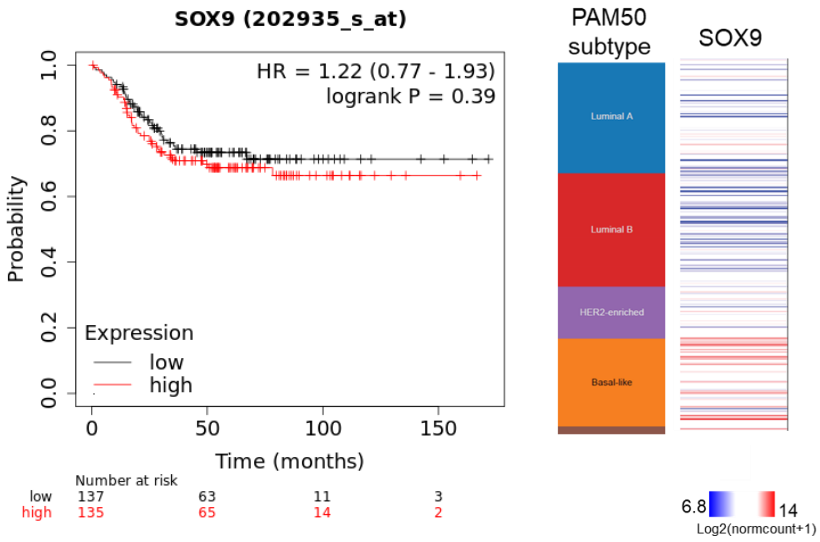


Figure 9 SOX9 expression in breast cancer. *SOX9 expression correlates with poor prognosis of TNBC patients (left): Kaplan-Meier plot computed on KMplot.com using data from 272 ER negative (array), PR negative (IHC), HER2 negative (array), basal breast cancer patients Distant Metastasis Free Survival (DMSF). SOX9 expression in basal subtype of breast cancer (right): Heatmap computed on XenaBrowser.net; RNA expression of 1247 TGCA Breast Cancer samples classified on the basis of PAM50_mRNA_nature2012 phenotypic classification.*

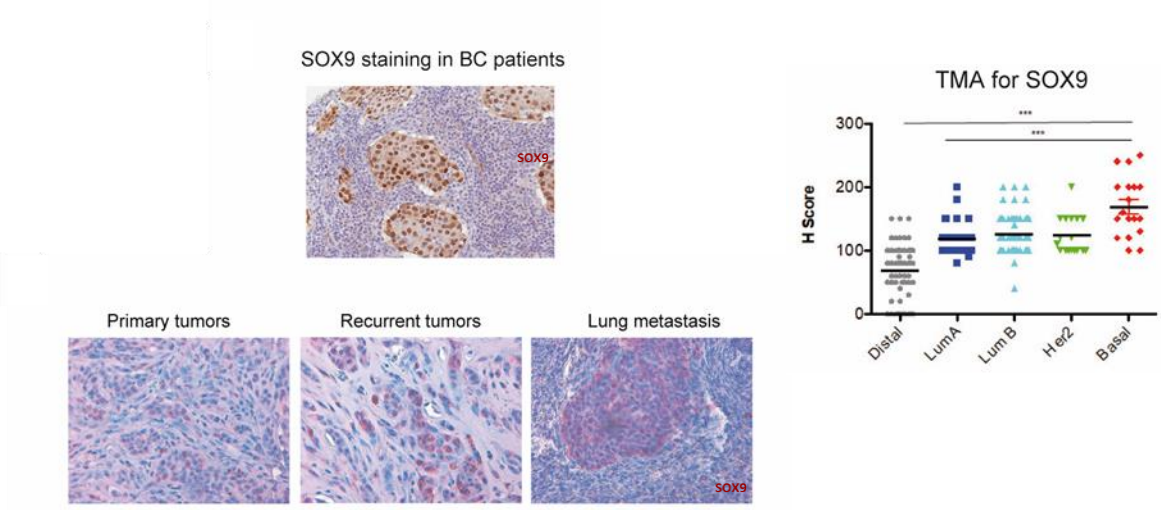


Figure 8 SOX9 IHC analysis of BC patients. *Immunohistochemical analysis of BC samples (n=124) and distal samples (n=74) in primary tumours, recurrent tumours and lung metastasis (left) and relative quantification (right). SOX9 staining in red/maroon (red to maroon toning follows dye exposure timing). One-tailed t-test; ***: P≤<0.0001. In collaboration with Prof. Bosari, Milan General Hospital, Pathological Anatomy.*

Notably, increased SOX9 expression was associated with upregulation of the $CD44^{hi}$ $CD24^{low}$ cancer stem cell phenotype as well as poor prognosis (Ma et al., 2020).

SOX9 is one of the multiple TFs with the ability to induce EMT in multiple types of tumours. Studies have shown that the EMT process is accomplished by direct binding of SOX9 to the promoters of SLUG, VIM (Vimentin), CLDN1 (Claudin-1), CTNNB1 (β-Catenin), and ZEB1 (Zinc finger E-box-binding homeobox 1) in various cancers. In

addition, SOX9^{+} cells possess a mesenchymal state upon stimulation with TGFβ, which shows that SOX9^{+} cells exhibit metastatic potential. Moreover, SOX9+ cells exhibit self-renewal capacity and differentiate into SOX9+ and SOX9− populations in hepatocellular carcinoma (Kawai et al., 2016). Furthermore, it was recently demonstrated that Notch signalling facilitates repositioning of specific enhancers to activate direct Notch target genes, among which SOX9 belongs to, independently of changes in enhancer H3K27ac level in TNBC cell lines (Petrovic et al., 2019).

SOX9 gene is located in a gene-desert region of the genome, thus probably enriched for regulatory elements. In this thesis, SOX9 is used as a readout of oncogenic activity of regulatory elements proximal to the SOX9 promoter in the context of TNBC metastasis. SOX9 enhancer dis-regulation during metastasis may be acknowledged as a proof of concept of chromatin remodelling instructing for metastatic-specific enhancer response.

1.5. A xenograft-derived model of basal breast cancer

In order to investigate enhancer reprogramming during tumorigenesis, we characterized a basal breast cancer model previously developed in the laboratory where I worked during my PhD (**Figure 10**) (Poli et al., 2018).

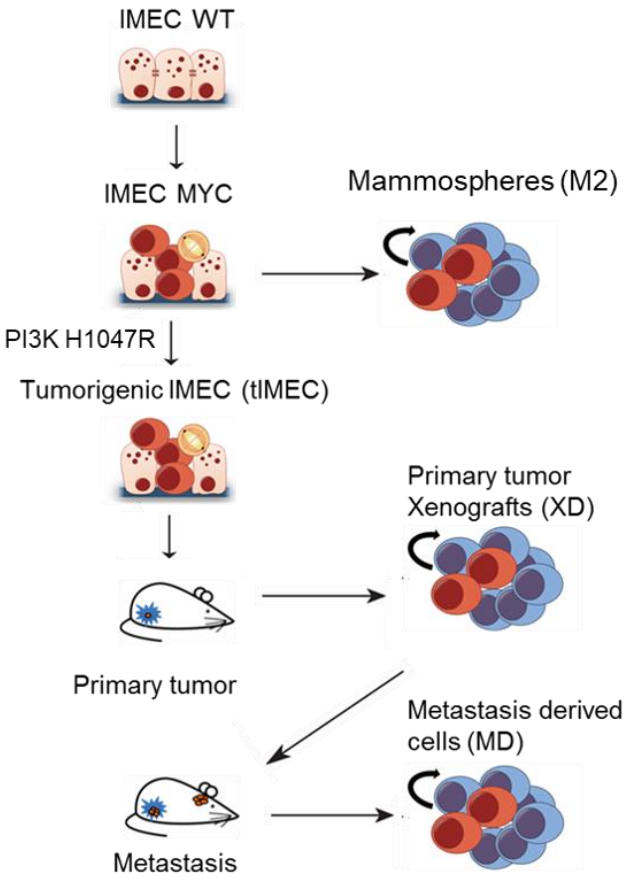


Figure 10 The basal breast cancer model derivation scheme. *hTERT-immortalized human mammary epithelial cells (IMEC WT) were transduced with a retroviral vector expressing low levels of the exogenous c-Myc. IMEC-MYC cells were allowed to grow for subsequent passages in low adherence conditions as mammospheres. IMEC-MYC were then transduced with an overexpressing PIK3CA^{H1047R} vector. IMEC-MYC-PIK3CA^{H1047R} cells (tIMEC) formed tumours from which Xenograft cells were derived (XD) which, when re-injected in the mice, gave rise to lung metastasis, which allowed the derivation of a metastatic cell line (MD). (adapted from Poli et al., Nature Communication 2017)*

Briefly, hTERT-immortalized human mammary epithelial cells (thereafter named IMEC) were transduced with a retroviral vector expressing low levels of the exogenous c-Myc. IMEC-MYC cells were allowed to grow for subsequent passages in low adherence conditions as mammospheres. IMEC-MYC were then further challenged with an

additional oncogenic insult by overexpressing PIK3CA^{H1047R}, which caused hyper-activation of PI3K pathway. IMEC-MYC-PIK3CA^{H1047R} cells and the corresponding controls were injected in the sub-renal capsule of immunocompromised mouse hosts. All mice injected with IMEC-MYC-PIK3CA^{H1047R} cells formed tumours composed of highly proliferative (KI67^{+}) and poorly dedifferentiated cells. Hereafter, we refer to IMEC-MYC-PIK3CA^{H1047R} as tumorigenic-IMEC (tIMEC). Primary tumours samples were processed for genome-wide analysis, and xenograft-derived cells from 4 tumours were derived. To determine long-term tumorigenic potential, xenograft-derived (XD) cells obtained from primary tumours were re-injected in the mammary gland of secondary recipient mice. Serial transplantations showed that the XD cells maintained tumorigenicity, forming tumours with features resembling the primary one. Notably, XD cells showed considerable migration and metastatic seeding capacity as, after surgical resection of secondary tumours, all treated animals developed macro-metastasis in the liver, lung, and spleen. Xenografts from the lung metastasis allowed a metastatic cell line derivation, henceforth referred to as metastatic-derived cells (MD).

IMEC WT, IMEC-MYC and mammospheres derived from IMEC-MYC were extensively characterized in the work from which the project of my PhD started (Poli et al., 2018) through gene expression profiling (RNA-seq) and histone marks enrichment analysis (ChIP-seq). From those results, we were already able to assess different histone marks distribution and enrichment for de-novo accessible regions in the chromatin, so we hypothesised that the cell types representing tumorigenesis steps derived from IMEC-MYC might be affected by additional changes at the chromatin level.

As depicted in **Figure 11**, when comparing IMEC WT to IMEC-MYC and mammospheres (M2), the distribution of those histone marks that individuate active or inactive enhancers is quite different between the cell types; in particular, the distribution of the deposition of H3K27ac and H3K4me1 in M2 highlighted the activation of the so-called “de novo enhancers” located in the proximity of promoters guiding the expression of transcription factors, among which, SOX9.

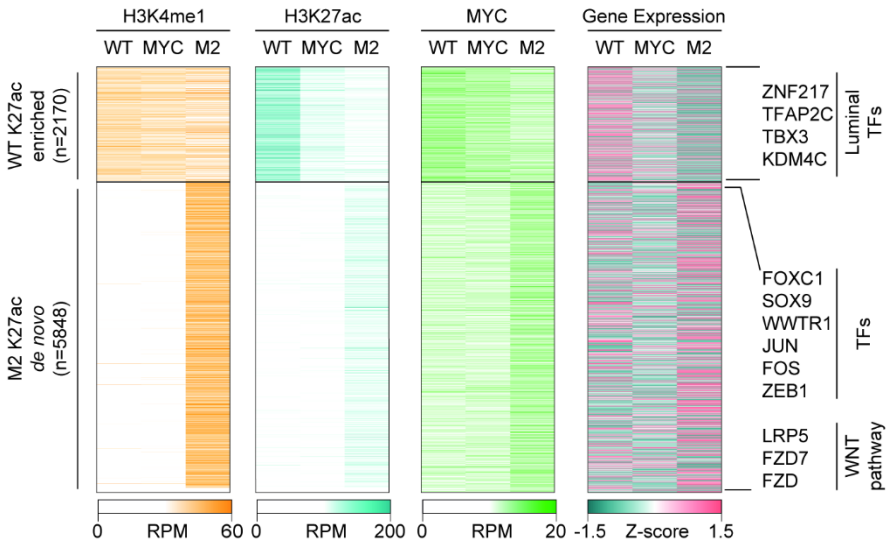


Figure 11 MYC induces an alternative epigenetic program in mammary epithelial cells. *Heatmap showing the dynamic behaviour of H3K4me1, H3K27ac, and MYC normalized ChIP-seq signals overidentified modulated enhancers. Expression of associated genes is reported. Key relevant genes associated to different enhancers groups are indicated on the right. RPM reads per million.*

2. Aim of the thesis

The main aim of this thesis is to elucidate the function of enhancer activity modulation in tumorigenesis and metastasis formation in the context of triple-negative breast cancer. We hypothesise that enhancer increased activity in tumour cells instructs for the acquisition of traits that facilitate the process of dissemination and metastasis formation.

To verify our hypothesis, we first aimed to establish a TNBC xenograft derived model system to characterize the phenotypic traits and epigenetic landscape that distinguish metastatic cells from primary tumour derived cells and tumorigenic cells. From the genome-wide analysis, *SOX9* emerged as one of the most active transcription factors in metastatic cells, and the *SOX9* locus arose as an example of increased accessibility and plasticity during tumour progression. We thus focused on the *SOX9* locus as a proof of concept of regulatory elements dysregulation during tumorigenesis in TNBC, increasing cell adaptability and cancer fitness.

Secondly, to investigate the establishment of the alternative epigenetic program in metastatic cells, we focused on a specific locus and the augmented accessibility and interaction of putative enhancer elements in its surroundings. We worked towards identifying putative enhancer regions acting on the SOX9 promoter and validating these regions as regulatory elements.

Finally, we sought to investigate the features acquired by metastatic cells and the mechanisms activated following induction of the retinoic acid pathway. In particular, we focused on characterizing the activation of a quiescence program specifically by MD cells in response to ATRA and in cooperation with SOX9 levels modulation.

3. Materials and methods

3.1. Cell lines

All experiments were performed in the following cell lines: hTERT-immortalized human mammary epithelial cells transduced with pMXs-c-Myc, PGK-H2B-mCherry and pBabe-puro-HA-PIK3CA^{H1047R} (tIMEC), primary tumour xenograft derived cell populations from mouse #5, #6, #7 and #8 (XD), lung metastasis derived cell population #6 (MD). Derivation of the cited cell types is described in Poli et al., 2018. MDA-MB-231 and MCF7 cells lines were retrieved from ATCC. These cell lines were used as representative of transformed mammary basal and luminal epithelial cells, respectively. All cell lines were tested for mycoplasma contamination and resulted negative.

3.2. Plasmids

For Luciferase Assays, pGL4.27[luc2P minP Hygro] and pGL4.75[hRluc/CMV] were kindly provided by Dr. Robert Clarke. To generate the positive control, pGL4_TCF, TCF sequence was amplified by PCR with primers containing AfeI and NheI restriction sites at 5' and 3' respectively (Fwd: CCGAGCTCTTACGCGAGATCAA; Rev: CAGTACCGGAATGCCAAGCTG). To generate pGL4_enhancers constructs, enhancers were amplified by PCR from IMEC WT cells genomic DNA with primers containing AfeI and NheI restriction sites at 5' and 3' respectively:

EN	Fwd	Rev
EN 10	GCCAGCGCTGCTGAGGAAAAGACTGTATCTCCAAA	GCCGCTAGCCTCCTCTATTAACAACCCTGCTCTA
EN 17	GCCAGCGCTGGGTTAAAACATTTCCGGGACTG	GCCGCTAGCTGACAACTTGTCAGCATCCCTACTT
EN 13	CCTTTAGCTAGCTGGTTTCCAGTT	GCCGCCAGCGCTGATACCTAGTAAGGATTTTTCTACATAT
EN 15	CTAGCCAGCGCTGAAACATAGATAGACCAGCTAATTGG	GCCGCTAGCGCAAAATAAATAAAATCTGCTGT

For *SOX9* knock-down and overexpression, TetO Fuw Sox9 and pLKO shSox9 were purchased from Addgene (#41080 and #40644 respectively).

For the enhancer reporter system, ppbnl-CAG-mVenus and pCMV-PBase were a kind gift from Luca Tiberi's lab; positive control PB-TCF-GFP was constructed by subcloning TCF-minimal promoter-GFP-Thymidine-kinase through SalI and NotI restriction sites, added by amplifying the insert by PCR (Fwd: GCCGTCGACAGCGCTGAAGCTTGCATGCCTGCAGGGTACC; Rev: GGAAAAAAGCGGCCGCTCAGTTAGCCTCCCCCATCTCCCGG) and subcloning it in place of the CAG-mVenus. hPGK promoter and Hygromycin resistance were amplified with primers containing EcoRV and SalI-AgeI restriction sites and 5' and 3' respectively (Fwd: GCCGATATCGGGGTTGGGGTTGCGCCTTTTCCA; Rev: GCCGTCGACCCCCGCCTAACCGGTTTCCTTTGCCCTCGG) and subcloned upstream of TCF; SNAPtag was amplified by PCR with primers containing HindIII-AgeI and FseI-SalI restriction sites at 5' and 3' respectively (Fwd: GCCAAGCTTACCGGTATGGACAAAGACTGCGAAAT; Rev: GCCGGCCGGCCTTAACCCAGCCCAGGCTTGCCCAGTCTGTG) and subcloned downstream of Hygromycin resistance. For PB-enhancers-GFP, TCF fragment in PB-TCF-GFP was substituted with enhancers amplified from IMEC WT genomic DNA with the same primers used for Luciferase Assay plasmids cloning.

3.3. Cell culture, stable cell lines and transfection

tIMEC, XD and MD cells were cultured at 37 °C and 5% CO2 in 1:1 DMEM/F-12 medium (gibco #11320-074) supplemented with 100 U/mL Penicillin/Streptomycin (gibco # 15140122), insulin (Clonetics, MEGM SingleQuots #CC4136), EGF (Clonetics, MEGM SingleQuots #CC-4136), hydrocortisone (Clonetics, MEGM SingleQuots #CC-4136), B-27 Supplement (Gibco # 17504044) and 20 ng/ml human FGF-basic (PeproTech #100-18B). Cells were seeded at 20 $*10^3$ cells/mL and split every 72-96 hours.

MCF7, MDA-MB-231 and metastatic derived (LM2 – lung; BrM2 – brain) cells were cultured at 37 °C and 5% CO2 in DMEM high glucose (Euroclone #ECB7501L) supplemented with 100 U/mL Penicillin/Streptomycin (gibco # 15140122), 10% fetal bovine serum (Euroclone #ECS0180L), 1 mM sodium pyruvate (Euroclone #ECM0542D) and 2 mM glutamine (Euroclone #ECB3000D).

For enhancer reporter experiments, 100 * 10^3 cells were seeded in 6W plates (corning) in 1 mL culture medium. After 24h from seeding, cells were transfected with Lipofectamine 3000 (#L3000015 ThermoFisher Scientific) following manufacturer's instructions with the different enhancer reporter constructs. 12h after transfection, culture medium was refreshed. 48 hours after transfection, cells were collected for SNAPtag staining and images were acquired.

Generation of stable sh-SOX9 and tetO-SOX9 cell types was carried out by transduction at multiplicity of infection (MOI) 1, 3 or 10 with lentiviruses expressing either shRNA against human SOX9 or the inducible tetO-SOX9 construct. Lentiviral particles used to transduce TetO Fuw Sox9 and pLKO shSox9 were kindly gifted by Luca Fagnocchi.

Testing for RARα antibody specificity was performed through transient transfection of tIMEC cells with 75pmol RARα siRNA (#4392420 Life Technologies). Transfection was performed using Lipofectamine RNAiMAX transfection reagent (#13778075 Life Technologies) following manufacturer's instructions. Cells were fixed with 4% PFA and immunostained 48h after transfection.

3.4. Sphere-forming assay

Cells were plated in four technical replicates in 24W ultralow attachment plates (Corning) at a density of 2.5×10^3 viable cells/ml in 1:1 DMEM/F12 medium supplemented with insulin (Clonetics, MEGM SingleQuots #CC-4136), EGF (Clonetics, MEGM SingleQuots #CC-4136), hydrocortisone (Clonetics, MEGM SingleQuots #CC-4136), B-27 Supplement (Gibco # 17504044), 20 ng/ml human FGF-basic (PeproTech #100-18B) and 2% Matrigel (BD Biosciences #354230). Formed spheres were analysed

after 72h; images were acquired with ImageXpress Micro Confocal High-Content Imaging System (Molecular Devices), and analysed using the MetaXpress 6 software (Molecur Devices). Objects were segmented using mCherry fluorescent signal and spheres were defined as round object with an area>500 μm^2. Sphere forming efficiency (SFE) was calculated as the ratio between the number of spheres per well and the number of seeded cells, for at least three biological replicates.

3.5. Migration assay

2×10^3 cells were plated in medium without growth factors and placed onto Matrigel-coated (BD Biosciences #354230) transwells of 8-µm pore size (Corning #3422). In the lower part of the transwell the complete medium was placed as a chemo-attractant. The number of migrated cells on the bottom of the membrane was calculated up to 72 h by microscope observation: briefly, the upper part of the transwell was cleaned with a cotton swab after 4% PFA fixation of the membrane; then, cells on the lower part of the transwell were stained for DAPI and acquired at SP8 Confocal Microscope. Quantification of migration capacity was calculated as the percentage of migrated cells over the number of seeded cells, for at least four biological replicates.

3.6. Invasion Assay

Cells were plated as for short-term sphere-forming assay. After 72 hours from seeding, spheroids were embedded in pepsinized collagen type I from bovine skin diluted in growth factors depleted culture medium (final concentration 1.6 mg ml−1, PureCol, Advanced BioMatrix) with vehicle, 1µM ATRA (DBA #10-1138) or 1µM RARα antagonist BMS614 (Tocris #3660). Fibrillar collagen matrices containing spheroids were polymerized at 37 °C for 1 hour, and then covered with complete culture medium, as chemoattractant. Invasion capacity of cells was evaluated 72h after embedding of spheres, by staining with CellTrace™ Calcein Green, AM (Invitrogen #C34852) 1:2000

and acquiring six fields of view per cell line with 50-100 2.6µm stacks at SP8 Confocal Microscope, 10X objective with 2X zoom for at least three biological replicates. Quantification of the spheroid area and perimeter, the number of disseminated cells and migration distance was performed with ImageJ.

3.7. Flow Cytometry analysis (FACS)

Dye retention assay was performed with CellVue Maroon Cell Labelling Kit (Invitrogen #88-0870-16) on tIMEC, XD and MD cells following manufacturer's instructions. After the staining, cells were plated as standard with or without indicated ATRA treatment and acquired at FACS Canto A after four or eight days. Gating for each experiment was performed as described in the corresponding figure legends.

3.8. RNA sequencing and differential gene expression analysis

Cells were seeded at maintenance density four days before collection. Three biological replicates per condition were collected. Cells were directly lysed on plates, and tissue collected from primary lung tumours (PT) (Poli et al., 2018) was lysed with TRIzol (Thermo Fisher cat. #15596026), and total RNAs were extracted according to the manufacturer's instructions. Contaminating genomic DNA was removed by DNase (Qiagen cat. #79254) digestion. RNA quality and concentration were assessed using the 2100 Bioanalyzer (Agilent cat. #G2939BA) and the Qubit fluorometer (Thermo Fisher cat. #Q33226), respectively. RNA-seq libraries were prepared by using the TruSeq® Stranded Total RNA (Illumina #20020596) supplemented with Illumina Ribo-Zero plus rRNA Depletion Kit (Illumina #20037135) starting from 500 ng of total RNA.

Libraries were sequenced on an Illumina HiSeq 2500 with SE 50 bp. Quality of the reads was checked using FastQC and trimmed using Trimmomatic v (Bolger et al., 2014). Reads were aligned to the human genome hg19 using Tophat2 (Kim et al., 2013). Further

analysis was performed using HOMER (PMID: 20513432). In brief, Tag Directories were created for each of the three replicates per cell type and genes were annotated using analyzeRepeats.pl rna hg19 –rpkm. For the differential expression analysis, the same command was used with–raw instead of – rpkm , to count the unnormalized number of reads needed as an input for edgeR.

Differential expression analyses were undertaken using the edgeR v3.20.9 and limma v3.34.9 software packages (McCarthy et al., 2012; Ritchie et al., 2015). Lowly expressed genes were filtered with the filterByExpr function from edgeR. Compositional differences between libraries were normalized using the trimmed mean of log expression ratios method (TMM). All counts were then transformed to log2-CPM and differential expression was assessed using linear models and robust empirical Bayes moderated t-statistics with a trended prior variance. P-values were adjusted using the the Benjamini and Hochberg method to control the FDR below 0.05%. Peaks with absolute fold change > 0.5 were considered differential.

Gene ontology analysis was performed with EnrichR (https://maayanlab.cloud/Enrichr/) using default settings.

3.9. ATAC sequencing

ATAC-seq was performed as previously described (Buenrostro et al., 2015; Corces et al., 2017). Briefly, four biological replicates for each cell population were considered. 50000 cells for each replicate were harvested and washed once with PBS; nuclei were isolated together with homogenized tissues collected from primary lung tumours (PT) (Poli et al., 2018). Samples were resuspended in 50uL Tn5 transposase mixture: 1x Tagment DNA Buffer, 0.5uL Tagment DNA Enzyme (Nextera DNA Library Preparation Kit, Illumina). Transposition reaction was incubated at 37°C for 45 minutes, followed by DNA isolation using a Qiagen MinElute PCR purification kit (QIAGEN, Hilden, Germany). Construction of ATAC-seq libraries included an initial round of PCR in a total volume of 50uL using the NEBNext High-Fidelity 2X PCR Master Mix (New England Biolabs, MA, USA) with primers (0.5uM each) from (Buenrostro et al., 2015) with the following

thermal cycles: 5 minutes at 72°C, 30 s at 98°C, followed by 5 cycles [98°C for 10 s, 63°C for 30 s and 72°C for 60 s] and a final extension at 72°C for 5 minutes. PCR products were purified and size-selected using Agencourt AMPure XP beads (Beckman Coulter) (0.65x and 1.8x volume to remove long and short fragments respectively) and eluted in 50uL of EB (QIAGEN). To avoid over amplification of libraries which result in GC bias, 2uL of the eluted DNA were subjected to qPCR; amplification curves were analyzed and the optimal number of PCR cycles for each sample were estimated with cycle thresholds reaching 1/3 of the maximum fluorescent intensity. Upon selecting the cycle threshold, 45uL of the eluted DNA were subjected to a second round of PCR in a volume of 50uL with NEBNext High-fidelity 2x PCR master mix, respective primers (1.25uM each) and the following thermal cycles: 30 s hot-start at 98°C, followed by 7-13 cycles [98°C for 10 s, 63°C for 30 s and 72°C for 60 s] and a final extension at 72°C for 5 minutes. The libraries were purified by Agencourt AMPure XP beads (x1.5 vol.), eluted in 30uL Tris HCl 10mM pH 8 and quantified by qPCR using Power SYBR Green PCR Master mix (ThermoFisher). Libraries were sequenced on an Illumina HiSeq2500 with SE 50bp.

3.10. ATAC-seq data pre-processing and peak calling

Quality of the reads was checked using FastQC and trimmed using Trimmomatic v0.39 (Bolger et al., 2014). Trimmed.fastq files for each cell line were merged for subsequent steps. Reads were aligned to the human genome hg19 using BowTie2 (Langmead and Salzberg, 2012). SAM files were processed using samtools v1.10 (Li et al., 2009). Further analysis was performed using HOMER (Heinz et al., 2010). In brief, Tag directories for the merged files were created and peaks were called with callPeaks hg19 –peak size 500 –distance 1000 –L 0 –C 3.

3.11. Analysis of the chromatin accessibility data

Differential accessibility analysis was undertaken using the edgeR v3.20.9 and limma v3.34.9 software packages (McCarthy et al., 2012; Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, 2015). Since some of the samples showed high compositional bias, normalization factors were calculated on the filtered counts. For this, lowly accessible peaks were first filtered using the filterByExpr function in edgeR without recomputing the libsize (keep.lib.sizes=TRUE). Normalization factors were calculated with calcNormFactors using the TMM on the filtered object, and then assigned back to the original non-filtered object to obtain TMM-normalized counts (non-filtered). Final filtering was performed in the TMM-normalized counts using the filterByExpr function in edgeR. A mean-dependent trend was fitted to the negative binomial dispersions with the estimateDisp function and, and differential accessibility between all cell types was assessed using the quasi-likelihood (QL) framework of the edgeR package. P-values were adjusted for multiple testing using the Benjamini-Hochberg method. Peaks with a FDR below 0.05% and absolute fold change > 0.5 were defined as differentially accessible regions.

Peaks were annotated as Promoter, Downstream promoter, 5' UTR, 3' UTR, exonic, intronic, or intergenic using the assignChromosomeRegion function in the ChiPpeakAnno (Zhu et al., 2010) (v 3.18.2') package in R. Chromatin states were annotated using the ChIP-seq-defined ChromHMM states from the Roadmap Epigenomics Project, where 15 state models were downloaded from the chromatin state learning site for the 'E028 Breast variant Human Mammary Epithelial Cells' (E028) (https://egg2.wustl.edu/roadmap/web_portal/chr_state_learning.html).

Using the umap function in the umap R package(Konopka, 2020) (v0.2.7.0), tIMEC, XD and MD accessibility data were projected along with publicly available accessibility data on primary human tumours (Cancer Genome Atlas (TCGA)), human primary non-cancerous tissues from different locations (GSE165659) and claudin-low cell line MDA-

MB-231 and its metastatic subpopulations (Cai et al., 2020). In brief, we downloaded the TCGA_ATAC_PanCan_Raw_Counts file containing chromatin accessibility profiles of 410 tumour samples spanning 23 cancer types from The Cancer Genome Atlas as well as the pan-cancer peak calls list (TCGA_ATAC_PanCancer_PeakSet) containing the hg38 coordinates of a list of 562,709 transposase-accessible DNA elements from the GDC (https://gdc.cancer.gov/about-data/publications/ATACseq-AWG). GSE165659 and GSE129646 accessibility datasets were downloaded from Gene Expression Omnibus (GEO) and were pre-processed as per above.

To obtain the number of independent Tn5 insertions in each of the 562,709 peaks, each end of a fragment of BAM files were counted using "countOverlaps" from the GenomicRanges (v1.36.1) package(Lawrence et al., 2013) in R. We then built a RangedSummarizedExperiment object including peaks as GenomicRanges, a counts matrix, and metadata detailing information for each sample using the SummarizedExperiment (v1.14.1) package(Morgan M, Obenchain V, Hester J, 2020) in R. The counts matrix was then normalized using the cpm function of the edgeR (v3.26.8) package(Robinson et al., 2010) followed by a quantile normalization using normalize.quantiles function in preprocessCore (1.46.0) package(Bolstad, 2019) in R. For the tIMEC, XD and MD the hg19 coordinates were first converted to hg38 using liftOver from liftOver package (v1.8.0)(2019) and the hg19ToHg38 liftOver chain in R. For measuring the distance or similarity of chromatin accessibility between the tIMEC, XD and MD samples and each of the other samples in the study we measured the Spearman Rank-order correlation between of all samples. Samples were then ranked by their Spearman coefficients in descendant order such as each pairwise comparison was attributed a rank number where smaller the rank number the higher the similarity. For the intersection of the RNA, ATAC and HiChIP dataset we used the GenomicRanges (v1.36.1)(Lawrence et al., 2013) package in R.

3.12. Visualization of the data and plots

Plots were performed using the ggplot2 (v 3.3.3) package (Wickham, 2016) in R. Heatmaps were constructed using the coolmap function in the limma (v3.40.6) package in R (Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, 2015). Multi-dimensional scaling (MDS) plots were constructed using the plotMDS function in the limma v3.40.6 on the filtered and normalized log2-counts-per-million values for each library. The removeBatchEffect function of the limma package was used to correct for sequencing batch effect. RNA-seq and ATAC-seq coverage plots, and bed and bedpe files were plotted with the plotBedgraph, plotBed and plotBedpe functions, respectively, from the Sushi (v1.22.0) package (Phanstiel, 2019) in R.

3.13. HiChIP sequencing

HiChIP was performed as described (Mumbach et al., 2016) with some modifications. Briefly, ChIP protocol was replaced with the one reported in Crispatzu et al., 2021, using H3K27ac antibody (Abcam #ab4729), starting from $1*10^6$ cells per condition. Digestion and ligation were performed overnight (NEB T4 Ligase, #M0202 instead of Invitrogen T4, 15224-041) and DNA was extracted with phenol-chlorophorm. After biotin pull-down, beads were resuspended in 25 µL of 2X TD Buffer, 0.5uL of Tn5 (Illumina #1503861), and water to 50 µL. After transposition, beads were resuspended in 50 µL of Phusion HF 2× (New England Biosciences). Generally, 14 cycles were used for Tn5 Nextera PCR amplification (Illumina Nextera DNA UD Indexes Kit) with the following PCR program: 72 °C for 5 min, 98 °C for 1 min, then cycle at 98 °C for 15 s, 63 °C for 30 s, and 72 °C for 1 min. 50ul of PCR product were size-selected with 25ul Ampure XP Beads and separated on a magnet. The supernatant was incubated again with 90ul Ampure XP Beads. The supernatant was then discarded, beads were combined, washed with Ethanol and eluted in 30ul of Elution buffer. 1ul was loaded onto D1000 Screen Tape (TapeStation)

for a second size check, where library size and molarities were extracted. Then, qPCR was performed with standard Illumina primers in order to quantify the libraries. Finally, samples were pooled with different Index/Adapter sequences and sequenced on a NovaSeq 6000, first at shallow depth (1-5 million reads) to judge proximity ligation efficiency and ChIP signal, then paired-end deep sequenced with read lengths of 75 and 100 million reads.

3.14. HiChIP data pre-processing and loop identification

HiChIP sequencing reads were pre-processed, aligned and filtered using *HiC-Pro* (Servant et al., 2015). We specified the genome assembly (hg19) and the restriction fragment coordinates file for the enzyme used in the processing of the samples (DpnII), while all the other parameters were kept with default values. Filtered reads where then used as input to *FitHiChIP* (Bhattacharyya et al., 2019) for the identification of interacting regions. Since ChIP-seq peaks were not available for these experiments, we used ATAC-seq ones as reference to improve accuracy of loop calling. Loops were identified at different resolutions (bin sizes of 10 and 25 kb) and we used both peak to peak and peak to nonpeak loops for background modeling together with the coverage bias regression to assess the significance of identified interactions. We ultimately used interactions at 10 kb resolution for all the downstream analysis; 25kb resolution was used to intersect HiChIP bins with ATACseq peaks for looping *SOX9* enhancers identification.

3.15.HiChIP differential analysis, bin hubs and annotation

To identify changes in 3D conformation between the different samples, we used the script provided by *FitHiChIP* (*DiffAnalysisHiChIP.r*) to compare tIMEC vs XD, XD vs MD and finally tIMEC vs MD. As for the identification step, we used ATAC-seq reads as reference and all parameters were kept with default values.

Significant loops (FDR ≤ 0.01) were compared to evaluate the number of loops identified among samples. To this end, we used Unix shell scripting to compare and count the interactions either specific to each sample or identified in two or all three of them.

To evaluate the presence of bin hubs in our three samples, we took all bins included in the loops individually and counted the number of times each was included into a significant interaction.

To assess whether the significant loops were bringing into special proximity promoter regions with regulatory element, we annotated each interacting bin. Promoter regions were defined as 2.5kb upstream up to 100bp downstream of the transcription start site (TSS), which were obtained from Ensembl (GRCh37 v. 103). Regulatory elements were defined as ATAC peaks that do not overlap a promoter region within a window of 1.5 kb. Bin annotation was then performed using the *intersect1D* and *intersect* functions from *pgltools*. (Greenwald et al., 2017)

3.16. Integrated analysis of motif activity and gene expression changes of transcription factors (IMAGE)

IMAGE is a recently developed tool that allows precise prediction of causal transcription factors based on transcriptome profiling and genome-wide maps of enhancer activity (Madsen et al., 2018). Briefly, IMAGE searches for motifs using extended PWM database with P-value-based cut-offs using HOMER (Heinz et al., 2010). Subsequently, motifs without any hits in the supplied sequences and motifs mapping to transcription factors with low expression are removed. To predict target enhancers, IMAGE performs ridge regression. The motif matrix is centered, and the user-supplied enhancer activity matrix of a normalized tag is centered and scaled. The enhancer activity at a specific position in a particular sample is given by the sum of all motif activities multiplied by their motif frequency at that site. Target enhancers are identified by leave-one-out-based analysis. IMAGE defines target enhancers of each motif as sites where the motif is present and where the accuracy of the IMAGE model decreases upon leaving out that motif of the analysis. IMAGE uses the predicted sites to calculate motif activities for gene expression using an integrated model of enhancers using a similar ridge regression scheme. Target genes are identified by leave-one-out-based analysis: IMAGE calculates a P-value-like score based on the drop in prediction accuracy decreases upon leaving out that motif, as well as the predicted presence of binding sites near the gene. Genes with a score below 0.005 that are differentially regulated, as well as expressed above 1 normalized reads per kb, are defined as target genes. To analyze the statistical significance of transcription family enrichment, a hypergeometric t-test was performed. All TFs included in the IMAGE dataset which were considered a true TF by (Lambert et al., 2018)were used as input (#840). Identified TFs were subdivided by TF family and the enrichment of the number of TFs of each family identified in the given dataset was calculated using the hypergeometric test.

For TF network analysis, target genes of transcription factors assigned with causal motifs identified by IMAGE were retrieved and selected for statistically significance of the target (trueTarget = 1). True targets of each motif were merged with the list of true transcription factors defined by (Lambert et al., 2018), expressed in MD cells and showing a higher activity in MD compared to XD and tIMEC samples. If one TF was assigned multiple motifs with causality, the lists of true target genes were merged and duplicates were removed. Network was drawn using Cytoscape 3.9.0.

3.17.RNA extraction and expression level quantification

Cells were seeded as for maintenance and eventually treated with 1μM RARα antagonist BMS614 (Tocris #3660) or with 1μM ATRA (DBA #10-1138) as indicated in the relative figure legends. Total RNAs were extracted from log-phase cells with TRIzol (Ambion #15596018), according to the manufacturer's instructions. Quantitative real-time PCR analysis was performed with SuperScript III One-Step SYBR Green kit (Invitrogen #11746) and C1000 Touch termocycler - CFX96 Real Time System (Biorad). Relative gene expression levels were determined using comparative Ct method, normalizing data on endogenous GAPDH. Primers used in this thesis are: *SOX9* Fwd: GTACCCGCACTTGCACAAC Rev: TCTCGCTCTCGTTCAGAAGTC GAPDH Fwd: AGGTGAAGGTCGGAGTCAAC Rev: CCATGTAGTTGAGGTCAATGAAG CXCR4 Fwd: TGGAACCGATCAGTGTGAGT Rev: GGGCAGGAAGATCCTATTGA RARα Fwd: CCCCTGGAGATGGATGATGC Rev: TCCGCACGTAGACCTTTAGC

3.18.Protein extraction and Western Blot analysis

Total protein extracts were obtained as follows. Cells were washed twice with cold PBS, harvested by scraping in 1 ml cold PBS and centrifuged for 5 min at 1500 rpm. Harvested cell pellets were lysed by the addition of 5× v/v ice-cold RIPA buffer (50 mM Tris-HCl pH 8.0, 150mM NaCl, 1% NP-40, 0,5% Na-deoxycholate, 2% SDS) plus protease and phosphatases inhibitors 30 min at 4 °C. Lysates were cleared by centrifugation for 10 min at 14,000 rpm at 4 ° C and supernatant was collected on ice. Protein concentration of lysates was determined using PierceTM BCA Protein Assay Kit 24 (Thermo Scientific, 574 #23227), according to the manufacturer's instructions. The absorbance was measured at λ = 595 using SAFAS spectrophotometer (SAFAS, Monaco). Values were compared to a standard curve obtained from the BSA dilution series. For western blots analysis, 40 µg of protein samples were boiled and then subjected to SDS-PAGE in a 12% polyacrylamide gel. After electrophoresis, proteins were transferred to a nitrocellulose membrane. Membranes were blocked in PBS-Tween containing 5% milk for one hour at RT with constant agitation and incubated with indicated primary antibody O/N at 4 °C with agitation. The membrane was then washed three times with PBS-Tween, each time for 5 min, followed by incubation with secondary antibody HRP-conjugated for one hour at RT. ECL reagents (GE Healthcare #RPN2232) were used to initiate the chemiluminescence of HRP. The chemiluminescent signal was captured using LAS3000 system (GE Healthcare). Primary antibodies used are as follows: SOX9 (#AB5535 Millipore) 1:1000; GAPDH (#sc-32233 Santa Cruz) 1:1000. Relative optical density was quantified with ImageJ Software.

3.19.Immunofluorescence

tIMEC, XD and MDs were kept in culture from 72 to 96 hours or treated with 1µM ATRA (#10-1138 Focus Biomolecules) as indicated. Cells were then fixed for 15 min at

room temperature with 4% paraformaldehyde (Sigma-Aldrich #158127) and washed three times with PBS, and subsequently trypsinized in order to disaggregate spheres; trypsin was then inactivated with 10% FBS PBS, and disassociated cells were counted. 100 *10^3 cells were used for each in-suspension staining. Staining was then performed in 96W U-bottom plates (corning) according to the following conditions: permeabilization and blocking with PBS/5% Goat serum (#11475055 Fisher Scientific)/1% BSA (#126579 Millipore)/0.3% Triton X-100 (blocking solution) for one hour at room temperature in constant agitation, followed by incubation with primary antibody (diluted in the blocking solution) in agitation overnight at 4°C, three washes in PBS 1X and incubation with secondary antibodies and DAPI (diluted in the blocking solution) for 1 hour at room temperature, in constant agitation. Pelleted cells were then resuspended in 5µL mounting medium (Fluormount G # 00-4958-02 Life Technologies) and spotted on support glasses; spotted cells were then covered with coverslips and left to dry in the dark overnight. Images were acquired using a Leica SP8 confocal microscope with HCX PL APO ×63/1.40 objective. In cases where image analysis was performed, image acquisition settings were kept constant. Primary antibodies are as follows: SOX9 (#AB5535 Millipore) 1:250, Ki67 (#KI67-MM1-L-CE Leica Microsystems) 1:1000, RARα (#NB200-322 Novus) 1:250, RXRα (#ab125001 Abcam) 1:250, RARγ (#HPA053883 Sigma) 1:100, RXRβ (#HPA063653 Sigma) 1:100, p-p38 (#44-684G ThermoFisher) 1:500; Secondary antibodies were goat-anti-mouse or -rabbit coupled to Alexa-488 or -647 (Invitrogen) and used 1:500.

SNAPtag staining was performed using Janelia Fluor SNAPtag substrate (#6419 Tocris) following manufacturer's protocol.

3.20.Confocal imaging data analysis

Confocal imaging data analyses were performed using ImageJ software. For 2D analysis, DAPI DNA dye was used to identify the nucleus and define the region of

interest. The fluorescence intensity and physical parameters were determined. The values of the fluorescence intensity were background subtracted.

To quantify the nuclear mean intensity of each staining, LIF files were converted to TIFF multichannel images, and then the following macro for ImageJ was applied:

```
function DAnalyze(input, output, filename, lothresh, hithresh) {
open(input + filename);
run("Duplicate...", "title=[TO MEASURE] duplicate");
run("Duplicate...", "title=C_nuc duplicate channels=1");
run("Median...", "radius=4");
setAutoThreshold("Default dark");
run("Threshold...");
setThreshold(lothresh, hithresh);
setOption("BlackBackground", true);
run("Convert to Mask");
run("Fill Holes");
run("Set Measurements...", "area mean integrated skewness area_fraction limit display redirect=[TO MEASURE] decimal=2");
run("Analyze Particles...", "size=size range display clear add");
close("Results");
selectWindow("TO MEASURE");
roiManager("Show None");
roiManager("Show All");
Stack.setChannel(x);
roiManager("Measure");
saveAs("Results", output + filename + ".csv");
close();
close();
close();
}
input = "PathInput/";
output = "PathOutput/";

lothresh = lower threshold;
hithresh = higher threshold;
setBatchMode(true);
list = getFileList(input);
for (i = 0; i < list.length; i++){
```

```
        DAnalyze(input, output, list[i], lothresh, hithresh);
    }
    setBatchMode(false);
```

The hitresh, lotresh, and size range parameters were determined manually for each set of images.

To quantify the nuclear to cytosolic localization of retinoic acid receptors, we adapted a MATLAB routine that is deposited in GitHub (https://github.com/SZambranoS/RoutinesNucCytoYAP). In short, images of the DAPI and RARs channels were saved as 16-bit TIFF files. To segment the nuclei, we used the signal from the DAPI channel. The nuclear masking was performed using as a threshold the mean intensity of the image plus twice the standard deviation. After thresholding, segmentation was carried out after a watershed transformation, so most of the few overlapping nuclei could be separated. The segmented nuclei were filtered for their size to exclude artefacts or improperly segmented clusters of nuclei. To estimate the average cytosolic intensity per cell, a ring of 30-pixel width (approximately 7 µm) around each segmented nucleus was found. Pixels of the ring with too low intensity of the RARs signal (below twice the value of the background signal) were discarded. The average cytosolic signal for each cell was the average intensity of the remaining pixels. We then calculated for each cell the nuclear to cytosolic intensity as the ratio of the background-corrected nuclear and cytosolic average RARs intensities.

2D cluster analyses were performed using ImageJ software. For the 2D cluster analysis of RARα and RXRα, background subtraction was applied. To calculate the background level, the mean of the minimum level of intensity of the specific staining was calculated among all of the nuclei analysed. Then, unsharp masking and median filters were applied. The clusters were identified with the Yen automatic threshold and the intensity of each cluster was calculated, redirecting the measurement on the unmodified images.

3.21.Luciferase Assay

Enhancer activity was determined by luciferase reporter assays. IMEC WT, tIMEC, MDA-MB-231 and MCF7 cells were nucleofected with 1ug of either the enhancer luciferase reporter plasmid (pGL4_enhancer), the Wnt pathway luciferase reporter (pGL4_TCF) or the empty vector (pGL4.27[luc2P/minP/Hygro]) and the pGL4-CMV-Renilla luciferase vector as a normalization control in a 30:1 ratio. $1x10^6$ cells were nucleofected using an Amaxa Nucleofector (program U-23, Lonza) and homemade buffer (KCl 5 mM, MgCl2 15mM, Glucose 1M, K2HPO4 120mM). After 24 hours, Firefly and Renilla luciferase activity was measured using the Dual-Luciferase® Reporter Assay System following manufacturer's instructions.

3.22.Statistical Analysis

Statistical parameters are reported in figure legends and include: number of replicates analysed (n), dispersion and precision measures (mean ± SEM) and statistical significance (p-value). Data have been statistically assessed by one-tailed Student's t-test; for multiple comparison, statistical analysis was conducted by one-way ANOVA followed by parametric or non-parametric post hoc test according to the distribution of the data. The use of the specific tests has been reported in each figure legend. In figures, asterisks mean *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ns = not significant. $p < 0.05$ and lower were considered significant.

4. Results

In this section of the thesis, I present the results I obtained during my doctoral studies, that have been carried out at the Department of Cellular, Computational and Integrative Biology (CIBIO, University of Trento) in the Laboratory of Chromatin Biology and Epigenetics led by Prof. Alessio Zippo, in collaboration with Dr Alvaro Rada-Iglesias, who leads the laboratory of Developmental Genomics at the Centre for Molecular Medicine Cologne (CMMC), University of Cologne, Germany, where I spent a research period abroad. Of note, as this project required a multidisciplinary approach encompassing molecular biology, cell biology bioinformatics and cancer cell biology, some of the experiments and data analyses were performed in tight collaboration with other members of the team: Dr Sven Beyes and Dr Naiara Garcia Bediaga performed the bioinformatic analysis of genome-wide experiments including ATAC-seq, RNA-seq and IMAGE; Dr Martina Dori performed the bioinformatic analysis of HiChIP data. Dr Luca Fagnocchi prepared sequencing libraries for ATAC-seq and RNA-seq analysis.

4.1 A xenograft-derived model for basal breast cancer

In order to study enhancer activity during the steps of tumorigenesis, I characterized the model developed in our lab (Poli et al., 2018). Since this model is derived from mammary epithelial cells and is characterized by a genetic background that reflects TNBC traits, we used this model to address our main research questions. In order to assess whether the metastatic derived cell line, in particular, would resemble the phenotypic characteristics of TNBC cells, we proceeded to characterize their sphere-forming capacity, migration and invasion capacity compared to the primary tumour derived cells and tumorigenic IMEC.

4.1.1. Phenotypic characterization of the model

In order to characterize the model system on which this work is based, we tested the clonogenic potential, migration and invasion capacity of tIMEC, XD and MD cells, reflecting different stages of tumour progression.

tIMEC cells are a non-clonal population of immortalized mammary epithelial cells, in which c-Myc was overexpressed, together with a mutated form of PI3KCA (PI3KCA ^{H1047R}).

XD cells are cells derived from xenografts of primary tumours obtained upon xenograft transplantation of tIMEC into the fad pad of the mammary gland of NOD/SCID mice. We were able to obtain multiple populations of XD cells: cells from 4 independent tumours were tested for most experiments, however, since characteristics of the different cell populations did not show any statistically significant difference (**Figure 25** and **Figure 26**), one representative population was chosen for the subsequent experiments.

MD cells refer to a population of metastasis-derived cells derived from liver metastasis developed from mice four week after surgical resection of the primary tumors obtained upon xenograft transplantation of tIMEC cells.

Metastatic cells possess advantageous traits that may originate in the primary tumour but continue to evolve during dissemination and colonization; specifically, the hallmarks of metastasis can be identified in four distinguishing features: motility and invasion, ability to modulate the secondary site or local microenvironments, plasticity, and ability to colonize secondary tissues (Celià-Terrassa and Kang, 2016; Welch and Hurst, 2019).

Sphere-forming assays are an *in vitro* technique to assay both normal and neoplastic cells for growth capacity in 3D settings that resemble more the primary tumour tissue microenvironment. Spheroids formed by tIMEC, XD and MD cells were analysed after 72 hours from seeding, at clonal density, in ultra-low attachment conditions. The

calculation of the sphere-forming capacity (SFE), which represents the percentage of the seeded cells which were able to form spheroids, indicated that there is not any statistically significant difference in the sphere-forming capacity of the three cell types, showing a high SFE (>40%) for all three cell types analysed (**Figure 12)**.

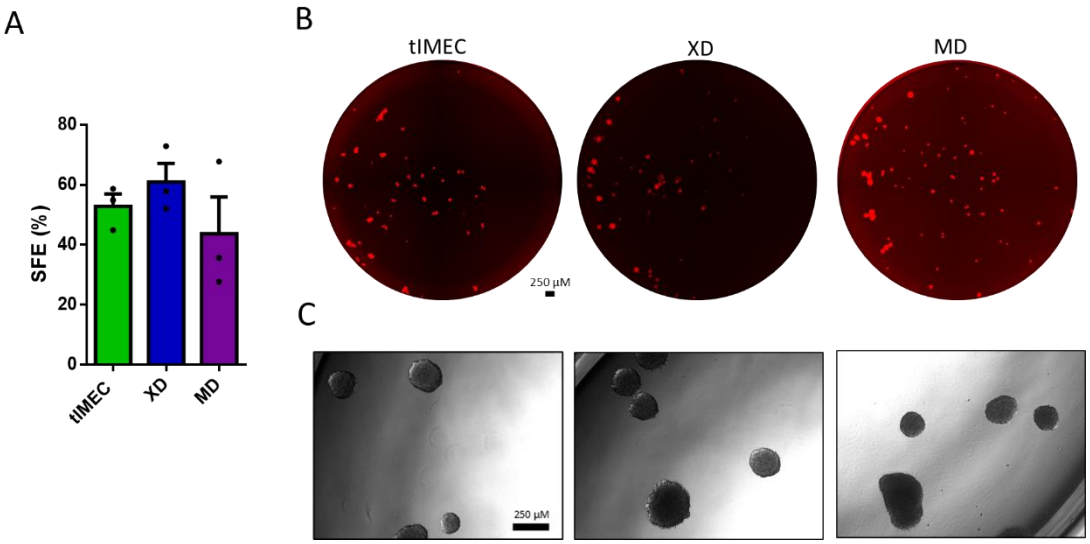


Figure 12 Sphere-forming assay of tIMEC, XD and MD cells. *A) Sphere-forming efficiency of the indicated cell types. AVG of 3 biological replicates +/- S.E.M. One-tailed unpaired student's t-test. Ns:P > 0.05; B) Representative images of tIMEC, XD and MD cells spheroids mCherry signal at 72h from seeding; 4X objective; scale bar: 250 µm. C) Representative brightflied images of tIMEC, XD and MD cells spheroids at 72h hours from sseding; 10x objective; scale bar: 250 µm.*

In order to quantify the motility of tIMEC, XD and MD cells towards a chemo-attractive agent, a transwell migration assay was performed. Specifically, we monitored and quantified the number of cells that migrated through Matrigel-coated 8µm pores towards a chemoattractant after 24, 48 or 72 hours from seeding, as shown in **Figure 13**. The obtained results indicated that XD and MD cells are characterized by a higher percentage of migrated cells with respect to seeded cells in comparison with tIMEC, especially at later time points; MD cells, in particular, were identified by the highest migration capacity at 72h after seeding.

Invasion of surrounding tissues is generally considered to be a key feature of malignant tumours. For some cancers, in particular, it is a cause of severe morbidity and can be life-threatening even in the absence of distant metastases. Therefore, there is an opportunity to target the invasion process to provide novel therapies that could be complementary to standard anti-proliferative agents. Until recently, this strategy has

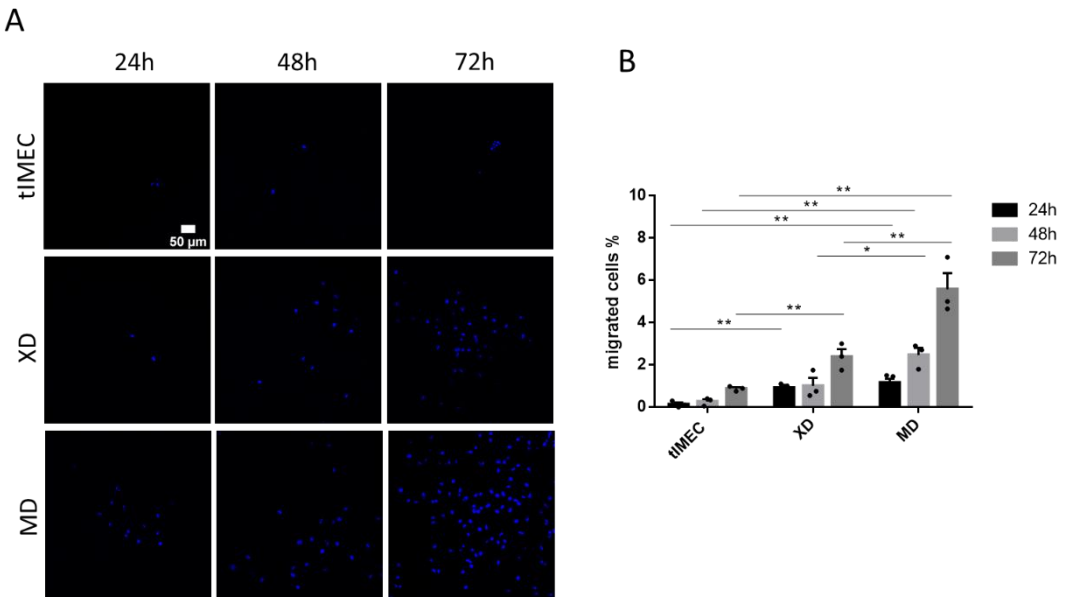


Figure 13 Migration assay of tIMEC, XD and MD cells at different time points. *Migration assay showing the capacity of tIMEC, XD and MD cells to migrate through Matrigel-coated 8 µm pores after 24, 48, or 72 hours. A) Representative 20X confocal images of membrane bottom attached tIMEC, XD and MD DAPI signal; scale bar: 50 µm. B) % of cells attached to the lower part of the membrane relative to the number of seeded cells. AVG of 3 biological replicates (N=3, +/- S.E.M. One-tailed unpaired student's t-test. Ns:P > 0.05; *:P ≤ 0.05; **:P ≤ 0.01.*

been hindered by the lack of robust, reproducible assays suitable for a detailed analysis of invasion. In recent times, quantitative assays to assess the invasive capacity of cells were established (Vinci et al., 2015). Hence, to quantify the capacity of these cells to invade the surrounding environment, moving through the extracellular matrix towards a chemo-attractant, we applied a three-dimensional tumour spheroid invasion assay (Vinci et al., 2015).

Images of spheroids formed by tIMEC, XD and MD cells when grown in ultra-low attachment conditions for 72 hours and subsequently embedded in a collagen matrix were acquired, and quantification of the spheroids morphological traits was possible after staining with Calcein AM. Through imaging, we assessed that the three cell types were able to form spheroids (**Figure 14**A) when cultured in low-attachment conditions, and when embedded in a growth-factors deprived collagen matrix, these spheroids induced the formation of protrusion to reach a chemoattractant-rich area of the culture plate (**Figure 14**A). Furthermore, cells detached from the spheroids and moved through the matrix, migrating towards chemoattractants as single cells (**Figure 14**A). When comparing the three cell types, quantification allowed us to assess that the area of the starting spheroid did not show any statistically significant difference (**Figure 14**B).

Nevertheless, the spheroid perimeter was significantly increasing in metastatic cells when compared to transformed cells and primary tumour xenograft derived cells (**Figure 14**C), suggesting that MD cells induced the formation of a higher number of protrusions. We were able to quantify single cells that detached from the spheroids and moved through the collagen matrix to a certain distance from the starting point: MD cells presented a significantly higher number of detached cells (**Figure 14**D) that were also able to migrate further away from the original spheroids when compared to XD cells and tIMEC (**Figure 14**E).

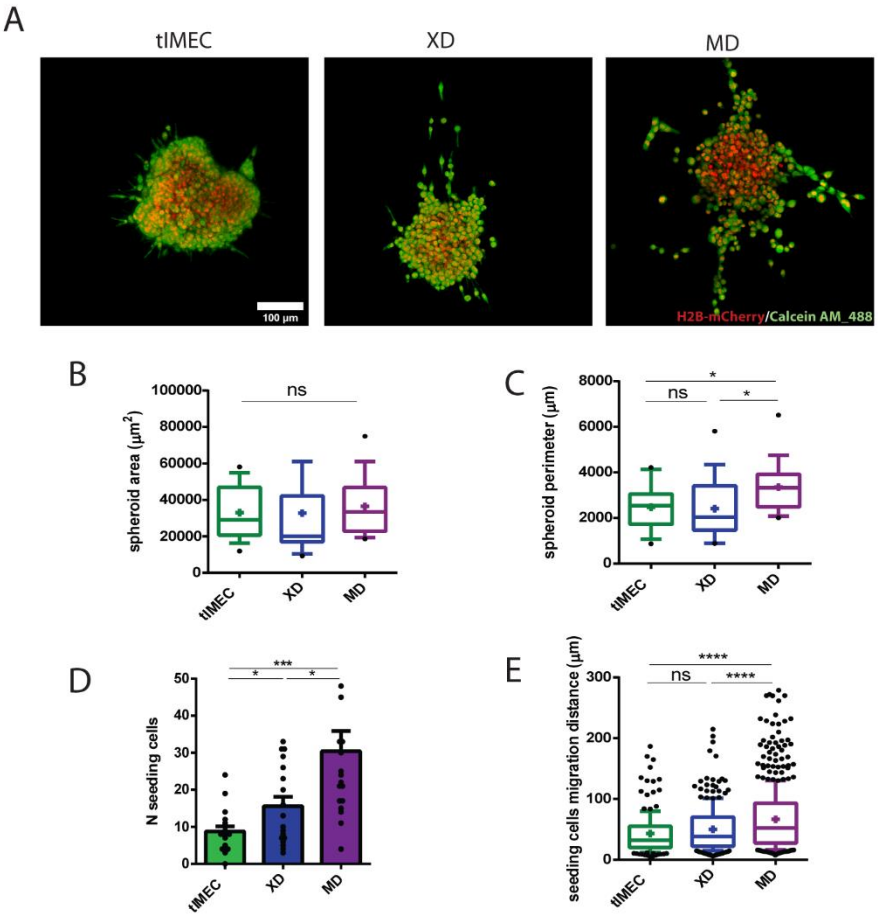


Figure 14 Invasive capacity of tIMEC, XD and MD cells. *A) Confocal 3D imaging of tIMEC, XD and MD spheroids embedded in a collagen growth-factor depleted matrix. Cells are invading the spheroid-surrounding matrix to reach a chemo-attractant rich area. Representative images of 3 independent experiments. Maximum intensity projection of 50-100 2.6µm Z-stacks acquired with a 10X objective, 2x zoom on a SP8 confocal microscope. Calcein AM (green); H2B-mCherry (red) B) Quantification of spheroid area. C) quantification of spheroid perimeter; higher perimeter is linked to higher number of protrusions. D) quantification of seeding cells (i.e. single cells detached from the originary spheroid mass and migrated towards chemoattractant); merge of analysis on 18 spheroids per condition, retrieved from 3 biological replicates; AVG +/- S.E.M. E) quantification of distance between seeding cell and originary spheroid mass. 3 biological replicates, student's t-test; Ns:P > 0.05; *:P ≤ 0.05; **:P ≤ 0.01;***:P ≤ 0.001;****:P ≤ 0.0001. The box plots in B), D) and E) indicate median values (middle lines), first and third quartiles (box edges) and 10th and 90th percentiles (error bars) of respectively area and perimeter of the analysed spheroids and distance of seeding cells from analysed spheroids.*

These results indicated that the model with which we are working reflects some of the distinguished features characterizing the steps of tumorigenesis. In particular, MD cells harbored (at least in vitro) the typical characteristics of metastatic cells: an intrinsic capability of forming spheroids, ability to migrate through small pores and invasion of surrounding space.

4.1.2. Epigenetic characterization of the model

Alteration of the epigenome has been revealed as a significant driver of oncogenic reprogramming towards enhanced cell plasticity in cancer progression (Wainwright and Scaffidi, 2017). Furthermore, epigenetic alteration has been shown to favour cancer cell survival during metastatic dissemination and adaptation to the microenvironments of the metastatic niche (Khan et al., 2021).

Thus, to understand the genetic and epigenetic background of our xenograft-derived cellular model, we proceeded to analyse gene expression profiling, chromatin accessibility and interactome profiling.

4.1.2.1. Chromatin accessibility and gene expression profiling individuates de-novo enhancers in MD cells

Thanks to the lab members in which I performed my PhD studies, Dr Luca Fagnocchi, Dr Sven Beyes and Dr Naiara Garcìa Bediaga, we performed and analysed gene expression profiling (RNA-seq) and the Assay for Transposase-Accessible Chromatin (ATAC-seq). Dr Fagnocchi prepared sequencing libraries; Dr Beyes and Dr Bediaga performed the bioinformatic analysis of the sequencing results.

ATAC-seq is a method for mapping chromatin accessibility genome-wide. This method assesses DNA accessibility with Tn5 transposase, which allows for tagging accessible regions of chromatin with sequencing adapters. Sequencing reads can then be used to extrapolate regions of increased accessibility and map nucleosome positions and regions of transcription factor binding (Buenrostro et al., 2015). In particular, we

aimed to assess differences in the chromatin landscape in our model and identify putative regulatory elements that could contribute to gene expression control.

We profiled chromatin accessibility in tIMEC, XD and MD cells, as well as in primary tumors (PT), which have been retrieved upon xenotransplantation of tIMEC into the fat pad of the mammary gland of NOD/SCID mice. The assessment of the diversity of the chromatin accessible regions in these cell populations through principal component analysis showed that chromatin accessibility of MD cells (**Figure 15**A) was much distinct in respect to the other three samples. Of importance, XD cells and primary tumour tissue (PT) clustered closely, suggesting that the brief in vitro amplification of the xenograft-derived sample did not affect the chromatin state. Cluster analyses of the differentially enriched regions showed that MD were characterized by the most distinctive accessibility pattern that distinguish them from the other cell populations (**Figure 15**B). Indeed, this analysis highlighted that MD were characterized by gaining chromatin accessibility at certain regions (reflected in clusters 1 and 3) as well as by losing accessibility in a distinct group of regions (clusters 2 and 4). Of note, the tIMEC-enriched regions showed a progressive reduction of chromatin accessibility, possibly indicating that these cis-regulatory elements are important to maintain cell identity and/or counteracting tumor progression (cluster 4). The XD and PT-enriched regions (cluster 1 and 2) showed a diverse behaviour in the MD: one group of regions resulted further-enriched (cluster 1) while a second group showed a reduction in chromatin accessibility, similarly to the level retrieved in tIMEC (cluster 2).

Annotation of the genomic features and ChromHMM states revealed that ATAC-seq peaks were enriched for promoters, (**Figure 15**C left), active TSS and enhancer chromatin states (**Figure 15**C right), in all the analysed cell populations. This analysis suggested that observed differences in the enriched chromatin accessibility regions among the cell types are not linked to a diverse genomic localization.

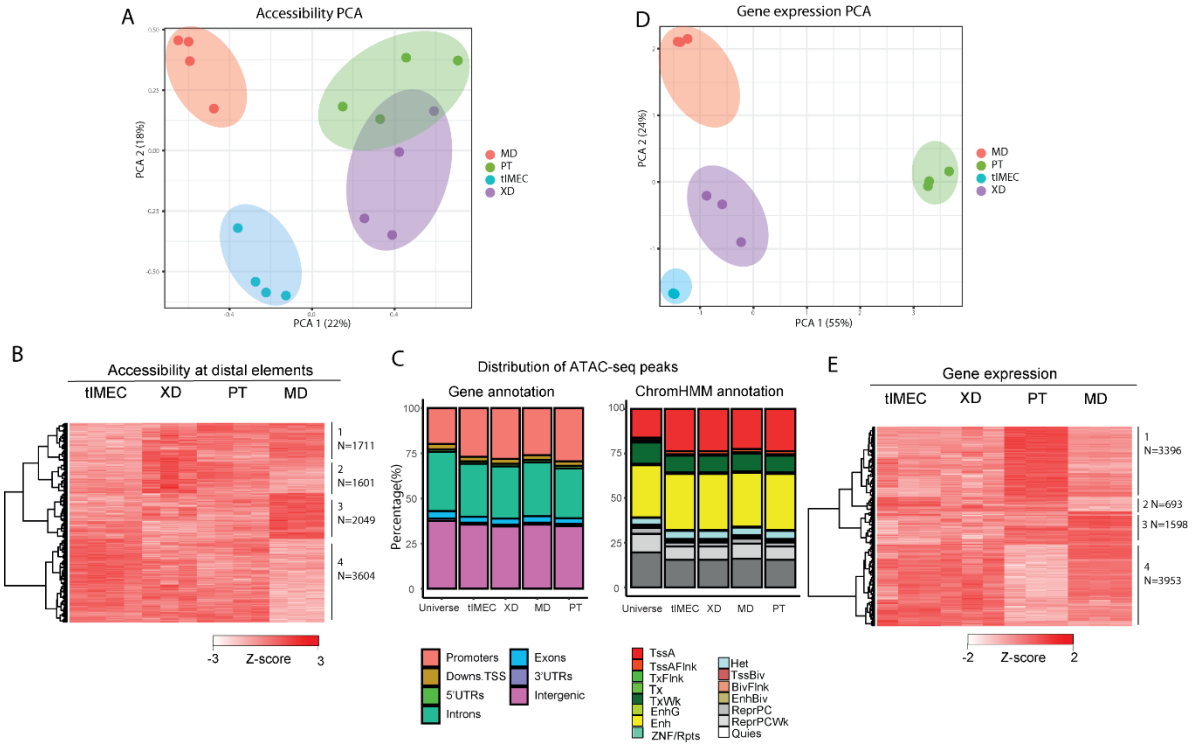


Figure 15 Characterization of chromatin accessibility and gene expression signatures in tIMEC, XD and MD. *A) Principal component analysis (PCA) plots of log-CPM values with samples coloured by cell type. The log-CPM values were corrected for sequencing batch. Distances on the plot correspond to the leading fold-change, which is the average (root-mean-square) log2-fold-change for the 5,000 most divergent peaks between each pair of samples. B) Heatmap showing chromatin accessibility of differentially accessible regions (DAR) for the MD vs. tIMEC, MD vs. XD, and tIMEC vs. XD and PT vs XD comparisons (Log2 transformed counts per million, log2CPM). 8965 DARs are represented. DARs number in each of the 4 major clusters are annotated. C) Barplot showing the distribution of genomic features and chromHMM annotations across peaks called in tIMEC, XD, MD and PT. ChromHMM annotation abbreviations are as follows: Row abbreviations as follows: TssA – active TSS, TssBiv – bivalent/poised TSS, EnhBiv – bivalent enhancer, ReprPC – repressed polycomb, ReprPCWk – weak repressed polycomb, Quies – quiescent/low, TssAFlnk – flanking active TSS, TxFlnk – transcription at gene 5' and 3', Tx – strong transcription, TxWk – weak transcription, EnhG – genic enhancers, Enh – enhancers, ZNF/Rpts – ZNF genes and repeats, Het - heterochromatin. D) Principal component analysis (PCA) plots of log-CPM values with samples coloured by cell type. The log-CPM values were corrected for sequencing batch. Distances on the plot correspond to the leading fold-change, which is the average (root-mean-square) log2-fold-change for the 500 most divergent genes between each pair of samples. E) Heatmap showing gene expression (Reads Per Kilobase of transcript, RPKM) of differentially expressed genes (DEG) for the MD vs. tIMEC, MD vs. XD, and tIMEC vs. XD and PT vs XD comparisons. 9640 DEGs are represented. DEGs number in each of the 4 major clusters are annotated.*

To evaluate the possible contribution of chromatin changes to the gene expression, we performed RNA-seq analyses on the same group of samples. Principal component analysis (**Figure 15**D), indicated that on the basis of component 1, which accounts for 55% of the differences, the primary tumour sample is significantly separated from tIMEC, XD and MD cells, which cluster closely on the basis of principal component 2. This may indicate that *in vitro* culture of xenograft-derived samples is partly introducing some modifications to the transcriptional landscape of our model. This was confirmed also when analysing the differentially expressed genes (DEGs) (**Figure 15**E): the primary

tumour sample is characterized by the highest number of positively (3396, cluster 1) and negatively (3953, cluster 4) differential genes in comparison to the three cell populations. Nevertheless, the cluster analyses showed that the MD cells are characterized by a distinctive gene expression pattern, with 693 depleted (cluster 2) and 1598 enriched (cluster 3) DEGs, with respect to other samples. In order to investigate what processes were involved in each of the enriched cluster, Gene Ontology (GO) analysis was performed (**Figure 16**). GO analysis underlined the enrichment, in cluster 1, of pathways involved in the interaction with the microenvironment, such as BM organization, collagen biosynthesis and ECM assembly. As this cluster corresponds to genes differentially expressed in the primary tumour tissue, this indication is in line with the fact that, contrarily to *in vitro* culture of the other cell populations analysed, this samples were retrieved from mice. GO analysis of cluster 4, which identifies genes whose expression is depleted in the primary tissue, highlighted metabolic processes related to mitochondrial activity and oxidative processes as the most enriched; this is in line with recent evidences showing the preferential production of energy through glycolysis than mitochondrial respiration in cancer (Gubern et al., 2016; Sajnani et al., 2017). Finally, cluster 2, which identifies genes depleted in MD cells, resulted to be enriched for cell-adhesion, MET and cytoskeleton related pathways, whereas analysis of cluster 3, which includes MD-enriched genes, revealed enrichment for pathways associated with migration, polarity and retinol metabolic processes.

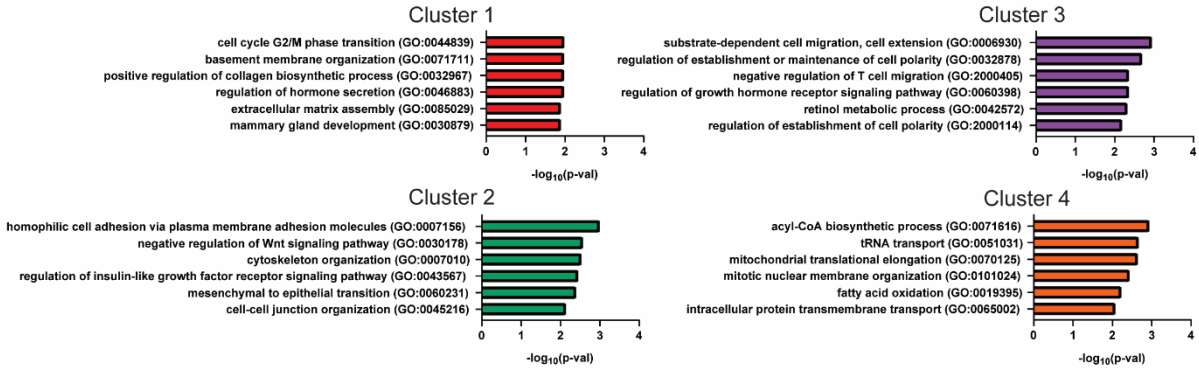


Figure 16 Gene ontology on differential gene expression clusters. *Gene ontology showing the top six statistically significant (p-val < 0.05) biological processes enriched in each differential gene expression cluster.*

With these results, although a characterization of histone marks distribution is still missing, we could state that the chromatin landscape of tIMEC, XD and MD cells is significantly different. Enrichment for chromatin differentially accessible regions

suggests that MD cells are characterized by the modulation of the activity of multiple regulatory elements.

This led us to ask whether the enrichment for regulatory elements activity may represent the metastatic cell state and if it would generally reflect basal breast cancer traits in patients.

4.1.2.2. Regulatory elements are conserved in basal BC patients

Since metastatic datasets are not available at this time, we aimed at assessing whether tIMEC, XD and MD cells epigenetic landscape would reflect the traits of primary cancer tissues, with a focus on basal breast cancer patient tissues. In order to investigate this, we performed clustering analysis on publicly available accessibility datasets representing chromatin accessibility of 410 tumour samples spanning 23 cancer types, non-cancerous tissues from different locations and the claudin-low cell line MDA-MB-231 and its lung, brain and adrenal gland metastatic subpopulations (Cai et al., 2020; Corces et al., 2018; Zhang et al., 2021). Accessibility data of tIMEC, XD and MD were projected along with publicly available accessibility data (**Figure 17**A); Uniform Manifold Approximation and Projection (UMAP) plot of tIMEC, XD and MD cells showed that their accessibility data cluster along with the other datasets.

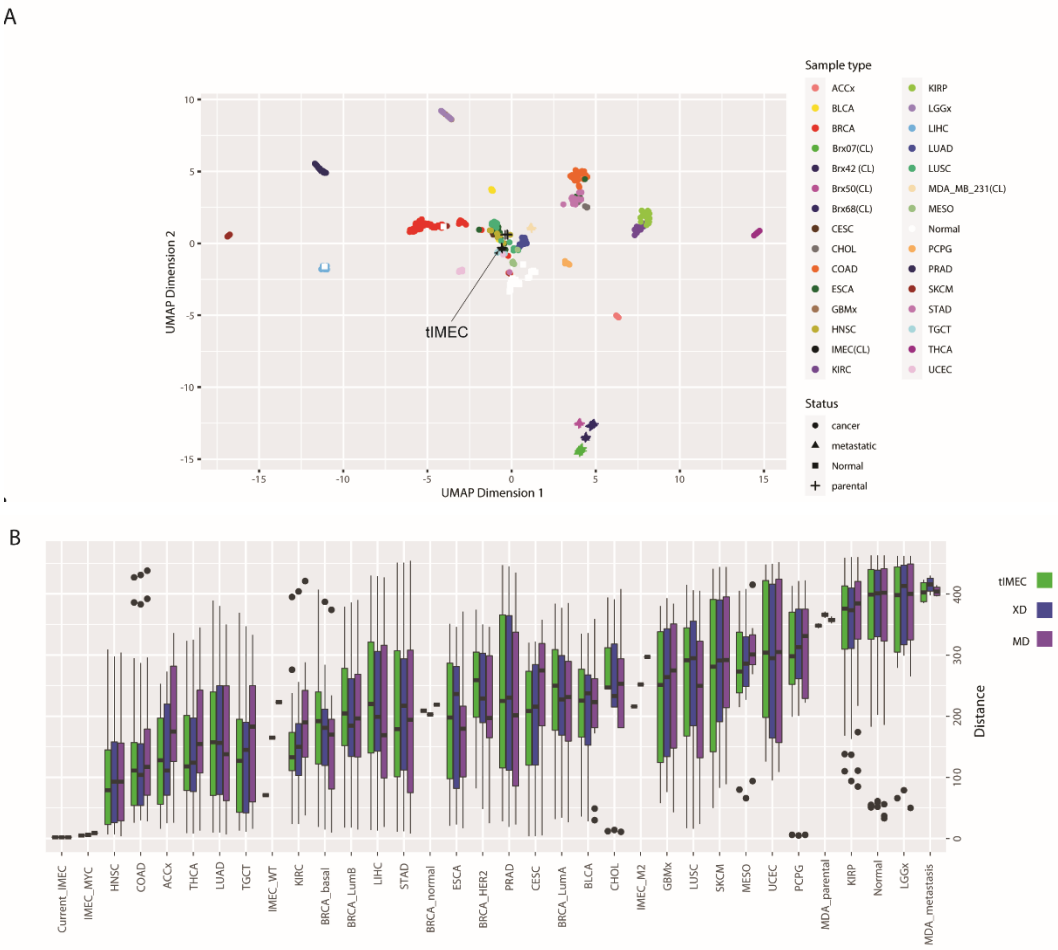


Figure 17 Cluster analysis on Pan Cancer chromatin accessibility dataset. *A) Uniform Manifold Approximation and Projection (UMAP) projections of tIMEC, XD and MD accessibility data along with publicly available accessibility datasets representing human primary cancer samples from different locations (Corces et al 2018), human primary non-cancerous tissues from different locations (Zhang et al. 2021) and claudin-low cell line MDA-MB-231 and its metastatic subpopulations (Cai et al., 2020). B) Boxplot showing Spearman pairwise distance between tIMEC/XD/MD, and WT-IMEC/tIMEC/M2, MDA-MB-231 samples, and the human primary tumour tissues in the TCGA dataset ordered by similarity.*

We measured the similarity of chromatin accessibility between the tIMEC, XD and MD samples and each of the other samples in the study by calculating the Spearman Rank-order correlation between of all samples. Spearman pairwise distance between tIMEC, XD and MD and the publicly available datasets (**Figure 17**B) indicated that metastasis derived cells cluster closest to both primary tumour derived cells and tumorigenic IMEC, close to samples from various carcinoma, IMEC WT and basal breast cancer, followed by the other subtypes of breast cancer and the other carcinomas included in the publicly available datasets. Interestingly, healthy tissue samples showed one of the lowest degrees of similarity with the "tIMEC-XD-MD" group.

Altogether these findings suggested that overall, our model's epigenetic landscape resembles chromatin accessibility features of human primary cancer tissues and that its accessibility profile is relatively close to that of basal breast cancer samples.

4.1.2.3. *De-novo* enhancer interactions are established in metastasis

Our specific aim was to identify regulatory elements that are hyperactivated in metastasis: these might be relevant in activating oncogenic pathways that enhance metastatic cells plasticity and thus favour the seeding process. To reach this goal, we investigated the connectome of our model and took advantage of the novel HiChIP technique.

Chromosome conformation capture (3C) combined with sequencing (Hi-C) has been reframing in our capability to appreciate the 3D organization of the genome at high resolution (Rao et al., 2014). However, because Hi-C assesses all possible proximity ligations in the genome, deep sequencing is required to identify chromatin architectural features and to quantify the frequency of chromatin looping. To achieve increased specificity, enrichment strategies have been developed to target factor-directed interactions via chromatin interaction analysis by paired-end tag sequencing (ChIA-PET) and locus-specific interactions via Capture-C and related methods (Dekker and Misteli, 2015). However, despite recent advances, ChIA-PET still requires hundreds of millions of cells per experiment and results in a modest fraction of informative reads for a given sequencing depth (Tang et al., 2015). To address these problems, Mumbach et al. developed HiChIP, a method that uses in situ Hi-C and transposase mediated on-bead library construction principles. In HiChIP, long-range DNA contacts are fixed before lysis, reducing possible false-positive connections and improving efficiency. ChIP is then performed on the samples, capturing long-range interactions associated with a protein of interest. Paired-end sequencing is then needed to detect the two distantly located segments of the genome (Mumbach et al., 2016).

To investigate enhancer connectome in tIMEC, XD and MD, I prepared, in collaboration with the Rada-Iglesias group, the H3K27ac HiChIP samples; libraries were then prepared by the sequencing facility at the CMMC (Cologne) and sequenced by the Next Generation Sequencing Facility at CIBIO. The Bicciato group analysed the sequencing results. Although multiple replicates would surely strengthen the conclusions that can be driven from this experiment, assessing the variation of chromatin interaction in one sample might be sufficient to reveal the different chromatin landscapes in our model, we thus proceeded with the analysis of a single biological replicate.

H3K27ac HiChIP raw interaction maps indicated that tIMEC, XD and MD cell populations showed a similar global chromatin interaction pattern, as depicted using a low resolution (250kb). These observations indicated that genome folding is comparable in the three cell types and that major genome rearrangements did not occur during tumor progression. By analysing the retrieved contact matrixes at higher resolutions (5kb), we could depict the presence of TADs, as highlighted by the occurrence of multiple interactions in the SOX9 locus. Of note, comparative analyses among the different cell types showed distinct enhancer interaction patterns, including long-range interactions (**Figure 18**). To better define whether the performed HiChIP assays could retrieve relevant changes in chromatin looping between cis-regulatory elements, we focused on specific loci encompassing relevant genes involved in breast cancer and metastasis formation. CXCR4 locus was recently demonstrated to be subjected to enhancer

rewiring during tumour progression (Cai et al., 2020). We thus analysed CXCR4 locus enhancer interactions as a potential metastatic-specific reference locus (**Figure 19**).

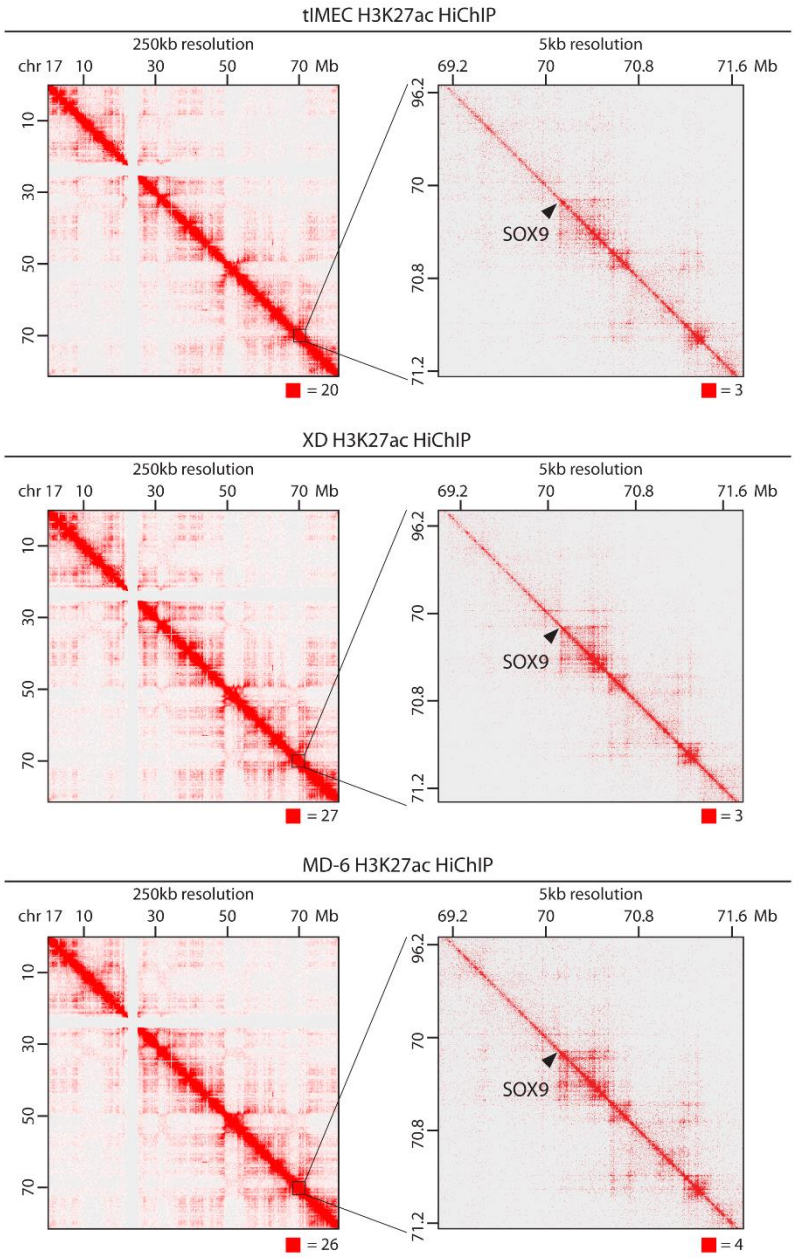


Figure 18 H3K27ac HiChIP raw interaction maps reveal different enhancer connectome in tIMEC, XD and MD cells: *H3K27ac HiChIP raw interaction maps of SOX9 locus in tIMEC, XD and MD cells (from top to bottom). 250kb to 5kb resolution (left to right). SOX9 locus (window = 80Mb left, 3 Mb right). Numbers below the interaction maps correspond to maximum signal in the matrix.*

C-Myc is also known to be a reference locus for enhancer connectome alteration during tumorigenesis in breast cancer (Ahmadiyeh et al., 2010; Zhang et al., 2016), we thus analysed the enhancer connectome landscape in its locus (**Figure 20**).

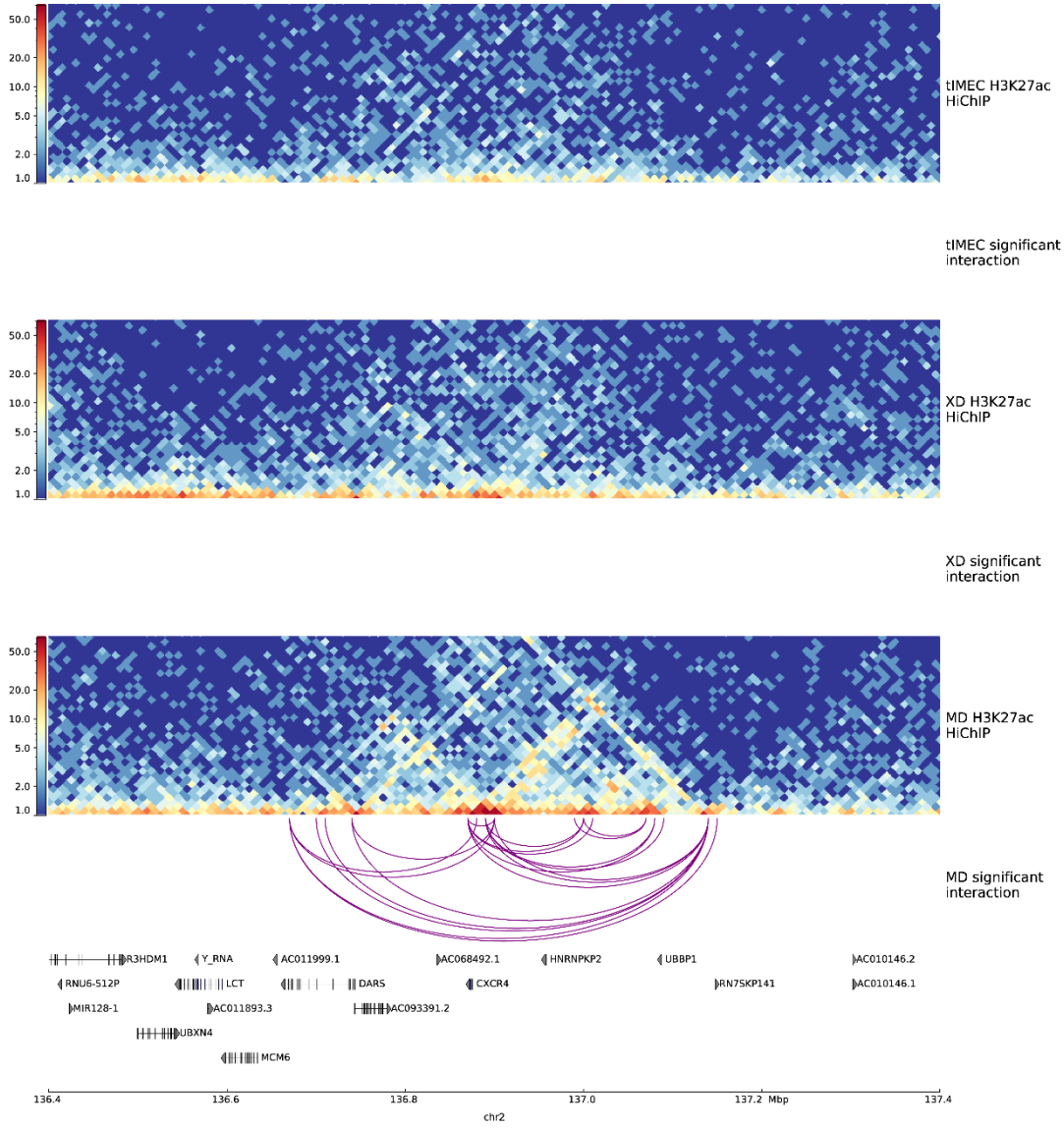


Figure 19 HiChIP contact matrices reveal different enhancer connectome in tIMEC, XD and MD cells: CXCR4. *HiChIP contact matrices generated with pyGenomeTracks of tIMEC, XD and MD cells (from top to bottom) with gene annotation-oriented significant (FDR ≤ 0.01) loops (tIMEC = green, XD = blue, MD = purple). Ensembl gene annotation (GRCh37 v. 103). CXCR4 locus (window = 1Mb). 10 kb resolution. Color range 1-50 (blue-red).*

In **Figure 19** and **Figure 20** contact matrices showing the enhancer connectome of reference loci in tIMEC, XD and MD cells are reported: HiChIP experiment was successful in identifying enhancer interactions in the three cell types. In the reported example contact matrices, we were able to identify loci that showed increased interaction in MD cells with respect to tIMEC and XD (**Figure 19**), or that showed steady interactions number between enhancers and promoters, but different interaction anchoring in MD, XD and tIMEC cells (**Figure 20**).

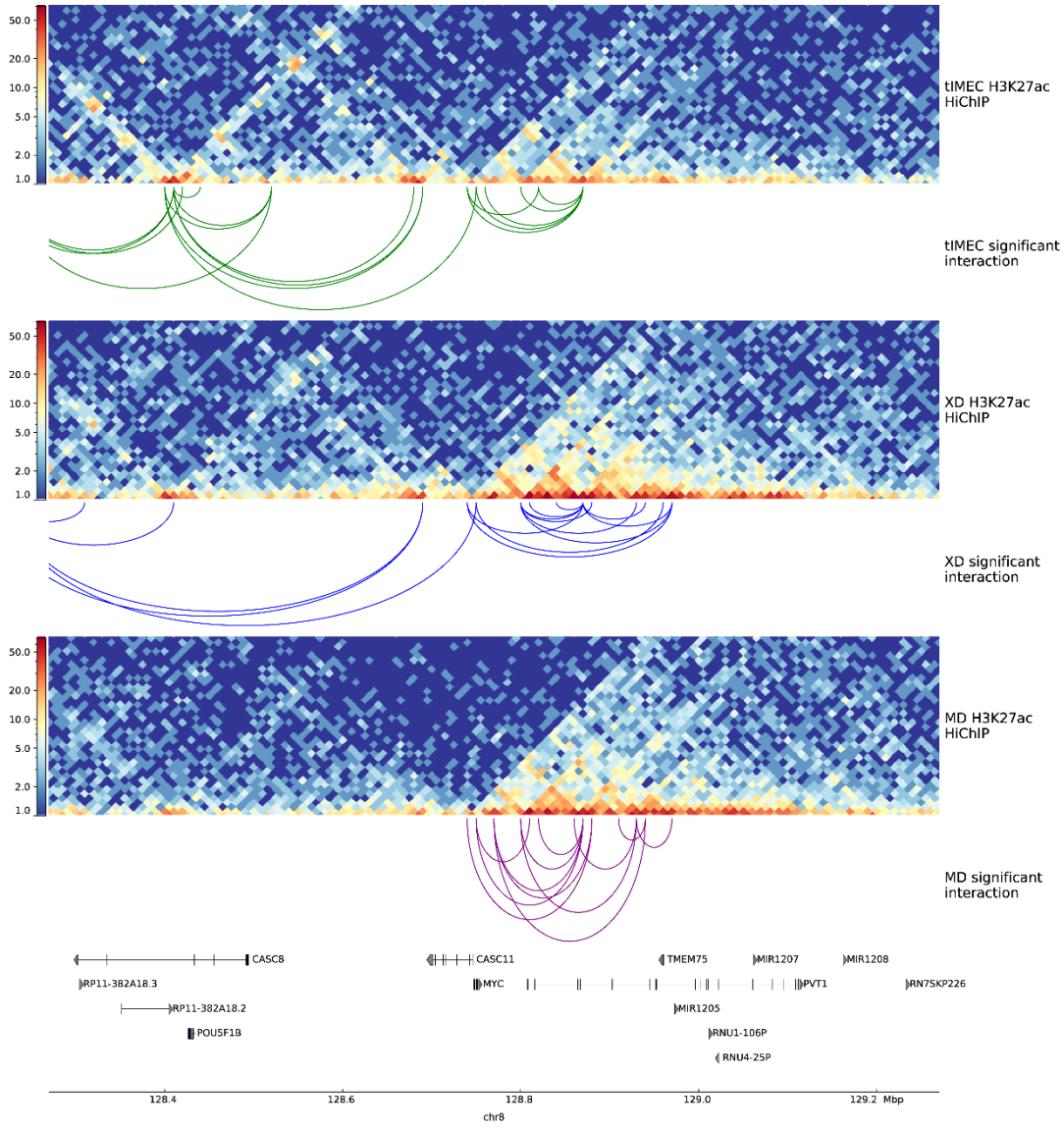


Figure 20 HiChIP contact matrices reveal different enhancer connectome in tIMEC, XD and MD cells: c-Myc. *HiChIP contact matrices generated with pyGenomeTracks of tIMEC, XD and MD cells (from top to bottom) with gene annotation-oriented significant (FDR ≤ 0.01) loops (tIMEC = green, XD = blue, MD = purple). Ensembl gene annotation (GRCh37 v. 103). c-Myc locus (window = 1Mb) 10 kb resolution. Color range 1-50 (blue-red).*

After verifying reference loci, we were confident that HiChIP data were reliable; we thus analysed the common and unique enhancer interactions in the three cell types. We identified a total of 39351 significant interactions in MD cells (**Figure 21**A), compared to the 32360 and 11977 in XD and tIMEC, respectively. The distribution of interaction annotation resulted similar among the three samples. Furthermore, we identified a total of 57034 unique interactions versus a total of 6466 common interactions across all samples (**Figure 21**B), indicating that the three cell types show a very different enhancer connectome.

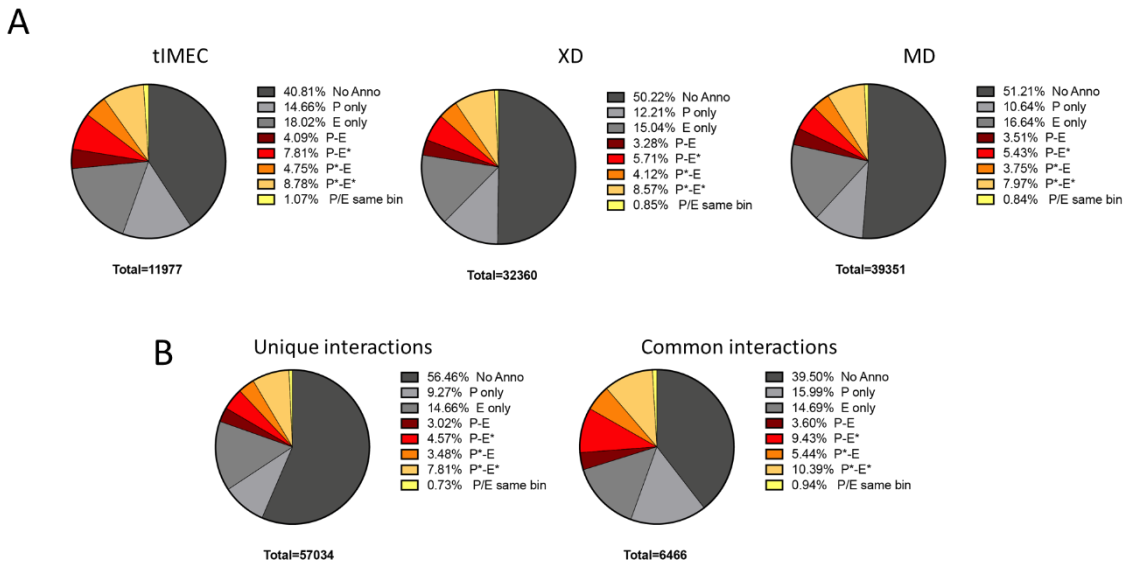


Figure 21 H3k27ac HiChIP reveals different enhancer connectome in tIMEC, XD and MD cells*. A) total interactions in tIMEC, XD and MD cells. B) Unique and common interaction across all cells. Promoter region were defined as TSS -2500/+150 bp, Enhancer regions as ATAC-seq peaks not overlapping promoter regions within a 1.5kb window. No Anno: both interacting bins do not overlap a promoter (P) or a regulatory element (E); P only: bins (either one or both) are overlapping a P (one or many); E only: bins (either one or both) are overlapping an E (one or many); P-E: one-to-one P-E interaction (one bin overlaps P, the other overlaps E); P-E*: one-to-many P-E interaction (one P and many E); P*-E: many-to-one P-E interaction (many P and one E); P*-E*: many-to-many P-E interaction; P/E same bin: single P and single E overlapped by the same bin.*

Unique interactions refer to interactions that have only been detected in one cell type, whereas common interactions refer to interactions that have been detected in two or more cell lines. We identified almost ten times more unique interactions with respect to common interactions; unfortunately, we cannot presume that this is a biologically relevant result or whether it is a technical bias; globally, unique interactions suggested that MD cells are associated with a rewiring of the 3D genome, although we cannot strictly conclude that this is MD specific. Nevertheless, it was clear that MD are associated with the establishment of a higher number of enhancer-enhancer (E-E) and enhancer promoter (E-P) interactions in comparison to its XD and tIMEC counterparts. Altogether, these results indicated that MD cells may be associated with rewiring the 3D genome at regulatory elements, although an orthologous experiment is needed to confirm these results.

4.1.2.4. Metastatic enhancers and TFs binding

To investigate the link between transcription factors binding and enhancer hyperactivation, we proceeded with the Integrated analysis of motif activity and gene expression (IMAGE). IMAGE allows precise prediction of causal transcription factors based on transcriptome profiling (RNA-seq data) and genome-wide maps of enhancer activity (enhancer marks ChIP-seq and DNAse-seq). High accuracy is obtained by the combination of a near-complete database of 896 position weight matrices (PWMs), produced by compiling public databases and systematic calculation of PWMs for unknown transcription factors, with a state-of-the-art approach for PWM scoring and a novel machine learning strategy to predict the contribution of motifs to transcriptional activity (Madsen et al., 2018).

Dr Beyes, part of Dr Zippo group, performed IMAGE analysis combining previously obtained ATAC-seq and RNA-seq data. IMAGE searches for motifs using an extended PWM database after removing motifs without any hits in the supplied sequences and motifs mapping to transcription factors with low expression.

IMAGE analysis allowed for the identification of transcription factors binding motifs in tIMEC, XD and MD cell types, excluding those motifs that are bound by low-expressed TFs. Target enhancers of the TFs for which motifs were present were identified, and target genes of the identified enhancers were predicted. IMAGE uses a machine learning model for transcriptional regulation that is based on the activity of both enhancers and promoters, and subsequently, integrates gene expression with motif activity analysis, thereby allowing for a prediction of causal transcription factors, which identify TFs important (causal) for the transcriptional network in the analysed cellular context as described in Rauch et al., 2019.

In **Figure 22**, the TF binding motifs enriched (FDR>0.01) in enhancers differentially accessible in MD cells with respect to XD cells are shown. Reported transcription factors are expressed in MD cells and are causal for enhancer activity. These filtering steps

allowed to define the 89 most significant binding motifs out of the 992 initial hits. Clustering analyses grouped a series of TFs which showed a relative low activity level in the tIMEC that increased in MD cells. Among these, IMAGE analyses retrieved TFs that have been previously shown to play a major role in epigenome reprogramming during metastatic progresssion such as NFIB, BACH1, TWIST2, ZEB1 and the homeobox factors SOX2 and SOX9 (Fang et al., 2011; Liang et al., 2012; Ma et al., 2020; Piva et al., 2014; Zhang et al., 2018; Zilli et al., 2021). These results underlined the robusteness and the significance of the obtained results. Of note, IMAGE analysis shows enrichment for EMT-involved transcription factors, such as TWIST and MECOM, and lineage-specifying transcription factors such as FOX-, HOX- and SOX- families, and PRDM family (Bleu et al., 2021; Wang et al., 2016). We also noted enrichment for signalling-responsive TFs such as BACH1 and NFE2L2, involved in the oxidative stress response and relevant in breast cancer progression (Lee et al., 2019).

After analysis of TF families for which binding motifs are represented, we could assess that MD enhancers are enriched for binding sites of TFs belonging to different families. Specifically, we assessed that the most significantly represented TF family is represented by the nuclear receptors (p value= 0.0005), primarily enriched for retinoic acid receptor group. This finding fostered our interest, as it is known that the retinoic acid pathway is involved in the metastatic process (Sosa et al., 2015). We thus hypothesised that RARs might play a role in activating a signalling response in the metastatic cells.

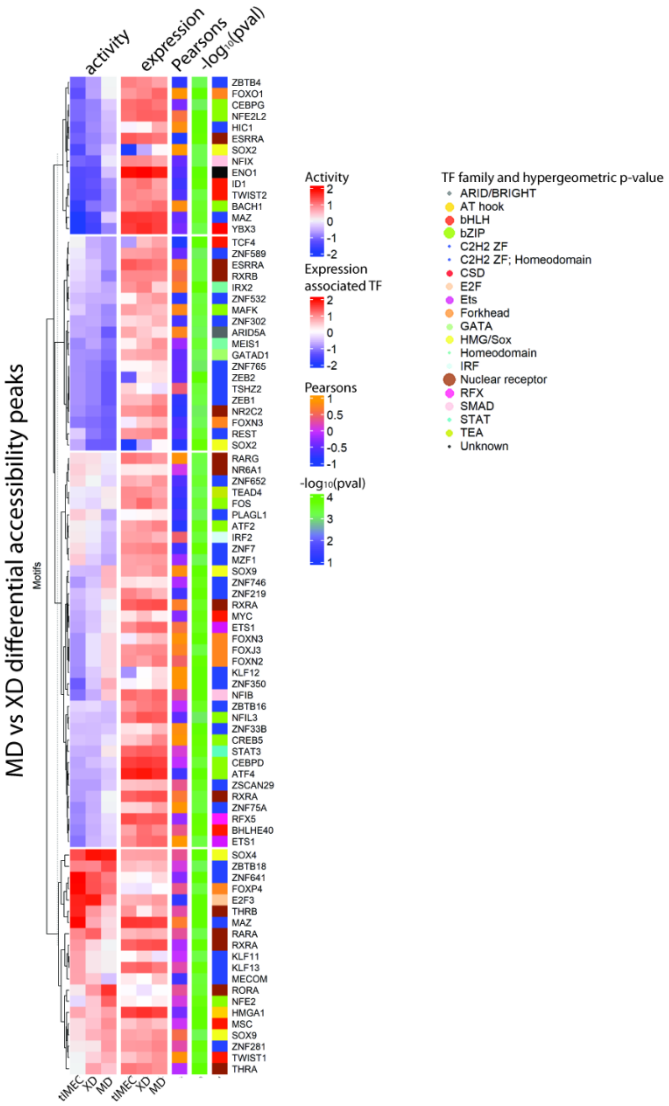


Figure 22 TF motifs enrichment analysis in metastatic derived cells differential enhancers highlights nuclear receptors and SOX9*. Heatmap showing the scaled activity of motifs detected in differential (MD vs XD) ATAC-seq peaks. Only TFs that are expressed in MD cells and that considered causal for target gene expression modulation are shown. Additional columns show Pearson coefficient (range -1 to 1), -log10(p-val) (range 1-4), TF family (based on Lambert et al., 2018) with relative hypergeometric p-value (higher the radius of the dot, higher the significance) range 0.0005-0.99; "Unknown" family p-value = N/A.*

Interestingly, one of the most represented TFs that was highlighted by this analysis is *SOX9*, which was in line with our hypothesis of this factor being important in terms of regulation of enhancer activity and gene expression in our model (**Figure 11**)

In order to assess whether the most significantly enriched TFs that emerged from the IMAGE analysis were also part of significant networks, we proceeded with network analysis on the 89 transcription factors shown in **Figure 22**; we chose to analyse this list of TFs which is the result of very stringent filtering steps, and it thus represents a small subset of the actual IMAGE output, because we aimed at studying the most significant and thus presumably relevant TFs, at the cost of losing part of the information that can be retrieved by the IMAGE analysis.

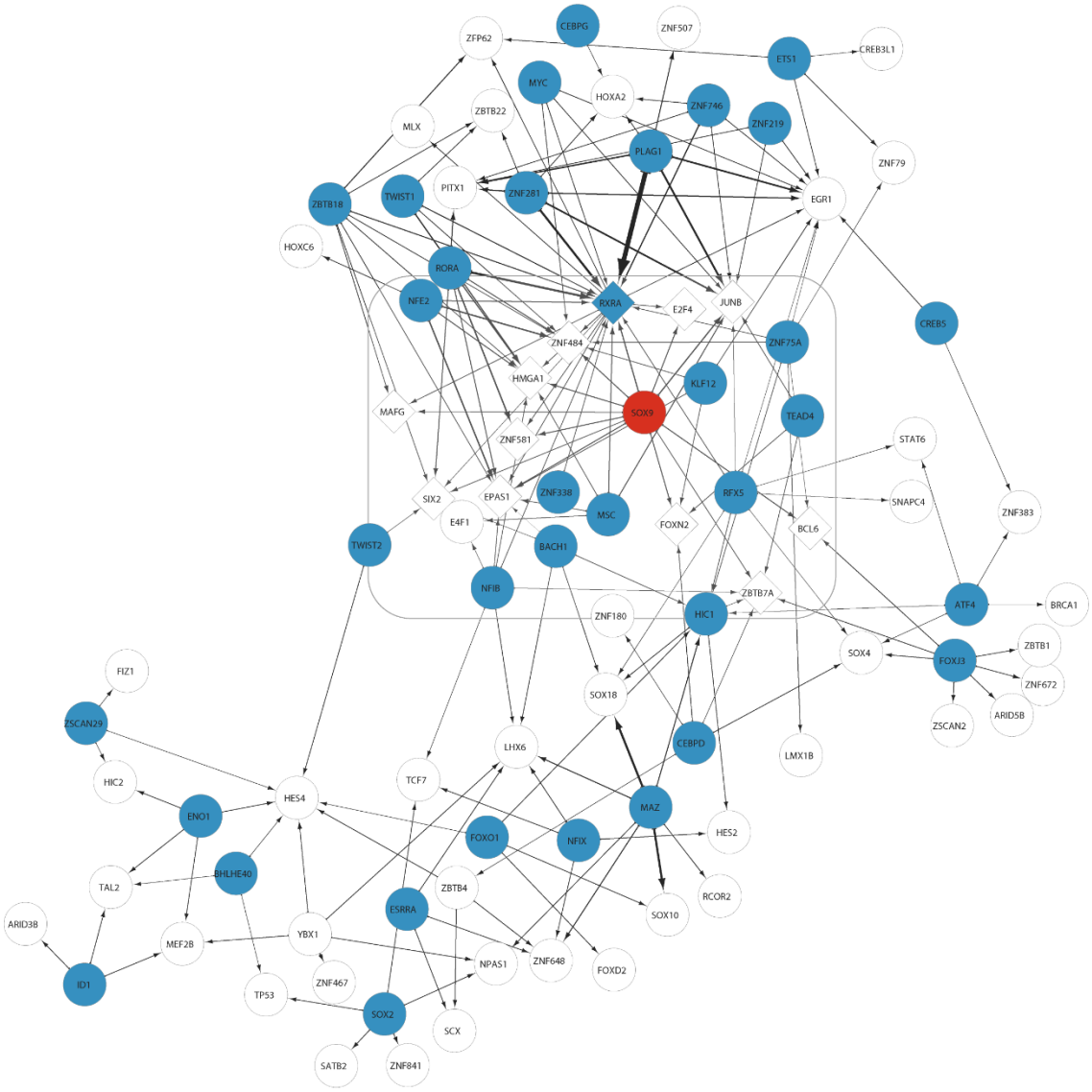


Figure 23 TF network analysis *TF network analysis overview and focus on central neighbourhood, containing SOX9 (red). TF network analysis shows target genes of causal, MD-expressed TFs with motif enrichment in MD>XD ATAC-seq peaks that were used as input for the analysis. Circle shape= TFs; Blue = causal TFs; Diamond shape = SOX9 target TFs; Edge sizes = target gene score (range 0-3).*

Transcription factor network analysis revealed the presence of multiple neighbourhoods, where central nodes were revealed as PLAG1, ZBTB18, RORA, and, surprisingly, SOX9, whose neighbourhood is central to the network (**Figure 23**). Among SOX9 target genes, RXRα represents the only causal transcription factor: we thus hypothesised that SOX9 activity may be linked to the modulation of the retinoic acid pathway. Furthermore, we confirmed that SOX9 may be worth to be investigated as part of the group of transcription factors that are indicated to play a role in distinguishing the metastatic cell population from tumorigenic cells and primary tumour derived cells.

4.1.2.5. The metastasis enhancer hyperactivation hypothesis

We hypothesised that dysregulation of regulatory elements is a major driver for establishing an alternative epigenetic program, due to rearrangements at the chromatin level and alterations of chromatin loops, which in turn modify promoter-enhancer interactions. This may lead to the activation of oncogenic pathways and an increase of cellular fitness by facilitating adaption to environmental cues during cell dissemination and metastatic colonization.

We focused on the *SOX9* locus to provide a proof of concept of enhancer dysregulation driving metastatic capacity, and we were expecting to assess the establishment of different scenarios: 1) the activation, in a specific step of tumorigenesis, of regulatory elements that act as additive to enhancers already active in parental cells or primary tumour 2) the repression of regulatory elements which show activity in one cell state, but are then silenced during tumour progression 3) the activation of *de novo* enhancers in a specific step of tumorigenesis (**Figure 24**).

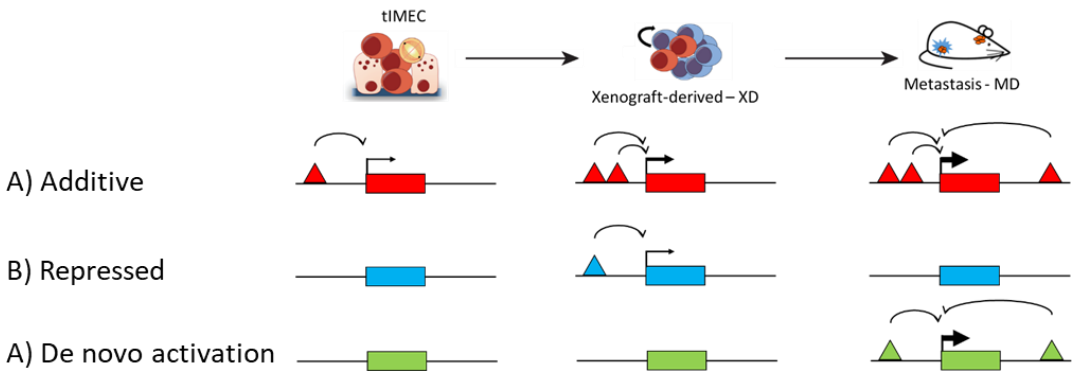


Figure 24 The enhancer dysregulation hypothesis. *Schematic representation of enhancer dysregulation in the tumorigenesis and metastasis steps. Red: the promoter (square) activity is increased from parental cells to metastatic cells through the activation of additional enhancer regions (triangle); Blue: the promoter activity is enhanced specifically at one step of the process, through the activation of regulatory element, which is then 'turned off' in the switch to another cells state; Green: the promoter is enhanced specifically at one step of tumorigenesis through the activation of a de-novo regulatory element.*

The first step to provide this proof of concept was to validate SOX9 as a good candidate for enhancer hyper/hypo-activation analysis, as discussed below.

4.1.3. *SOX9* is overexpressed in primary tumour and metastasis

In order to establish whether the increased accessibility at enhancer around the *SOX9* locus would effectively reflect on *SOX9* expression itself, and thus whether *SOX9* would be an efficient readout for enhancer activity, we analysed the expression of *SOX9* in the different cell types. We compared IMEC WT with transformed IMEC, three independent primary tumour xenografts derived cell populations and one population of metastatic derived cells.

After evaluating *SOX9* mRNA levels in the different cell populations through RT-qPCR (**Figure 25)**, we assessed that the *SOX9* transcript increased from wild-type cells to metastatic derived cells. We then tested whether, as a consequence of augmented *SOX9* expression, the protein levels of this transcription factor were also augmented in MD cells.

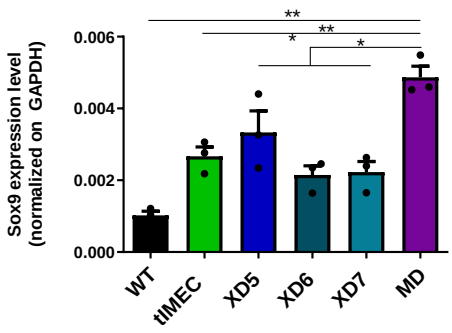


Figure 25 SOX9 expression level in WT, tIMEC, XD and MD. *SOX9 transcript levels in the indicated cell types, normalized on GAPDH transcript levels. 3 biological replicates, AVG +/- S.E.M. Unpaired one-tailed student's t-test. Ns:P > 0.05; *:P ≤ 0.05; **:P ≤ 0.01;***:P ≤ 0.001.*

Through Western Blot (WB) analysis (**Figure 26**A), we assessed that SOX9 total protein level was increased in MD cells with respect to XD cells, tIMEC and WT cells. Performing multiple replicates of WB experiments, we quantified the optical density of the SOX9 protein level relative to the housekeeping GAPDH (Figure 24B); this allowed us to assess that the increase in SOX9 total protein levels from wild-type cells to metastatic cells was statistically significant.

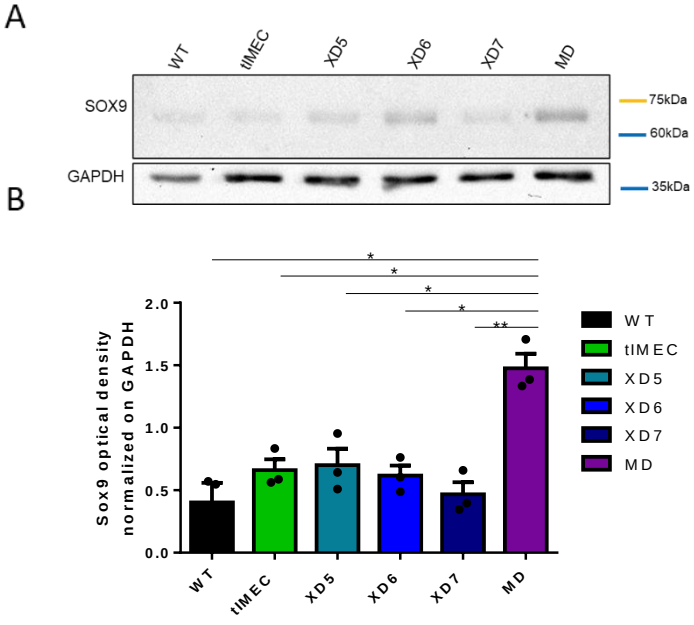


Figure 26 SOX9 protein level quantification in WT, tIMEC, XD and MD cells. *A) SOX9 total protein levels in the indicated cell types, representative image of 3 experiments. B) SOX9 relative total protein levels quantification in the indicated cell types. 3 biological replicates, AVG +/- S.E.M. One-tailed t-test. Ns:P > 0.05; *:P ≤ 0.05; **:P ≤ 0.01.*

Since SOX9 is a transcription factor that exerts its role mainly in the nucleus, we assessed nuclear protein levels, and in general protein localization, in WT, tIMEC, XD and

MD cell types through immunostaining (**Figure 27**A). Quantification of SOX9 nuclear mean intensity at the single-cell level (**Figure 27**B) showed that SOX9 nuclear protein levels were increasing from wild-type cells to metastatic cells, indicating that a nuclear accumulation of the protein also reflected the total protein levels augmentation as seen in WB analysis.

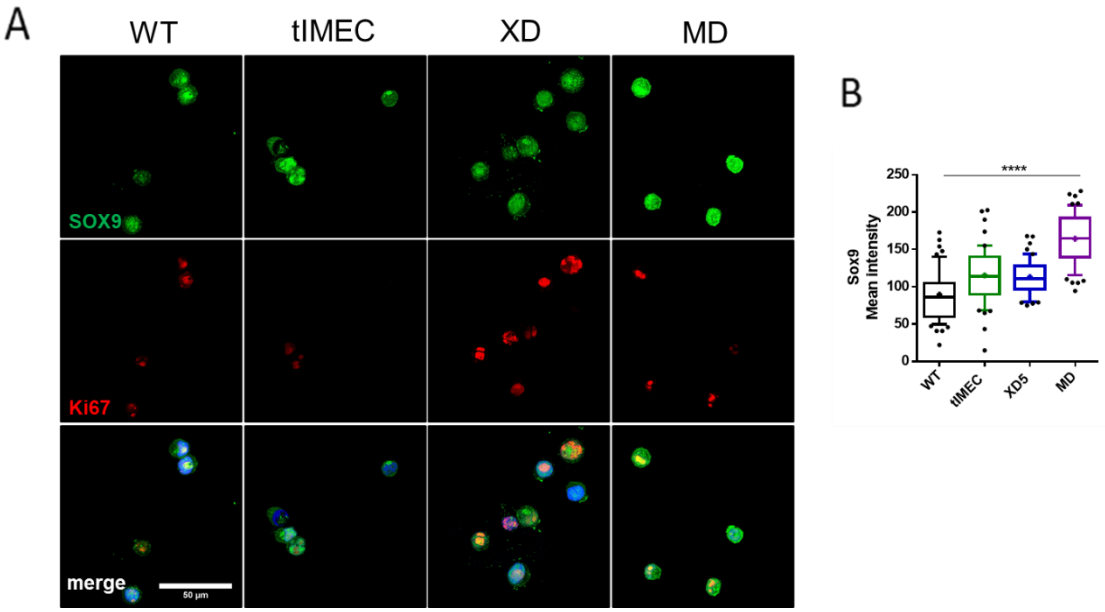


Figure 27 SOX9 immunofluorescence staining in IMEC WT, tIMEC, XD and MD cells. *A) Representative images of SOX9 and Ki67 immunostaining in the indicated cell types. Images were acquired with an SP8 confocal microscope, 63X objective. SOX9, green; Ki67, red; scale bar = 50 µm. B) Quantification of SOX9 nuclear mean intensity in the different cell types; 3 biological replicates; The box plots indicate median values (middle lines), first and third quartiles (box edges) and 10th and 90th percentiles (error bars) of the fluorescence intensity of analysed nuclei (n≥300). Unpaired one-tailed student's t-test. Ns:P > 0.05; *:P ≤ 0.05; **:P ≤ 0.01;***:P ≤ 0.001;****:P ≤ 0.0001.*

With these results, we confirmed that SOX9 is expressed in our TNBC model and that its expression resulted to be increased from WT to MD cells, reflecting the behaviour of this SOX family member in basal-like patients (**Figure 9**) and triple-negative breast cancer models (Ma et al., 2020; Mehta et al., 2019; Tang et al., 2020; Wang et al., 2021a). Furthermore, the expression level was sufficient to study SOX9 enhancer hypo-/hyper-activation in the model, since even slight changes in its expression, such as those that may be due to epigenetic alterations, could be detected.

4.1.3.1. *SOX9* enhancer identification and validation

As *SOX9* was expressed in MD cells at higher levels if compared to XD and tIMEC, and we also observed more interactions at the *SOX9* locus in MD compared to XD and tIMEC, we rationalized that *SOX9* can indeed be used as a readout of enhancer activity in our model. Thus, we first sought to identify those putative regulatory elements proximal to the SOX9 promoter and demonstrate their potential activity in metastasis.

4.1.3.1.1. *SOX9* putative enhancers chromatin accessibility and interaction

To investigate the chromatin accessibility around the *SOX9* locus, we analysed ATAC-seq data to identify chromatin hyper-accessible regions in MD cells. We were able to identify 20 peaks (**Figure 28** and **Figure 29**) defining chromatin accessible regions in MD cells. Differential analysis of the peaks identified in the three cell lines showed that among those, most resulted in being differentially accessible in the three cell populations (**Figure 29**), and regions 13-15, 20 show significantly higher chromatin accessibility in the metastatic derived cell line with respect to tIMEC and XD cells. We defined these peaks of increased chromatin accessibility as putative regulatory elements.

We then analysed the enhancer connectome in the *SOX9* locus by H3K27ac HiChIP; *SOX9* locus resulted to be highly plastic in terms of enhancer-promoter and enhancer-enhancer chromatin loops when comparing the three cell types (**Figure 28**). Interaction frequency between highly accessible chromatin regions surrounding the SOX9 locus indicated that MD cells are characterized by a higher number of chromatin regions interacting between each other and with the SOX9 promoter, when compared to both

tIMEC and XD cells. Nevertheless, H3K27ac increased anchoring may derive from an increment in acetylation levels; thus, further investigation will be needed to confirm the augmentation of enhancer-promoter interactions.

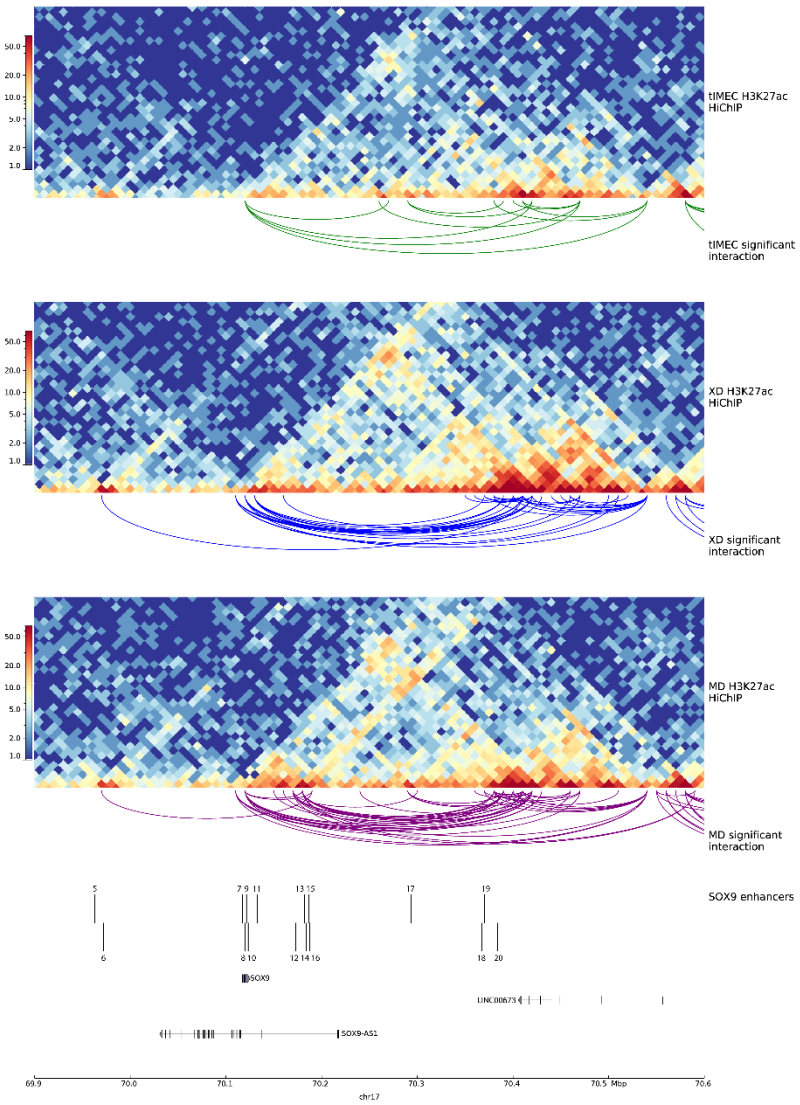


Figure 28 HiChIP contact matrices reveal high plasticity of SOX9 locus connectome in tIMEC, XD and MD cells. *SOX9 locus HiChIP contact matrices generated with pyGenomeTracks of tIMEC, XD and MD cells (from top to bottom) with gene annotation-oriented significant (FDR ≤ 0.01) loops (tIMEC = green, XD = blue, MD = purple). ATACseq peaks are reported as 'SOX9 enhancers'. Ensembl gene annotation (GRCh37 v. 103). Window = 700kb. 10 kb resolution. Color range 1-50 (blue-red)*

Although enhancer-enhancer interaction would be interesting to be further investigated, we decided to focus, in the first place, on specific enhancer-promoter interaction, in order to filter for those putative enhancer regions that would more likely act on SOX9 expression. Through the intersection of ATAC-seq peak calling with HiChIP 25kb bins relative to looping putative regulatory elements (**Figure 29**), we assessed that regions 13, 14, 15 and 20 show significantly increased accessibility in MD cells (**Figure 29** B), and that most MD chromatin accessible regions (3-6, 12-20) interact with the *SOX9* promoter in MD cells (**Figure 29** A).

Having assessed that the identified MD accessible regions are also physically interacting with *SOX9* promoter, independently of differential chromatin accessibility, we hypothesise that the regulatory elements 3 - 6 and 12 - 20 can be considered as potential *SOX9* enhancers, although regions 13, 14, 15 and 20 result to be the most promising ones in terms of SOX9 expression modulation. To further validate the potential enhancer activity of identified peaks, we tested the regulatory elements for acting as an enhancer of gene expression.

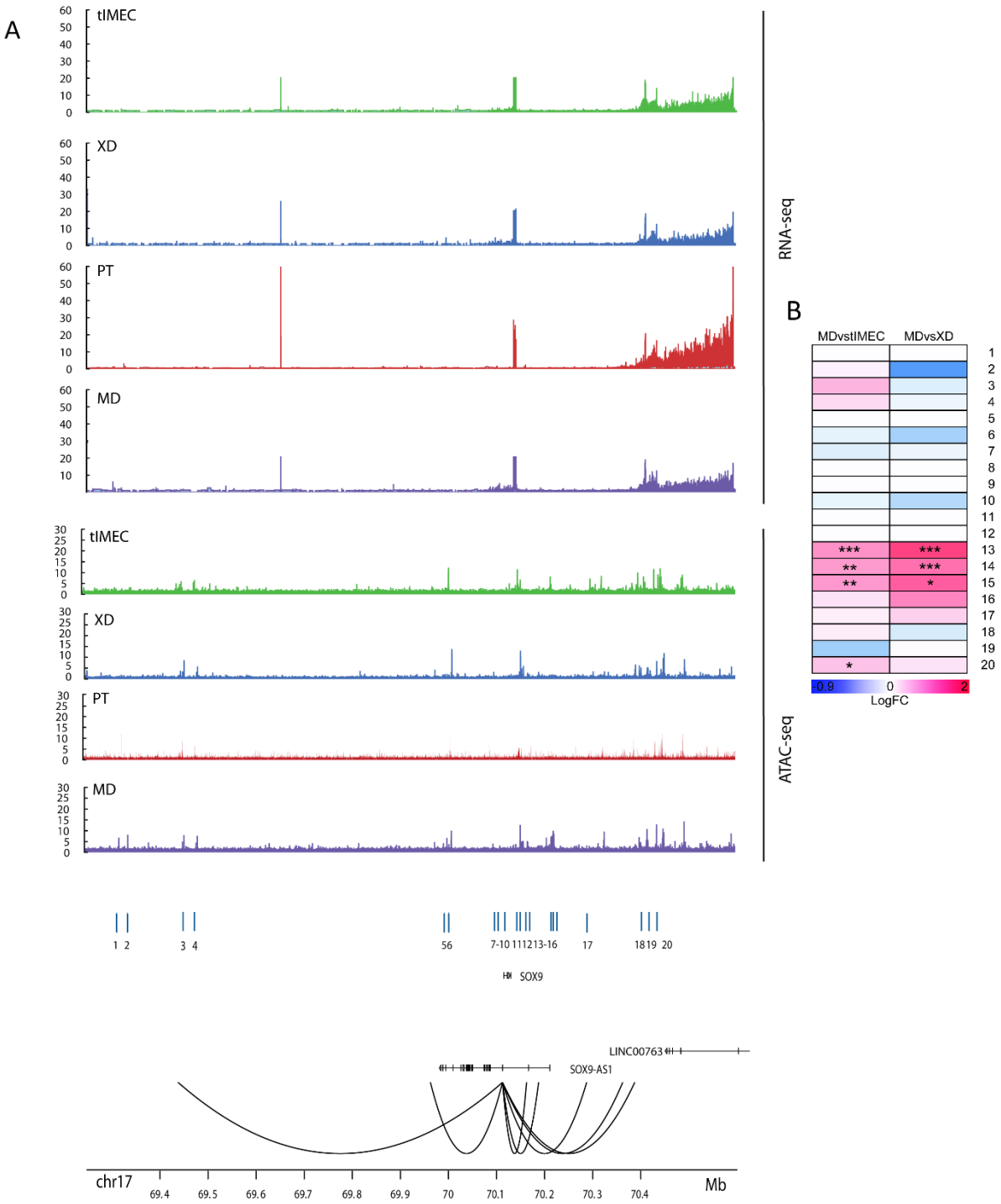


Figure 29 Intersection of ATACseq and HiChIP loops identify SOX9 looping enhancers. *A) ATACseq (upper part) and RNAseq (lower part) coverage plots of the SOX9 locus in tIMEC (green), XD cells (blue),MD cells (purple) and Primary Tumor tissue (PT, red). SOX9 enhancers 1-20 are indicated as blue lines over the gene annotation. MD regulatory elements interactions with SOX9 promoter at 25kb resolution are shown below the gene annotation. RefSeq gene annotation. Window = 1 Mb. B) Heatmap of Log_2(fold change in chromatin accessibility) of indicated ATAC-seq peaks at the SOX9 locus for MDs vs tIMEC and MDs vs XDs respectively. Colour coding of the log_2(FoldChange) ranges from -0.9 to 2. ***: FDR<0.005; **: FDR< 0.05; *: FDR<0.1.*

4.1.3.1.2. *SOX9* putative enhancers activity

In order to verify whether the identified putative enhancers at the *SOX9* locus could effectively act as regulatory elements by inducing the activity of a promoter, luciferase assays were employed. Those were performed by using a pGL4 backbone with each enhancer fragment cloned upstream of a minimal promoter, which guides the expression of a Firefly luciferase (**Figure 30**). As a positive control, the TCF repeated motif was cloned in place of the enhancer: if the Wnt pathway is activated (following treatment with Chiron), β-catenin binds to the TCF motif and induces the activity of the downstream promoter.

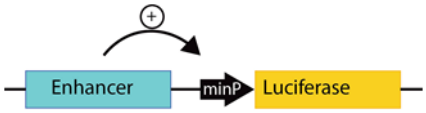


Figure 30 Schematic representation of luciferase expressing vector. *Enhancers (blue) were cloned upstream of a minimal promoter which guides the expression of a Firefly luciferase, in a pGL4 backbone. Positive control vector contains the TCF repeated motif, negative control is represented by the empty vector.*

Regulatory activity of the putative enhancers was tested in IMEC WT and tIMEC, but not in XD and MD. In fact, since xenograft-derived and metastasis-derived cells derive from *in-vivo* experiments, they were transduced with a constitutive Luciferase expressing vector for bioluminescence tracking of tumour cells *in vivo*. Thus, we could not distinguish between specific enhancer luciferase signals and constitutive signals. Nevertheless, we also tested BC cell lines representing the basal or luminal subtypes, MDA231 and MCF7, respectively.

Relative luminescence (RLU) is calculated as the ratio between the Firefly Luciferase signal, whose expression is guided by the upstream regulatory element, and the Renilla Luciferase, whose expression is regulated by a constitutive promoter and acts as a normalizer. Calculation of the RLU allowed to assess that all cells transfected with pGL4_TCF and treated with Chiron (C+) responded with at least a ten-fold increase of

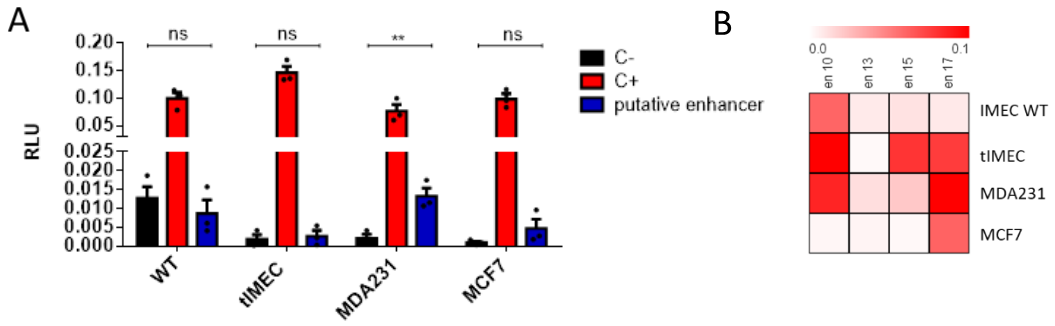


Figure 31 Relative Luminescence of an example putative enhancer and controls in BC cell lines. *A) Firefly luminescence normalized on Renilla luminescence of pGL4 constructs containing either positive control (TCF) (red), negative control (empty vector) (black) or enhancer region (blue) in IMEC WT, tIMEC, MDA231 and MCF7. Average of 3 independent biological replicates +/- S.E.M. One-tailed t-test. Ns:P > 0.05; *:P ≤ 0.05; **:P ≤ 0.01;***:P ≤ 0.001;****:P ≤ 0.0001. B) Firefly luminescence normalized on Renilla luminescence of pGL4 constructs containing indicated enhancer regions. Range 0-0.1; N=1.*

Luciferase transcription when compared to the negative control (C-) (**Figure 31**A). We tested multiple regulatory elements (**Figure 31**B) and, as for a representative enhancer region (en13) shown in (**Figure 31**A), the tested elements resulted in enhancing the minimal promoter activity of a minimum of 3-fold over the control in at least one cell line.

4.1.3.1.3. A reporter system for enhancer activity tracking

In order to test *in vitro* whether the identified enhancers are hyper/hypo activated in the different cell types of our model, we transiently transfected WT cells, tIMEC, XD and MD cells with an enhancer-reporter system (**Figure 32**), in which each enhancer (**Figure 33**C) activates a minimal promoter guiding the expression of GFP.

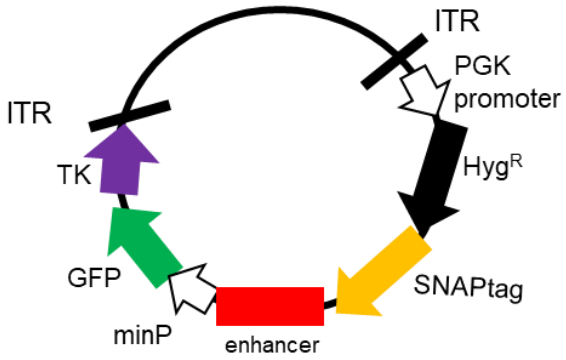


Figure 32 Enhancer reporter system schematic visualization: *In a PiggyBac transposon backbone, a PGK promoter (white) guiding the expression of Hygromycin resistance (black) and SNAPtag (yellow) coding sequences were inserted, upstream of the target enhancer (red) acting on a minimal promoter (white) guiding the expression of a GFP (green) and of the Thymidine-kinase (purple). The whole insert is flanked by inverted repeats (ITR) that allow for transposition when the vector is transfected together with a transposase coding plasmid.*

Upstream of the enhancer region, a SNAPtag coding sequence was cloned to visualise efficiently transfected cells. Furthermore, the enhancer reporter system was cloned into a Piggybac transposon backbone, allowing for the integration of the insert in the genome. Furthermore, the vector contains a Hygromycin resistance for selection of stable clones and a thymidine-kinase coding sequence: this would allow inducing cell death in response to enhancer activation upon ganciclovir administration. This reporter system could allow to establish a causal relationship between enhancer activation and metastasis formation, *in vivo*. As for the Luciferase Assay experiments, we cloned the TCF repeat motif in place of an enhancer as a positive control, and the empty vector was used as a negative control.

The PiggyBac transposon system was chosen because of the relatively low copy number of insertions that characterizes it since we want to avoid excessive disruption of higher-order chromatin structures. Nevertheless, this approach presents multiple weaknesses, such as the non-targeted insertion position and the possibility of impeding efficient visualization of the GFP expression increase following enhancer hyperactivation due to the low copy number of transposition.

The transient transfection of the enhancer reporter into WT, tIMEC, XD and MD cells allowed for visualization of GFP expression for the positive control for all the cell types analysed, although with different efficiencies, indicating that the reporter system was working correctly (**Figure 33**A). SNAPtag signal allowed for the identification of efficiently transfected cells. GFP signal deriving from enhancer activity was normalized on the intensity of SNAPtag signal to take into consideration augmentation of GFP expression due to the variability in the transfected DNA quantity. Enhancer activation–driven increase in GFP expression was detectable in MD cells in steady-state for just one out of the four regions tested (**Figure 33**A), and the fold change in respect to the

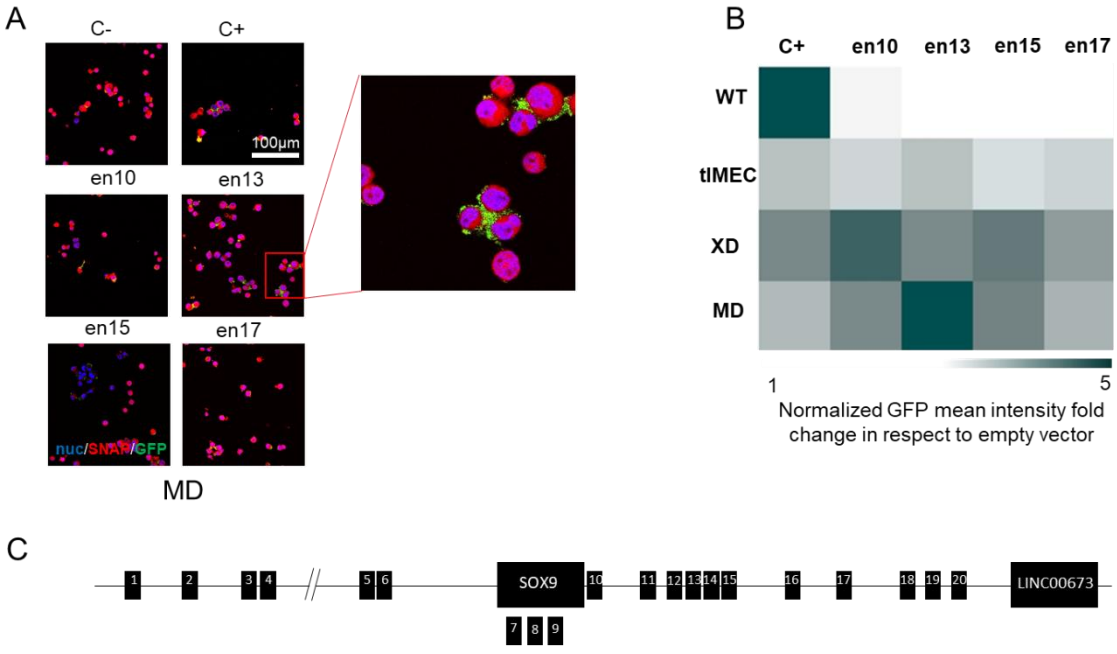


Figure 33 Enhancer reporter indicated enhancer activity in MD cells. *Transient transfection of enhancer reporter vector containing positive control, negative control and indicated enhancer regions (en10/13/15/17) in WT, tIMEC, XD and MD cells. A) panel of representative images acquired with SP8 confocal microscope, 63X objective, of MD cells transiently transfected with control plasmids and enhancer reporter plasmids. Red squares indicate 4x zoom. Nuclear pseudocolor (blue), SNAPtag staining (red), GFP (green). Scale bar =100µm. B) heatmap of normalized GFP mean intensity fold change respect to negative control of indicated enhancer regions and positive control in WT, tIMEC, XD (merge of 3 cell types) and MD cells. range 1 to 5, average of 3 independent biological replicates. C) scheme of SOX9 enhancers (1-20) locations with respect to SOX9 gene.*

negative control was slight (**Figure 33**B), indicating the possibility that one regulatory element alone may not sufficient to drive the enhancement of the reporter expression.

Based on these results and considering previous analysis on all MD putative enhancers (**Figure 22**), we asked whether enhancer activation could be regulated by cellular pathways, exerting their activity through transcription factors binding to the regulatory element. We thus proceeded with analysing the enhancers' DNA sequence focusing on transcription factor binding sites enrichment in SOX9 putative enhancers.

4.1.3.2. *SOX9* enhancers are enriched for RARs binding sites

In order to elucidate whether *SOX9* enhancers are enriched explicitly for binding sites for TF families that may suggest pathways whose modulation would result in *SOX9* expression alteration, we filtered the IMAGE results for the specific regions that overlap with the *SOX9* regulatory elements that show an interaction with *SOX9* promoter by H3K27ac HiChIP. Specifically, we assessed the distribution of MD-expressed TFs binding sites detected (FDR<0.01) in MD enhancers, filtering as in **Figure 22**, for motifs linked to MD-expressed TFs, that are enriched in MD differentially accessible ATAC-seq peaks, and that result to be causal in regulating target gene (*SOX9*) expression.

This analysis detected that all enhancers contained at least one significant TF binding site (**Figure 34**), with all putative enhancers except for region 17 showing binding sites for *SOX9*. Furthermore, we assessed that, besides *SOX9*, other transcription factors motifs are common to different regions, such as *SOX2*, EMT transcription factors (*TWIST1* and 2), KLF family and FOX family of transcription factors. Nevertheless, the retinoic acid responding transcription factors emerged as the most represented family of TFs for which SOX9 putative enhancers, especially the differentially accessible ones, show binding sites. We detected RXRβ motif in regions 6, 14, 15, 17, 19 and 20, RARα motif in regions 3, 6, 12, 14, 17 and 20, RXRα motif in regions 5, 12, 18 and 19, and RARγ motif in region 16. This result suggested that RARs binding sites, already highlighted in the genome-wide analysis (**Figure 22**), were also enriched in *SOX9* enhancers.

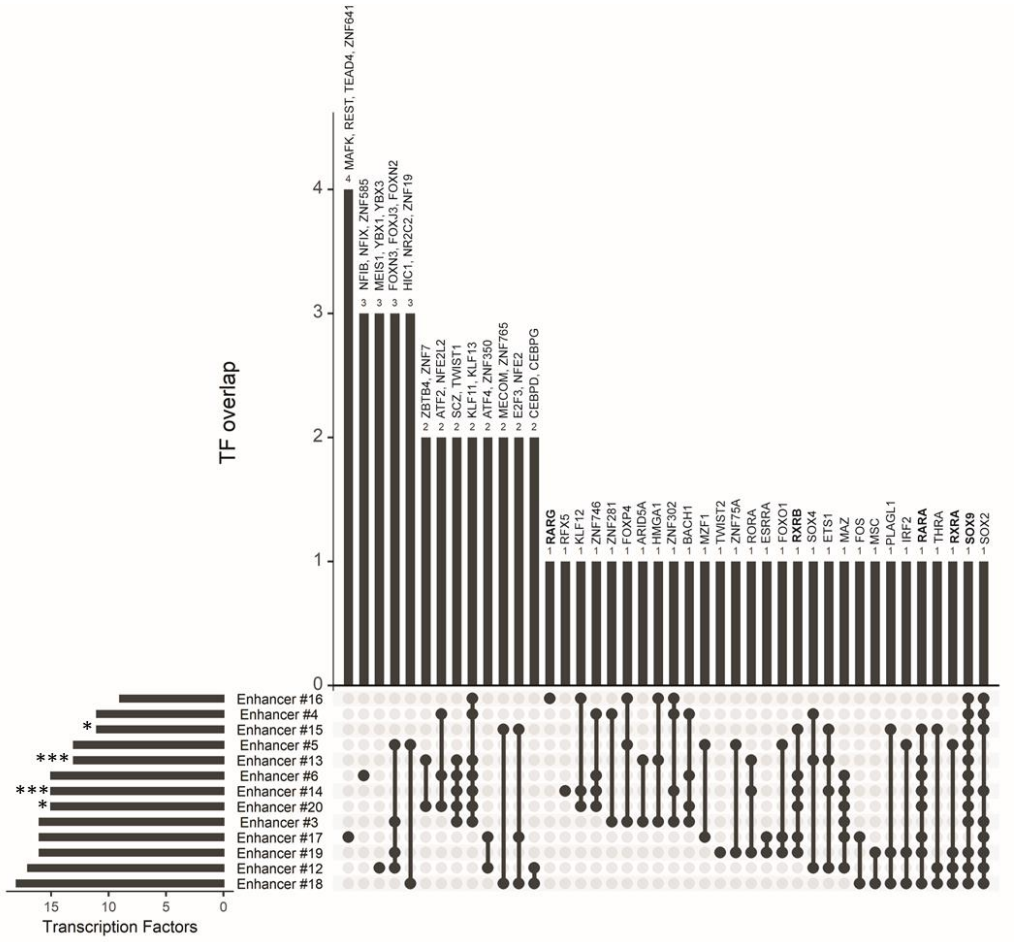


Figure 34 UpSetR plot showing common and unique TF binding sites in SOX9 promoter interacting enhancers. *UpSetR plot showing the overlap of causal transcription factors expressed in MD cells with a linked motif activity MD > XD& tIMEC, whose corresponding motifs were identified in the indicated enhancer regions with FDR<0.01. TFs of interest are highlighted. Enhancers are represented as grey dots, black dots represent enhancers in which the corresponding TFs are identified; lines connecting multiple enhancers indicate groups of enhancers that share a common list of TFs binding sites. Significantly differential enhancers (in MDvsXD or MDvstIMEC comparisons) are highlighted (***: FDR<0.005; *: FDR<0.1).*

To further confirm RARs binding sites enrichment in *SOX9* regulatory elements, we analysed the DNA sequence of enhancers 1-20 with an orthogonal approach, based on the Find Individual Motif Occurrences (FIMO) tool from the Motif-based sequence analysis tools suite (MEME). FIMO tests a set of sequences for individual matches to each user-provided motif, treating each motif independently.

We thus scanned the *SOX9* putative enhancers for retinoic acid pathway – involved TFs binding motifs, including NR2F1, RARα, RARβ, RARγ, RXRα, RXRβ and RXRγ. We also tested five regions shuffling peaks on the whole genome and used a control set of motifs to establish a threshold based on the p-value over which we may consider TF binding site enrichment statistically significant.

We found that binding sites for at least one of the mentioned transcription factors were present on *SOX9* putative enhancers (**Figure 35**), although with different confidence (p-value: black=zero, white=$1*10^{-4}$). Of interest, regions 1 and 2, which were assessed as 'non-looping' by HiChIP analysis, showed extremely low number (1 and 5, respectively) of retinoic acid receptors binding sites, which were not overcoming the significance threshold based on control regions (shuffled regions) and on the control set of motifs. On the other hand, many of the regions that showed looping to *SOX9* promoter resulted enriched for some RARs binding site with high confidence, especially regions 13, 14 (which also showed differential accessibility in MD cells) and 16.

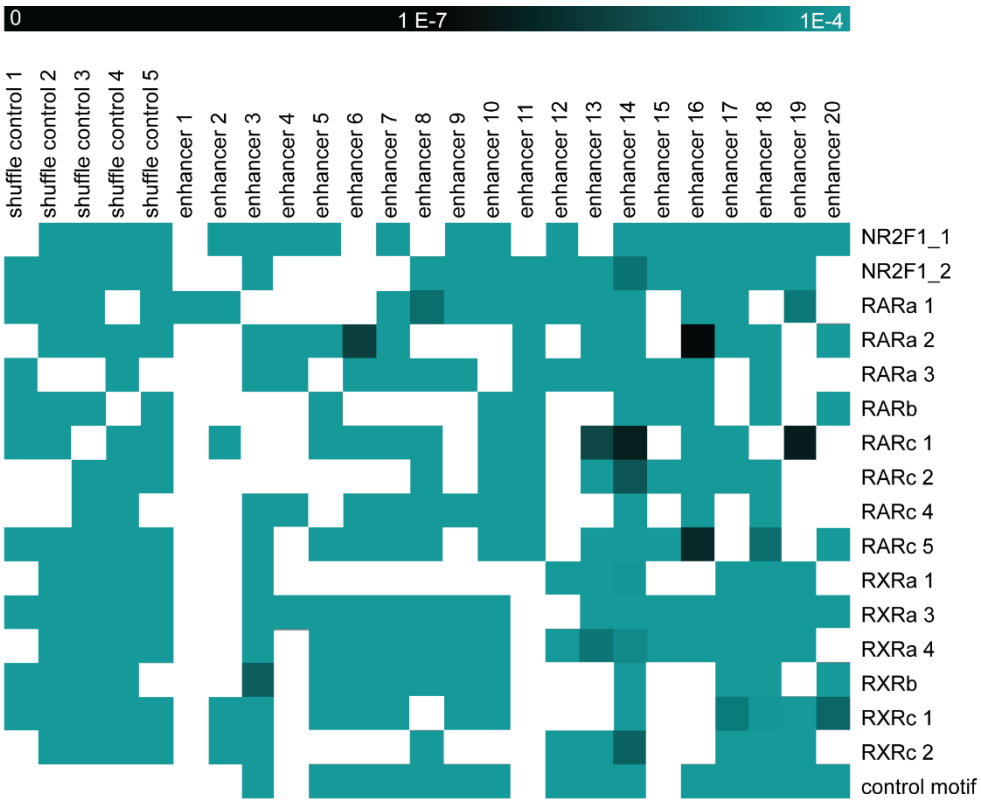


Figure 35 Confidence heatmap of RARs binding sites in SOX9 enhancers. *Heatmap representing p-value of RARs binding motifs and a control motif set in SOX9 enhancers (1-20) and shuffled control regions retrieved by FIMO analysis. Colour code = p-value 0 (black) to 1*1 $^{-4}$ (light blue) or missing motif (white).*

Many retinoic acid receptor binding sites, although consisting of complex DNA sequences, present strong similarities; results from MEME analysis may thus show a certain level of redundancy.

All in all, with these results, we could conclude that *SOX9* enhancers, especially the ones that show differential accessibility between MD cells, tIMEC and XD cells, were redundantly enriched for binding sites for TFs belonging to the retinoic acid pathway.

Thus, we asked whether retinoic acid may be involved in *SOX9* enhancer activity and *SOX9* expression and whether activation of this pathway may bring phenotypical traits to MD cells that would enhance their fitness and, therefore, their capacity to adapt to those foreign environments that they encounter during dissemination and metastasis formation.

4.1.4. *SOX9* is responsive to Retinoic acid

Since we found that the binding sites for retinoic acid receptor (RAR) family members were enriched in most of the putative enhancers *SOX9*, we hypothesised that the retinoic acid pathway might be involved in the activation or maintenance of activity of those regulatory elements. Thus, we aimed to test whether *SOX9* expression would be induced by retinoic acid treatment.

4.1.4.1. *SOX9* expression modulation by retinoic acid

Firstly, a time course treatment with 1µM all-trans retinoic acid (ATRA) was performed in tIMEC, XD and MD cells, and subsequently tested *SOX9* expression by RT-qPCR.

The obtained results showed that SOX9 expression was increased in the MD cells upon short treatment (1h) with ATRA, while tIMEC and XD seemed responsive only after prolonged stimulation (72h), as shown in **Figure 36**.

Although there was a clear trend, the *SOX9* expression increase in MD cells was slight when compared to the steady-state and not statistically significant when compared to the vehicle treatment. Since recent evidence suggests that chromatin reshaping can drive a transcriptional memory mechanism, and since retinoic acid receptors binding sites are redundant in *SOX9* enhancers, we hypothesised that combining a pre-exposure with a short treatment of ATRA might have a more substantial effect (Bevington et al., 2016; Siwek et al., 2020; Tu et al., 2017).

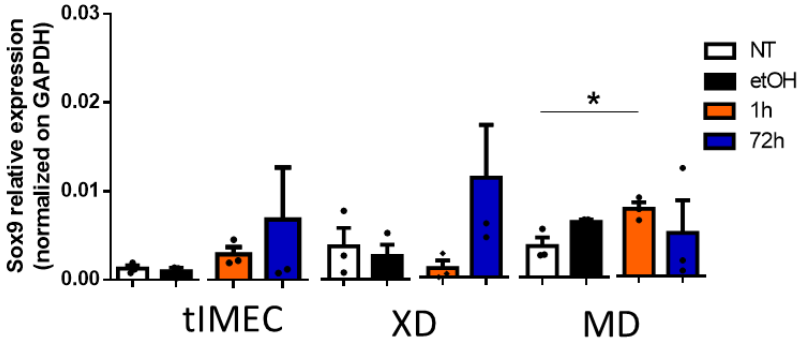


Figure 36 SOX9 expression is augmented by retinoic acid treatment. *SOX9 relative expression level in tIMEC, XD and MD following ATRA 1µM treatment in comparison to non-treated or vehicle treated samples. Average of 4 independent biological replicates +/- S.E.M. Unpaired two-tailed student's t-test. Ns:P > 0.05; *:P ≤ 0.05;*

4.1.4.2. Retinoic acid and SOX9 transcriptional memory

To investigate whether priming MD cells with a pre-treatment of ATRA combined with short time points ATRA treatment would stabilize the effect on *SOX9* expression increase, we designed the following treatment scheme, depicted in **Figure 37**A. Although in tIMEC and in XD cells no significant increase in the response was detected after the priming with ATRA, by quantifying SOX9 transcript levels we found that ATRA priming increased the responsiveness of MD to ATRA treatment, as indicated by the accelerated response and the increment of the transcript abundance. Indeed, upon priming, we measured a 2-fold more robust response than the 30 minutes treatment alone (**Figure 37**B). This suggested that a mechanism of transcriptional memory may strengthen the retinoic acid pathway response and the effect on SOX9 expression induction (Siwek et al., 2020; Zhao et al., 2020).

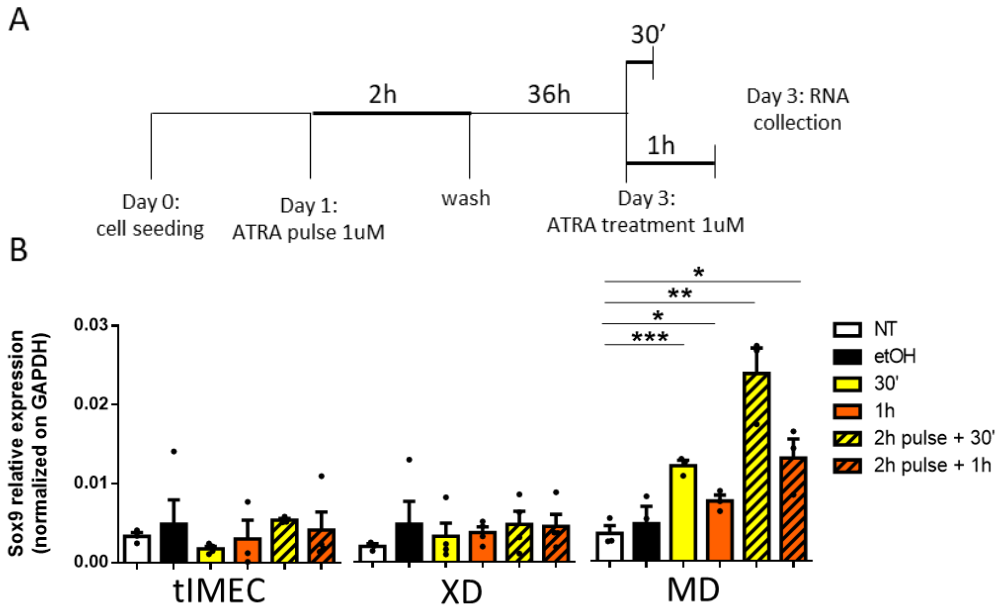


Figure 37: SOX9 expression is strongly induced by ATRA treatment combined with pre-treatment. *A) schematic representation of ATRA treatment: cells were seeded at day 0; ATRA 1µM was added ad day 1 and washed out after 2 hours incubation; after 36 hours, ATRA was re-added for either 30 min or 1 hour before RNA collection. B) SOX9 relative expression levels of tIMEC, XD and MD cells with non-treated, treated with vehicle, treated with short ATRA treatment (30min and 1h) or treated with a 2h ATRA pulse followed by short ATRA treatment (2h pulse +30min/1h). Average of 4 independent biological replicates, +/- S.E.M. Two-way ANOVA test. Ns:P > 0.05; *:P ≤ 0.05; **:P ≤ 0.01;***:P ≤ 0.001.*

To investigate whether SOX9 expression response to ATRA might depend on a specific retinoic acid receptor, we tested the effect of RARα knock-down with two different short interfering RNAs on the response of SOX9 to ATRA treatment.

Analysis of *SOX9* relative expression after ATRA treatment revealed that *SOX9* expression maintenance and response to ATRA treatment was dependent on RARα expression (**Figure 38**). Both short interfering RNAs targeting RARα efficiently and specifically reduced RARα levels in MD cells (**Figure 38**B), decreasing *SOX9* expression in the steady-state and impeding *SOX9* increase in expression following ATRA treatment (**Figure 38**A). Interestingly, ATRA treatment was reducing RARα transcript levels *per se*, suggesting a negative feedback loop.

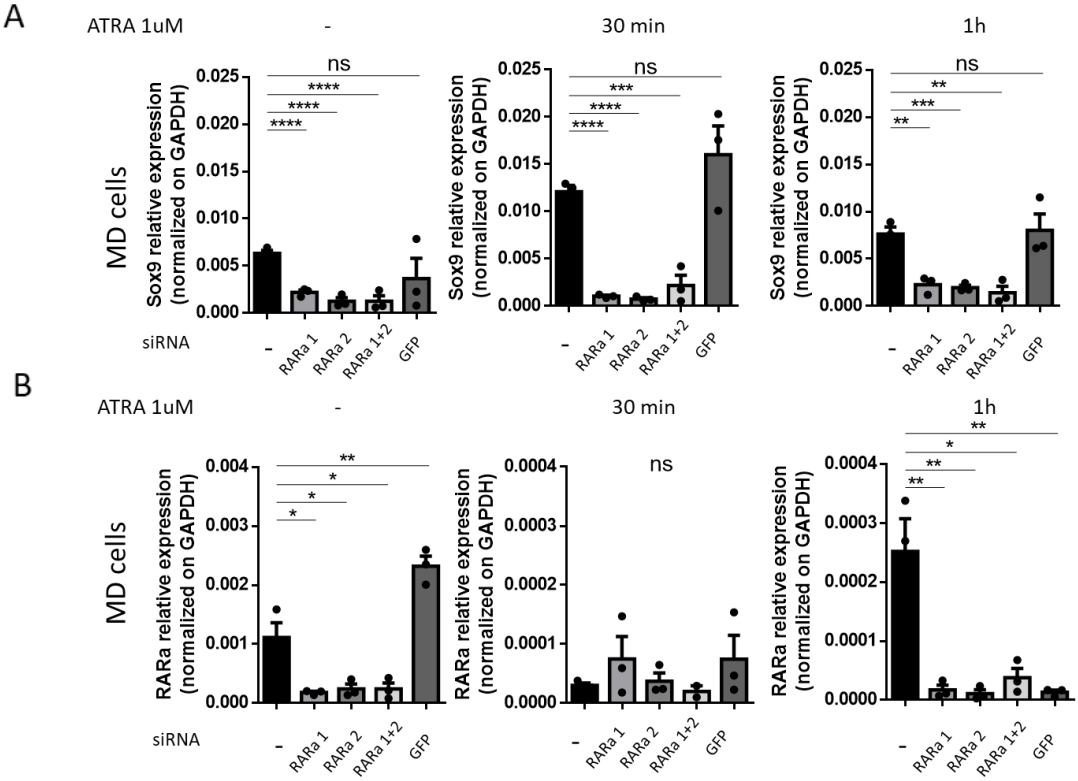


Figure 38 RARα knock-down impedes SOX9 response to ATRA. *A) SOX9 relative expression levels of MD cells with non-treated, treated for 30 minutes, treated for 1 hours with ATRA in cells non-transfected or transfected with RARα targeting siRNA 1, 2, or control siRNAs. B) RARα relative expression levels of the conditions in A). Average of 3 independent biological replicates, +/- S.E.M. Unpaired two-tailed student's t-test. Ns:P > 0.05; *:P ≤ 0.05; **:P ≤ 0.01;***:P ≤ 0.001.*

In order to confirm the cell specificity of SOX9 response to ATRA pathway modulation, we analysed SOX9 response to inhibition of RARα activity through treatment of cells with the RARα specific antagonist BMS614 (**Figure 39**). RARα agonist ligands induce a conformation in which the holo position of helix 12 is stabilized: this active conformation

provides a surface to which coactivators can bind via their NR boxes that contain LxxLL motifs.

In contrast, BMS614 binding allows helix 12 to bind to the static part of the hydrophobic groove of RARα and then blocks by competition both coactivators and corepressor binding.

Through RT qPCR on tIMEC, XD and MD cells, we compared SOX9 transcript levels in steady state or treatment with RARα antagonist at two different time points (**Figure 39**). RARα antagonist did not show any significant effect in tIMEC cells, whereas SOX9 expression was decreased in XD cells and strongly decreased in MD cells. We thus could not conclude that inactivation of RARα activity response was unique in MD cells, although the effect was clearly stronger in MD cells with respect to XD cells and not detectable in tIMEC.

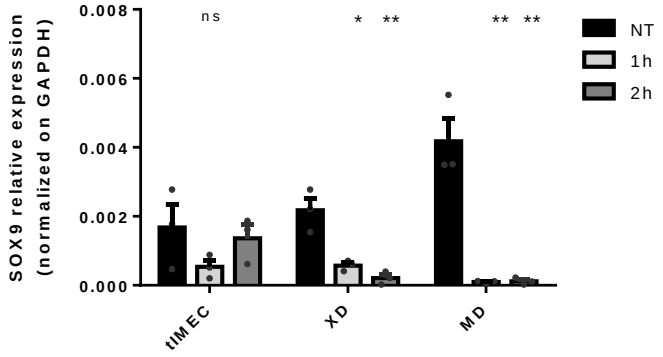


Figure 39 SOX9 expression is decreased following exposure to RARα antagonist *SOX9 relative expression levels in tIMEC, XD and MD cells following BMS614 treatment for 1 or 2 hours. Average of 3 independent biological replicates, +/- S.E.M. Unpaired two-tailed student's t-test. Ns:P > 0.05; *:P ≤ 0.05; **:P ≤ 0.01;***:P ≤ 0.001.*

4.1.4.3. Retinoic acid receptors expression

In order to assess whether *SOX9* expression induction could be due to the augmented expression or activation of the retinoic acid receptors, we proceeded to evaluate protein level and localization of the RARs family members. Therefore, we focused on those receptors that showed to be expressed in the three cell types (**Figure 40**). Steady-state levels and ATRA treatment response of RAR alfa (RARα), RAR gamma (RARγ), RXR alfa (RXRα) and RXR beta (RXRβ) were thus tested through immunostaining.

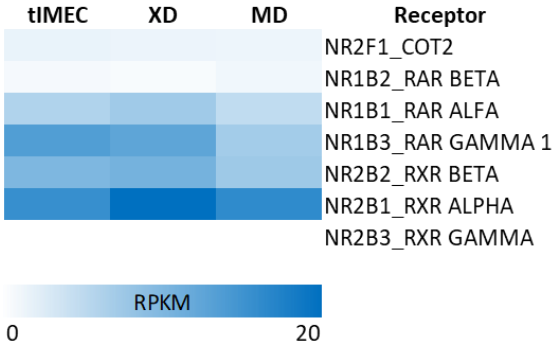


Figure 40 Gene expression heatmap of retinoic acid receptors family members. *Heatmap indicating gene expression data for nuclear receptors involved in the retinoic acid pathway in tIMEC, XD and MD. Average of 3 biological replicates. Color code = RPKM 0 to 20.*

Quantification of RARs nuclear mean intensity at the single-cell level allowed us to assess, in all cases, a bimodal distribution of nuclear mean intensity (panel B of **Figure 41**, **Figure 42**, **Figure 43** and **Figure 44**). Therefore, we analysed nuclei as either positive or negative to the receptor establishing a threshold based on nuclear mean intensity over which we considered cells as positive and under which we considered them negative.

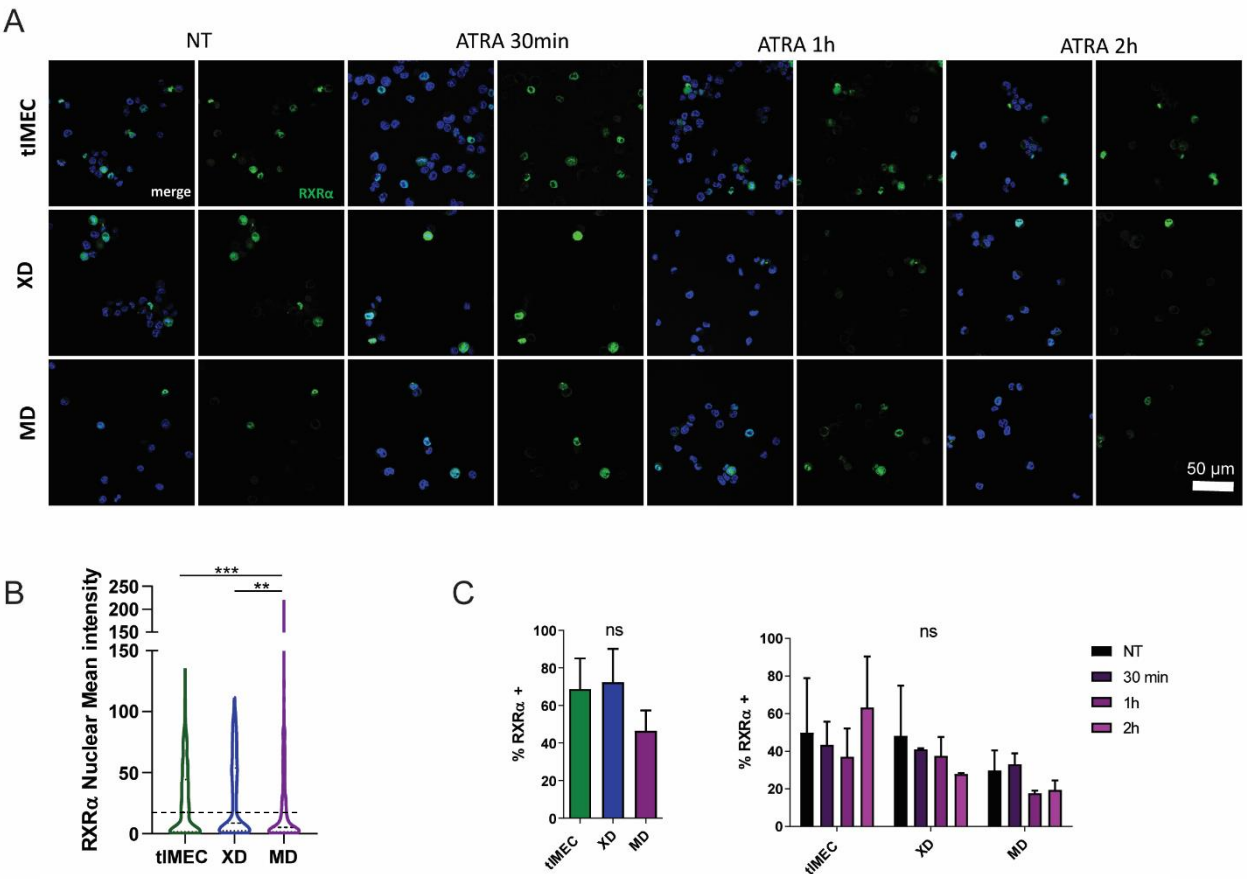


Figure 41 RXRα protein levels in tIMEC, XD and MDs after ATRA treatment. *A) Representative confocal microscope images of RXRα immunostaining in the indicated cell types. RXRα, green; DAPI, blue; scale bar = 50 µm. Images were acquired with a 63X objective. B) Quantification of RXRα nuclear mean intensity in the different cell types; 5 biological replicates; The violin plots indicate median values (middle lines), first and third quartiles (dashed lines) of the fluorescence intensity of analysed nuclei (n≥300). C) Quantification of RXRα+ nuclei percentage in steady state (left) and after ATRA treatment (right) in the indicated cell types; staining positivity was assessed on the basis of the distribution of data in B); for RXRα, nuclei with mean intensity>10 were considered positive. Average of 5 biological reps, +/- S.E.M. Unpaired two-tailed student's t-test. Ns:P > 0.05; *:P ≤ 0.05; **:P ≤ 0.01;***:P ≤ 0.001;****:P ≤ 0.0001.*

Quantification of RXRα nuclear mean intensity showed a slight decrease in MD cells (**Figure 41** panel B), although the difference when assessing the changes in the percentage of positive cells was not statistically significant (**Figure 41** panel C); the same holds true when comparing ATRA treatment time points with respective steady-state conditions. In the steady state though, MD cells showed a small population of very bright RXRα+ cells, which may be of interest for further analysis.

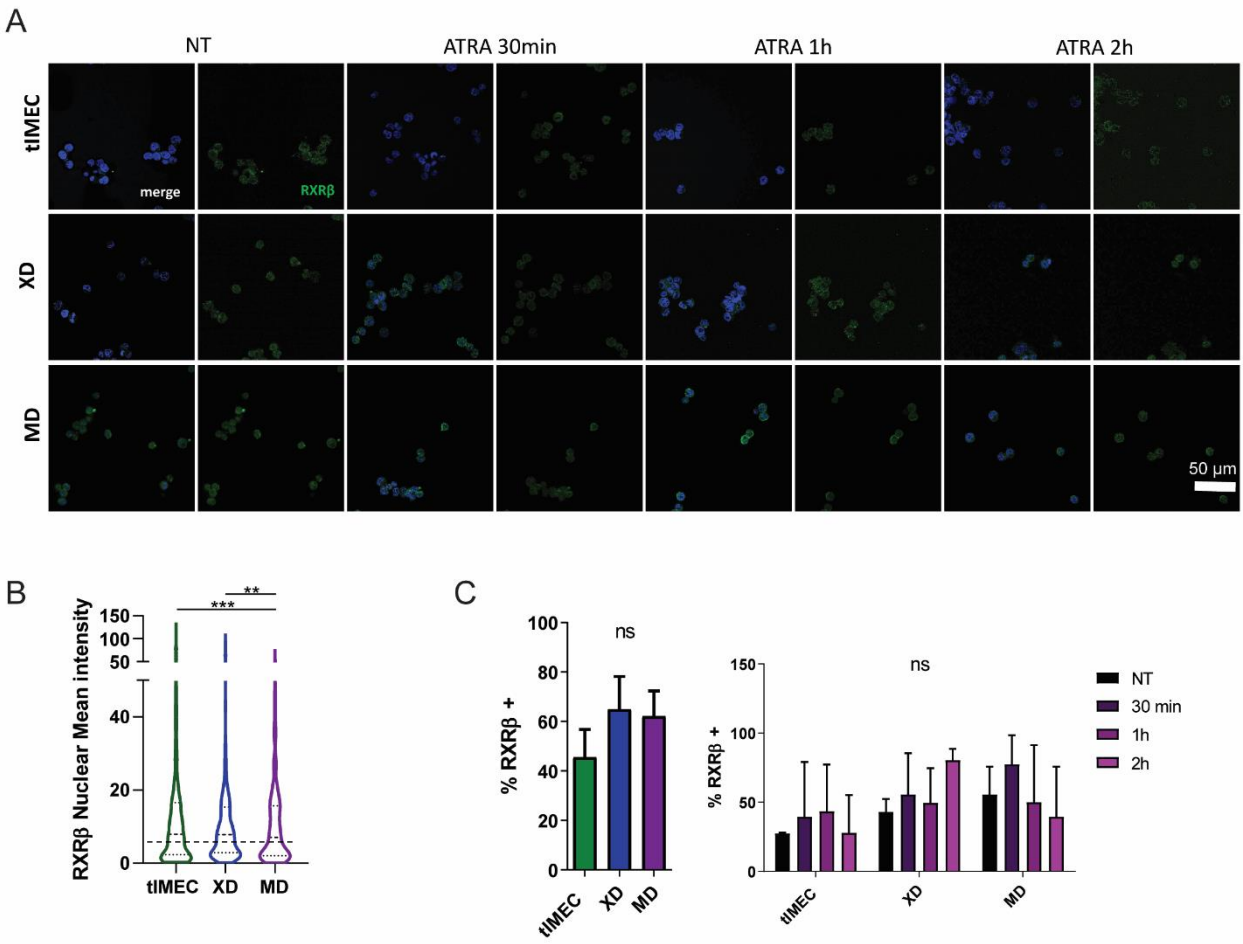


Figure 42 RXRβ protein levels in tIMEC, XD and MDs after ATRA treatment. *A) Representative confocal microscope images of RXRβ immunostaining in the indicated cell types. RXRβ, green; DAPI, blue; scale bar = 50 µm. Images were acquired with a 63X objective. B) Quantification of RXRβ nuclear mean intensity in the different cell types; 5 biological replicates; The violin plots indicate median values (middle lines), first and third quartiles (dashed lines) of the fluorescence intensity of analysed nuclei (n≥300). C) Quantification of RXRβ+ nuclei percentage in steady state (left) and after ATRA treatment (right) in the indicated cell types; staining positivity was assessed on the basis of the distribution of data in B); for RXRβ, nuclei with mean intensity>10 were considered positive. Average of 5 biological reps, +/- S.E.M. Unpaired two-tailed student's t-test. Ns:P > 0.05; *:P ≤ 0.05; **:P ≤ 0.01;***:P ≤ 0.001;****:P ≤ 0.0001.*

Quantification of RXRβ nuclear mean intensity allowed us to assess that nuclear protein levels were significantly lower than RXRα, but it showed a slight increase in XD and MD cells (**Figure 42** panel B): although the trend was clear, the difference was not statistically significant when comparing the percentages of positive cells between the cell populations (**Figure 42** panel C): this may be due to those extremely bright cells that elongated the distribution tail and that, when comparing the three cell populations, was decreasing from tIMEC to MD cells; the same holded true when comparing ATRA treatment time points with respective steady-state conditions.

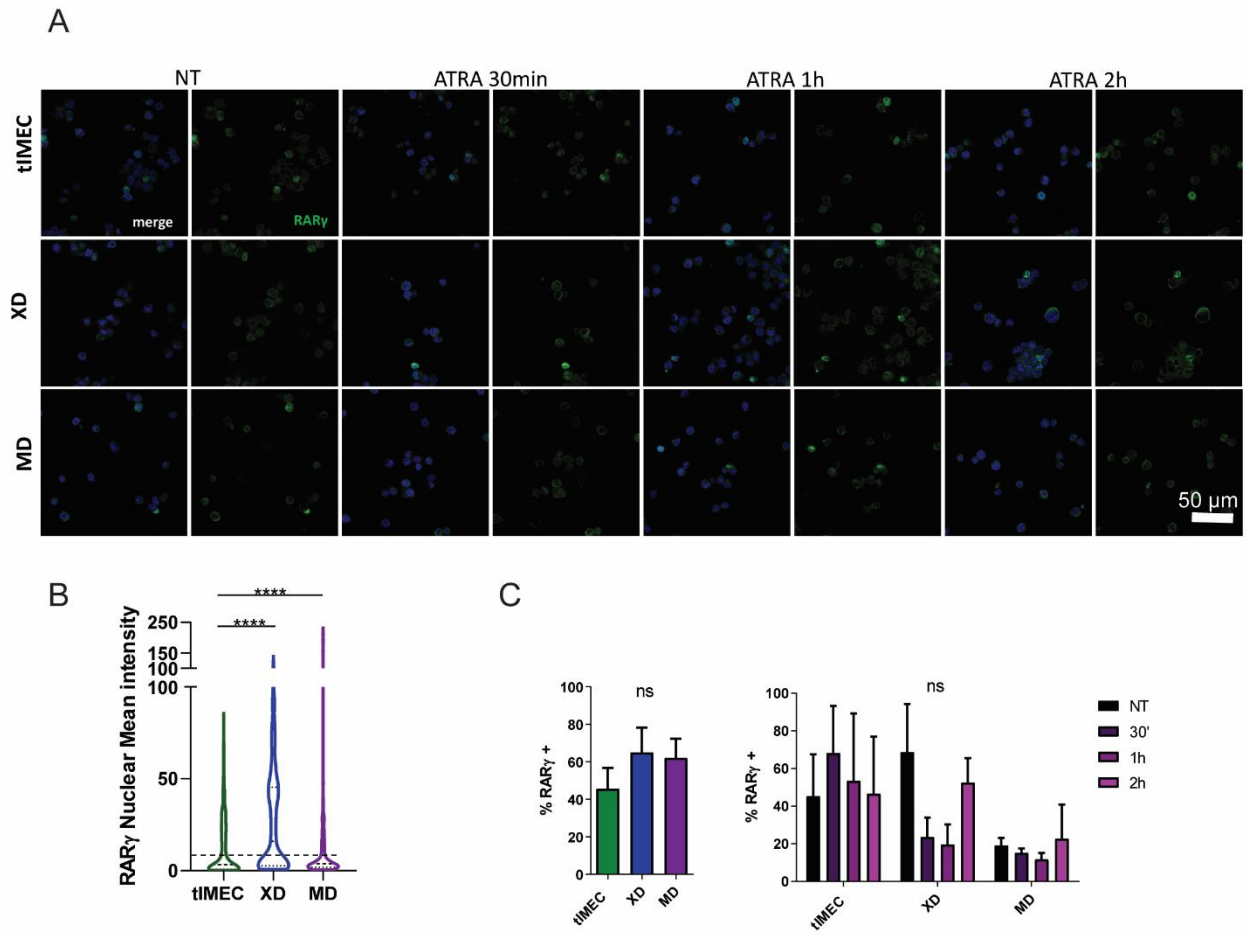


Figure 43 RARγ protein levels in tIMEC, XD and MDs after ATRA treatment. *A) Representative confocal microscope images of RARγ immunostaining in the indicated cell types. RARγ, green; DAPI, blue; scale bar = 50 µm. Images were acquired with a 63X objective. B) Quantification of RARγ nuclear mean intensity in the different cell types; 5 biological replicates; The violin plots indicate median values (middle lines), first and third quartiles (dashed lines) of the fluorescence intensity of analysed nuclei (n≥300).C) Quantification of RARγ+ nuclei percentage in steady state (left) and after ATRA treatment (right) in the indicated cell types; staining positivity was assessed on the basis of the distribution of data in B); for RARγ, nuclei with mean intensity>10 were considered positive. Average of 5 biological reps, +/- S.E.M. Unpaired two-tailed student's t-test. Ns:P > 0.05; *:P ≤ 0.05; **:P ≤ 0.01;***:P ≤ 0.001;****:P ≤ 0.0001.*

Quantification of RARγ nuclear mean intensity showed that RARγ protein level slightly increased in XD and MD cells (**Figure 43** panel B), but the difference was not statistically significant when quantifying the percentage of positive cells (**Figure 43** panel C); distribution analysis in panel B shows that although MD cells that result RARγ+ were brighter than in the other two cell populations, the positive population was smaller and thus less represented than in the other two samples; RARγ protein levels slightly decreased when comparing ATRA treatment time points with respective steady-state conditions.

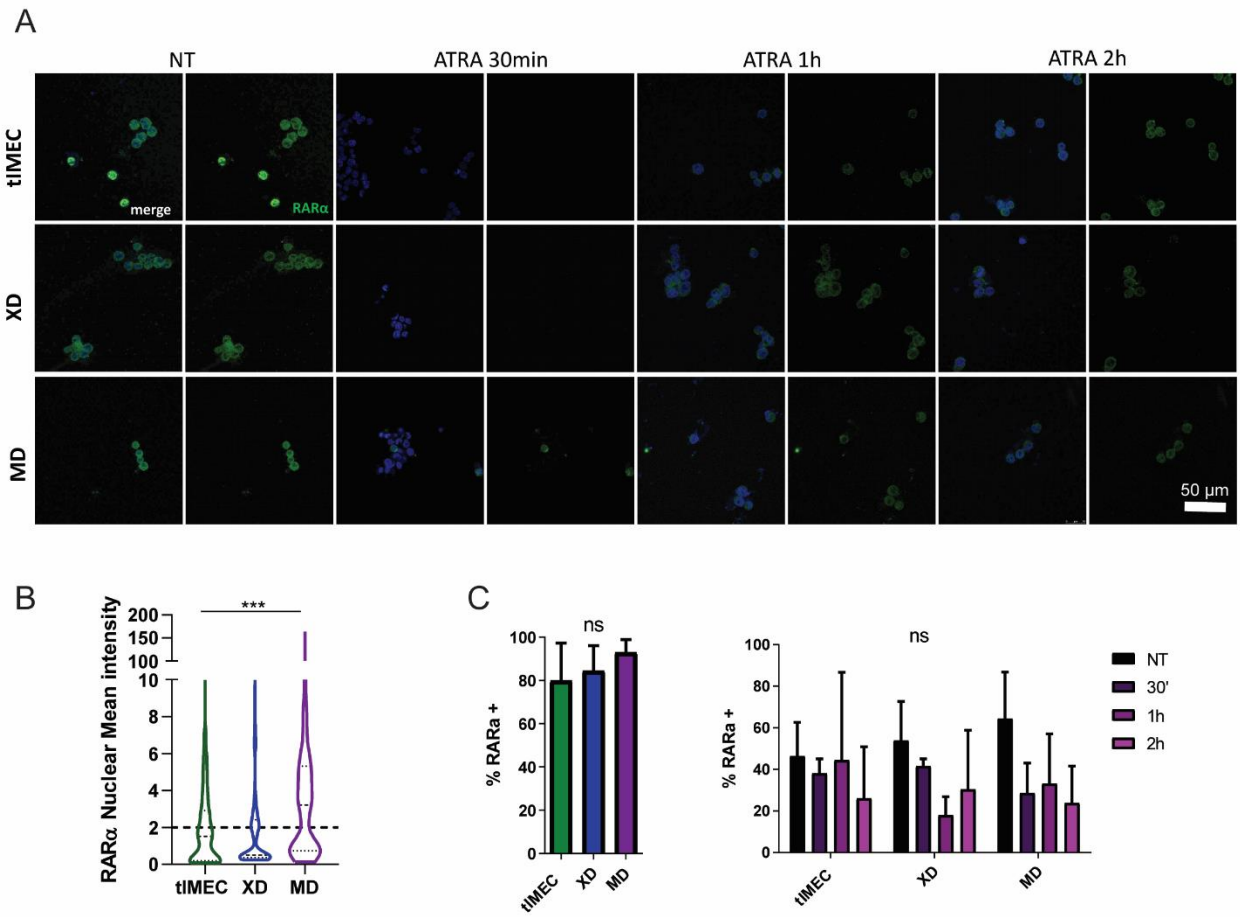


Figure 44 RARα protein levels in tIMEC, XD and MDs after ATRA treatment. *A) Representative confocal microscope images of RARα immunostaining in the indicated cell types. RARα, green; DAPI, blue; scale bar = 50 µm. Images were acquired with a 63X objective. B) Quantification of RARα nuclear mean intensity in the different cell types; 5 biological replicates; The violin plots indicate median values (middle lines), first and third quartiles (dashed lines) of the fluorescence intensity of analysed nuclei (n≥300). C) Quantification of RARα+ nuclei percentage in steady state (left) and after ATRA treatment (right) in the indicated cell types; staining positivity was assessed on the basis of the distribution of data in B); for RARα, nuclei with mean intensity>2 were considered positive. Average of 3 biological reps, +/- S.E.M. Unpaired two-tailed student's t-test. Ns:P > 0.05; *:P ≤ 0.05; **:P ≤ 0.01;***:P ≤ 0.001;****:P ≤ 0.0001.*

Finally, quantification of RARα nuclear mean intensity showed a slight increase in RARα nuclear protein levels in XD and MD cells (**Figure 44** panel B), but once again, the difference was not statistically significant when quantifying the percentage of RARα+ cells (**Figure 44** panel C); RARα protein levels slightly decreased when comparing ATRA treatment time points with corresponding steady-state conditions, in line with the RNA expression evaluation (**Figure 38**) and the hypothesis of a negative feedback loop regulating ATRA pathway induction and ATRA responsive receptors.

We noticed that RARα signal localization was predominantly cytosolic, especially before ATRA treatment: this was expected since multiple pathways regulate RARα nuclear to cytosol translocation (Braun et al., 2000; Santos and Kim, 2010). Nevertheless, we tested RARα antibody specificity by comparing immunostaining in tIMEC for RARα in RARα-knock-down cells versus non treated cells. As in tIMEC cells RARα localization was almost entirely cytosolic, instead of quantifying exclusively nuclear intensity, the total signal was quantified (**Figure 45**A). A residual mean intensity of RARα could be detected after knock-down (**Figure 45**B), suggesting that the efficiency of the repression of RARα expression may not be total. Nevertheless, RARα signal intensity was strongly reduced at 48 hours after transfection of tIMEC with RARα siRNA (**Figure 45**B), indicating that the antibody was specifically binding to the RARα receptor.

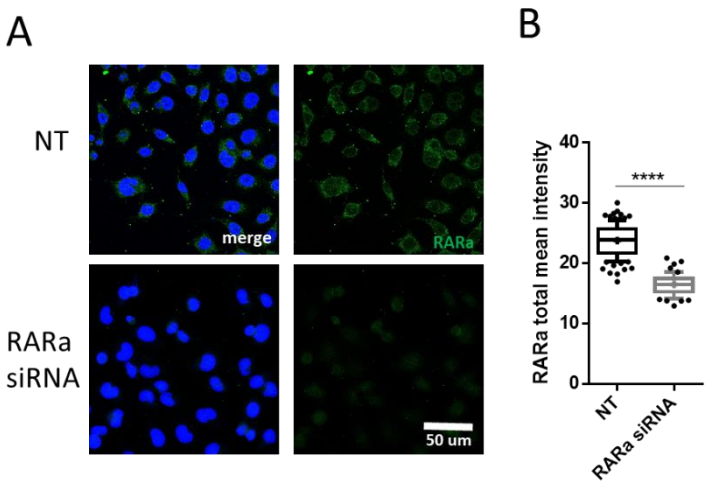


Figure 45 RARα antibody specificity testing. *RARα immunostaining in tIMEC transfected with RARα siRNA versus non treated cells. A) panel of representative images acquired with a confocal microscope showing RARα efficient knock-down in tIMEC cells; RARα, green; DAPI, blue; scale bar = 50 µm. Images were acquired with a 63X objective. B) Quantification of RARα total mean intensity in the two indicated conditions; 1 biological replicate, N=100; The box plots indicate median values (middle lines), first and third quartiles (box edges) and 10th and 90th percentiles (error bars) of the fluorescence intensity of analysed nuclei. Unpaired two-tailed student's t-test. Ns:P > 0.05; *:P ≤ 0.05; **:P ≤ 0.01;***:P ≤ 0.001;****:P ≤ 0.0001.*

Given that for all retinoic acid receptors we were not able to assess a substantial variation in protein levels or localization in the different cell types or following retinoic acid treatment, we hypothesised that significant changes, as for SOX9 expression, would be visible only after a pre-treatment with ATRA. Furthermore, we noticed that the distribution of RXRα and RARα in the nuclei suggested a non-homogeneous distribution.

RXRα was shown to contain an intrinsically disordered region (IDR) at the N-terminal domain, and recent findings suggest that IDRs may favour the assembly of molecular condensates, where TFs, cofactors and transcriptional machinery dynamically cluster at regulatory elements to regulate transcription (Belorusova et al., 2016; Cho et al., 2018; Chong et al., 2018). The RARα/RXRα dimer was suggested to play a crucial role in the recruitment of UTX, p300 and MLL4 and in regulating the active enhancer landscape by inducing the formation of a physical bridge that facilitates the formation of a physical bridge p300 and MLL4 functions (Wang et al., 2017). We thus speculated that clustering of those RA receptors more than changes in protein localization or levels would be responsible for the response to ATRA in terms of gene expression, possibly by facilitating the assembly of transcriptional condensates as previously assessed for numerous chromatin players and transcription factors (Boija et al., 2018; Fasciani et al., 2020; Sabari et al., 2018).

4.1.4.4. Retinoic acid receptors clustering

In order to test the behaviour of retinoic acid receptors by means of clustering capacity, immunostaining of RXRα and RARα receptors in MD cells was performed, comparing steady-state with very short ATRA treatments (5/15/30 minutes) preceded or not by a 2 hours ATRA pulse followed by 36 hours wash. The most effective time points were the shortest, *i.e*. 5 minutes of treatment combined (or not) with a two-hour pulse.

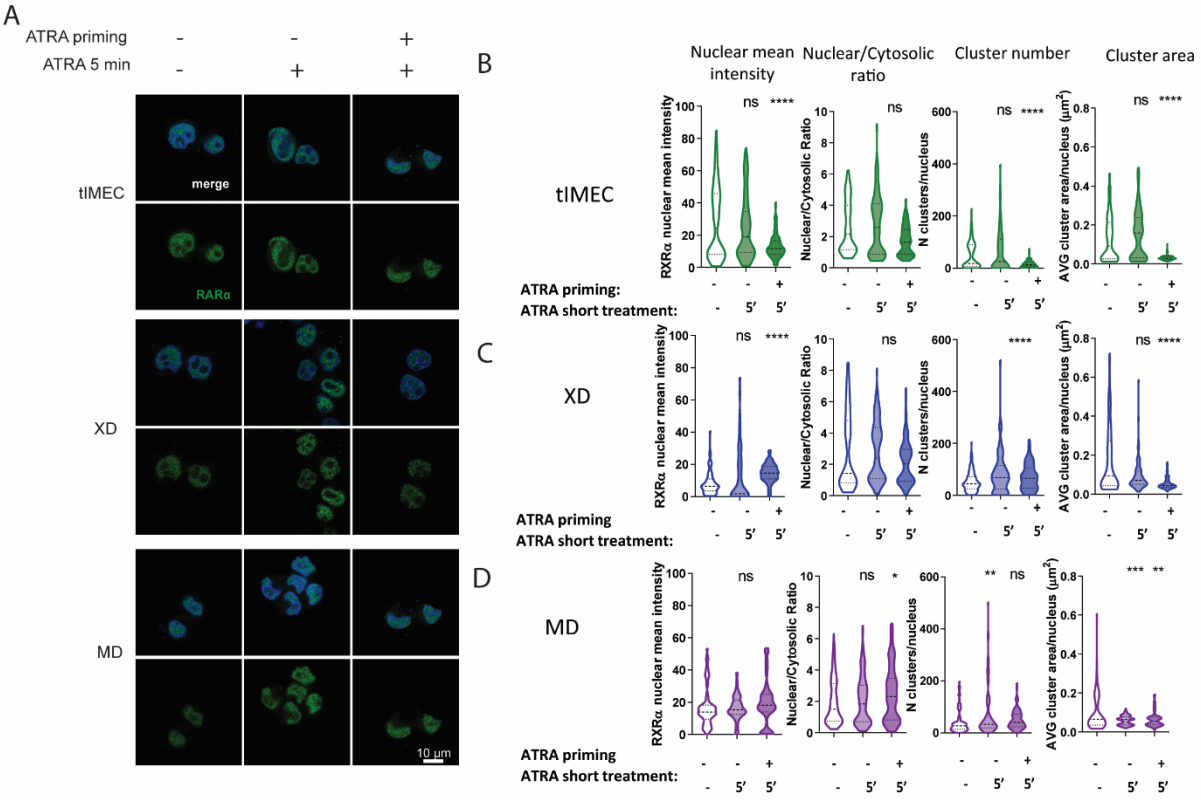


Figure 46 RXRα clustering in MD cells after ATRA treatment. *A) Representative confocal microscope images of RXRα immunostaining in MD cells at the indicated time points of ATRA treatment. RXRα, green; DAPI, blue; scale bar = 10 µm. Images were acquired with a 63X objective and 4X zoom. B-D) in tIMEC-XD-MD respectively, from left to right: quantification of RXRα nuclear mean intensity at the different time points of ATRA treatment; quantification of the nuclear mean intensity ratio between nucleus and cytosol; quantification of the number of clusters per nuclei; quantification of the average cluster area per nuclei. 2 biological replicates (N≥100); The violin plots indicate median values (middle lines), first and third quartiles (dashed lines) of the fluorescence intensity of analysed nuclei. Unpaired two-tailed student's t-test. Ns:P > 0.05; *:P ≤ 0.05; **:P ≤ 0.01;***:P ≤ 0.001;****:P ≤ 0.0001.*

RXRα nuclear protein level quantification showed a linear increase after ATRA treatment in XD and MD cells, displaying the highest nuclear intensity after the combination of ATRA priming and 5 minutes treatment (**Figure 46**A, C and D), although nuclear to cytosol ratio calculation reflected an increase of total protein levels in XD cells, whereas in MD cells, protein was more concentrated in the nucleus than in the cytosol after 5 minutes of treatment combined with the priming (**Figure 46** D). In tIMEC, RXRα total protein levels decreased after ATRA treatment (**Figure 46**A and B).

Furthermore, we performed a cluster analysis to assess the number, intensity and average size of RXRα clusters in tIMEC, XD and MD cells: a higher number of clusters per nucleus was detected after treatment when compared to steady-state (**Figure 46**C and D) in XD and MD cells, whereas tIMEC did not display an appreciable response (**Figure 46**B). Cluster quantification indicated that a significantly lower cluster area characterized the clusters formed by RXRα after ATRA treatment with respect to non-treated cells in the three cell types.

Although for MD cells and XD cells the diminishment of cluster area may be linked to the increased number, we could not explain the trend that characterized tIMEC: further validation through optogenetics approaches might thus be needed to ensure the dynamics of RXRα clustering. Furthermore, since the significant increase in RXRα nuclear protein level could not be explained by shuttling of the protein from the cytosol to the nucleus, and that was generally visible after the combination of the short treatment with the priming, further investigation and assessment of the protein levels after the priming alone should be considered for future experiments.

Quantification of RARα nuclear mean intensity allowed us to assess a decrease in nuclear protein levels after ATRA treatment in tIMEC and XDs (**Figure 47**A, B and C), and a slight increase in MD cells (**Figure 47**D); the ratio between protein levels in the nucleus and in the cytosol was slightly increasing in the three cell types, even though not significantly in XD cells (**Figure 47**B, C and D): this suggested a mechanism of shuttling of RARα protein from the cytoplasm to the nucleus after ATRA treatment.

Furthermore, we were able to detect RARα clustering in all three cell types (**Figure 47**B, C and D), but quantification of clusters indicated that ATRA treatment induced

reduction of cluster number in both tIMEC and XD cells (**Figure 47**B and C), inducing an increase of cluster number in MD cells (**Figure 47**D) instead; this was in line with the augmentation of nuclear protein signal in response to ATRA treatment. Cluster mean dimension decreased after ATRA treatment in tIMEC and XD cells independently of the

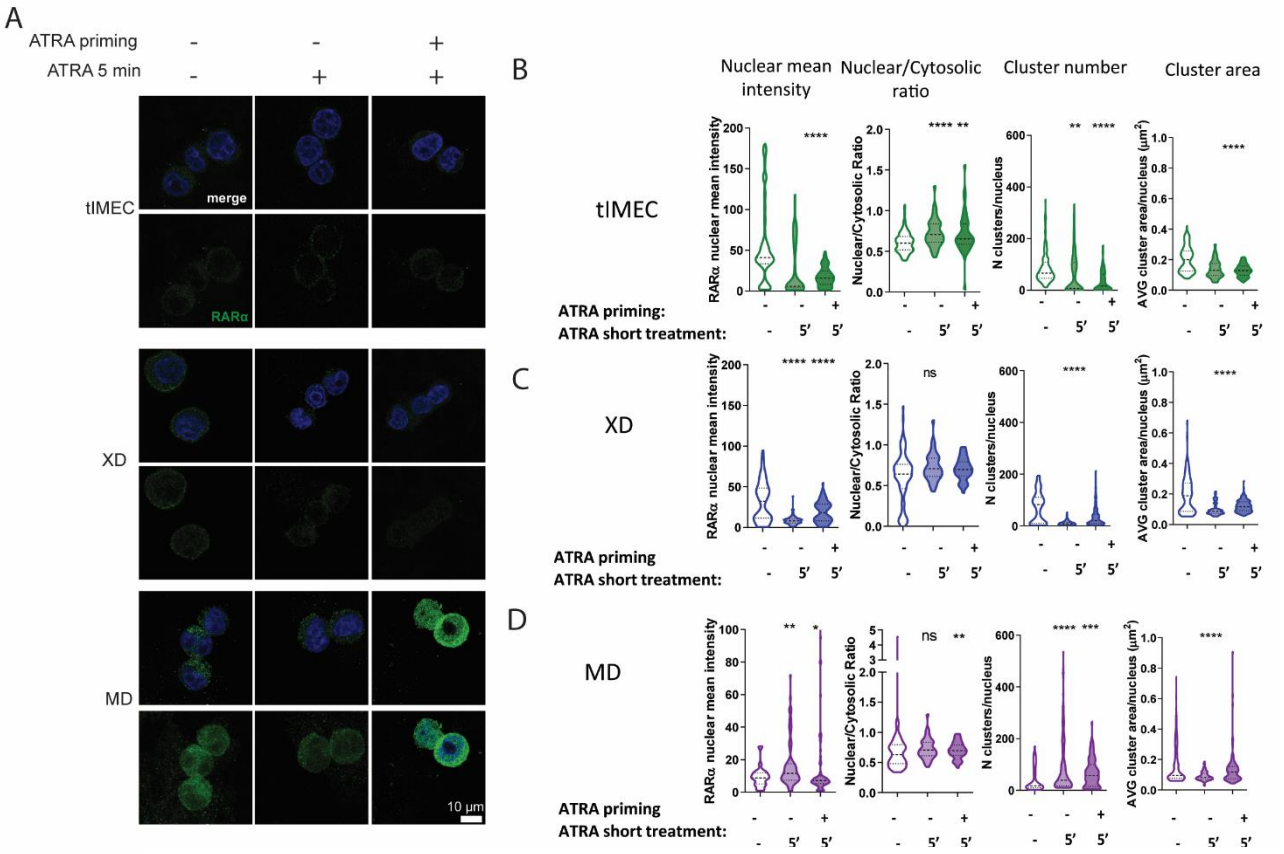


Figure 47 RARα clustering in MD cells after ATRA treatment**. **A)** Representative confocal microscope images of RARα immunostaining in MD cells at the indicated time points of ATRA treatment. RARα, green; DAPI, blue; scale bar = 10 µm. Images were acquired with a 63X objective and 4X zoom. **B-D)** in tIMEC-XD-MD respectively, from left to right: quantification of RARα nuclear mean intensity at the different time points of ATRA treatment; quantification of the nuclear mean intensity ratio between nucleus and cytosol; quantification of the number of clusters per nuclei; quantification of the average cluster area per nuclei. 2 biological replicates (N≥100); The violin plots indicate median values (middle lines), first and third quartiles (dashed lines) of the fluorescence intensity of analysed nuclei. Unpaired two-tailed student's t-test. Ns:P > 0.05; *:P ≤ 0.05; **:P ≤ 0.01;:P ≤ 0.001;****:P ≤ 0.0001.*

treatment, whereas in MD cells, although a decrease was visible after the short treatment, cluster area slightly increased when the short treatment was combined with the priming. Tuning the timing of the treatment might be necessary to better evaluate clustering dynamics, and coupling this experimental setting with optogenetics approaches might better elucidate the clustering response to activation of the retinoic acid pathway. Nevertheless, these results indicated that retinoic acid receptors protein levels change after ATRA short treatment combined with a pre-exposure to the

compound. Furthermore, both RARα and RXRα could form clusters following ATRA treatment, although with different dynamics.

4.1.4.5. Retinoic acid, SOX9 and proliferation inhibition

In order to investigate whether ATRA is linked to induction of proliferation inhibition in metastatic cells as suggested by previous evidence, we proceeded with testing the expression of proliferation markers (Sosa et al., 2015). We tested the abundance of the proliferation marker Ki67 in tIMEC, XD and MD cells after ATRA treatment for 72 hours (**Figure 48**A), and then tested cell proliferation through a dye-retaining assay in response to short-term ATRA treatment in both in primed and un- primed cells (**Figure 49**).

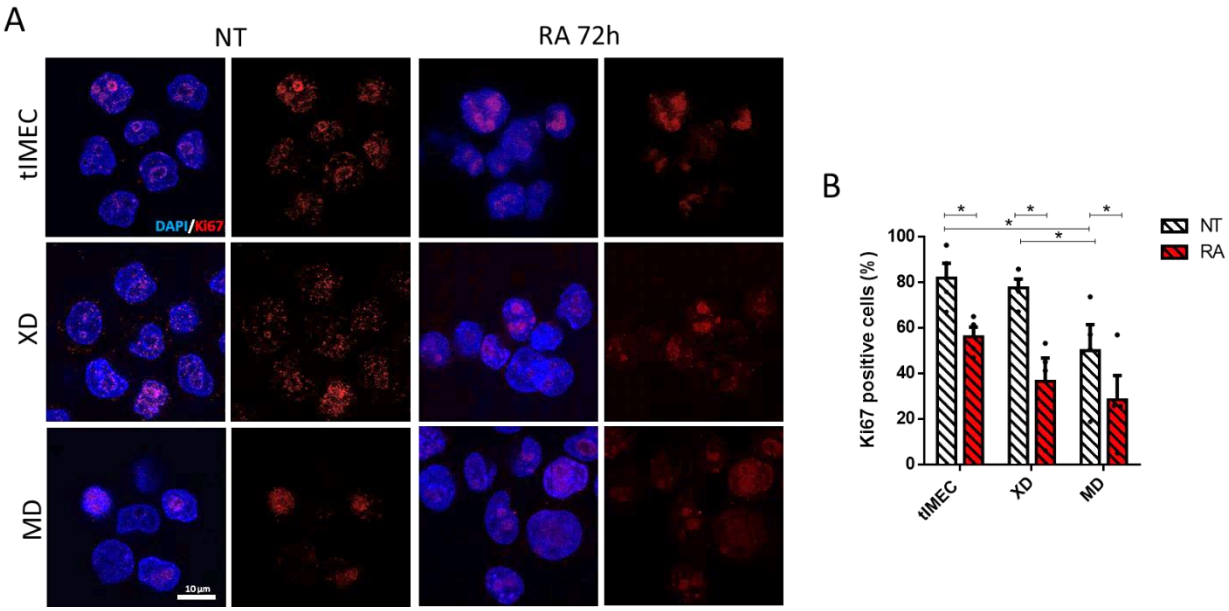


Figure 48 Ki67 Immunostaining in tIMEC, XD and MD cells after ATRA treatment. *A) Representative confocal microscope images of Ki67 immunostaining in tIMEC, XD and MD cells after 72h treatment with ATRA. Ki67, red; DAPI, blue; scale bar = 10 µm. Images were acquired with a 63X objective and 4X zoom. B) Quantification of percentage of KI67+ cells at the different time points of ATRA. Ki67 + cells were identified as mean intensity >10 nuclei. Average of 4 independent biological replicates +/- S.E.M.; Unpaired one-tailed student's t-test. Ns:P > 0.05; *:P ≤ 0.05.*

We quantified the percentage of Ki67+ cells with respect to the total number of cells analysed; when comparing the percentage of Ki67+ cells in tIMEC, XD and MD, we could

assess that in the steady-state, MD cells showed a decrease of Ki67 levels with respect to both tIMEC and XD cells. Furthermore, ATRA treatment significantly decreased Ki67 levels in the three cell types (**Figure 48**B), showing the most pronounced effect in MD cells, decreasing from 50% Ki67-positive cells to approximately 25% after ATRA treatment. This suggested that ATRA treatment may induce a decrease in the proliferation of tIMEC, XD and MD cells.

As an orthologous experiment, we analysed the proliferation of tIMEC, XD and MD cells after ATRA treatment through dye-retaining capacity evaluation: cells were stained with a far-red fluorescent cell membrane labelling reagent (CellVue ^{TM} Maroon) before seeding for ATRA treatment, and intensity staining is evaluated after four (**Figure 49** A, B, C and D) and eight days (**Figure 49** E and F). Actively proliferating cells will dilute the cell labelling reagent, resulting in lower general signal intensity, whereas quiescent cells will retain the initial staining intensity.

We quantified the cell population that retained after four days the same signal intensity as cells analysed immediately after staining with the dye in order to assess the percentage of quiescent cells (**Figure 49** A and B): quantification displays that in the steady-state XD cells and MD cells show a greater non-proliferating cell population with respect to tIMEC, even though only MD cells increase in percentage resulted statistically significant.

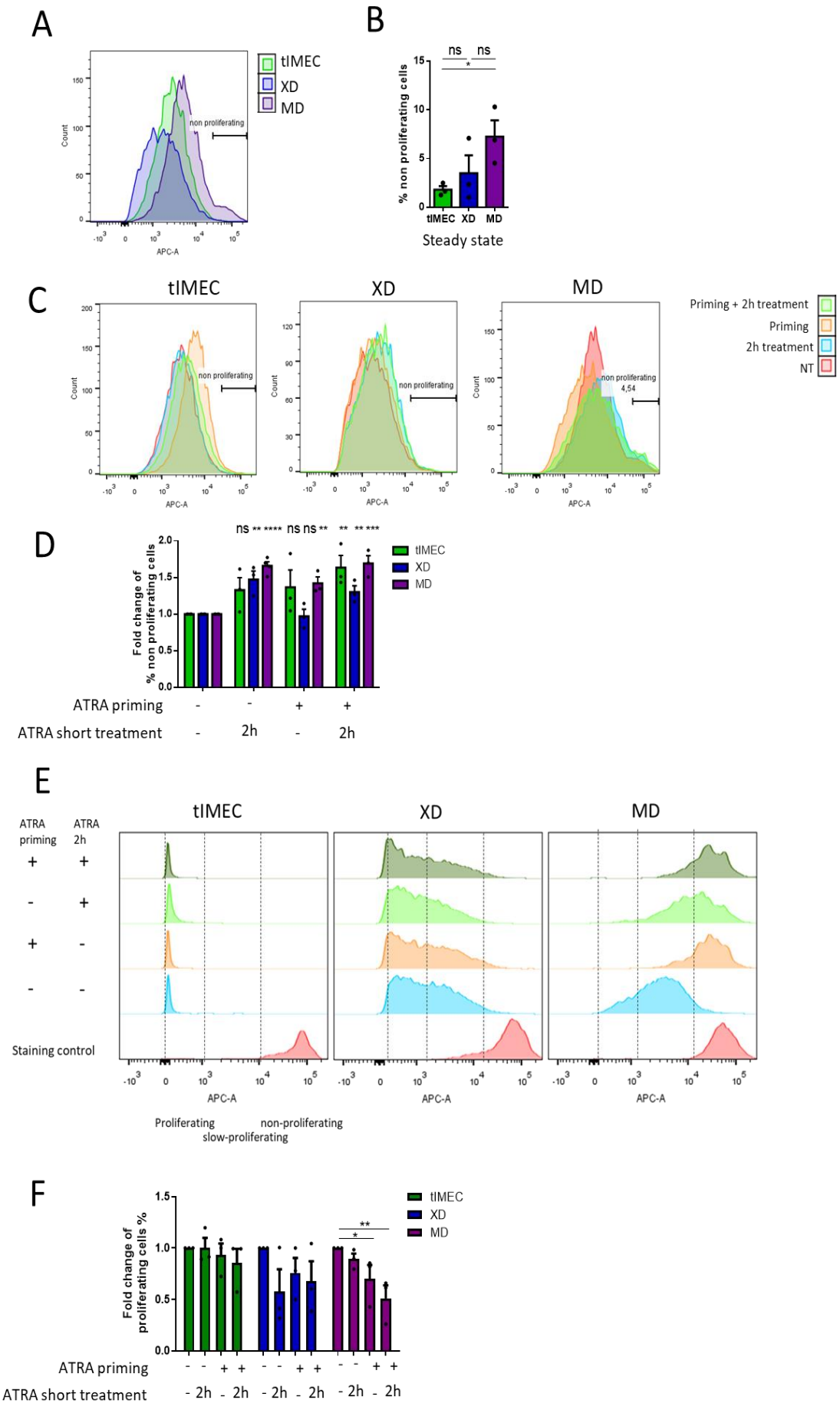

A
tIMEC
XD
MD
non proliferating
Count
APC-A
B
ns
ns
*
% non proliferating cells
tIMEC XD MD
Steady state
C
tIMEC
XD
MD
Priming + 2h treatment
Priming
2h treatment
NT
non proliferating
non proliferating 4,54
Count
APC-A
D
ns ** **** ns ns ** ** ** ***
Fold change of % non proliferating cells
tIMEC
XD
MD
ATRA priming - - + +
ATRA short treatment - 2h - 2h
E
tIMEC
XD
MD
ATRA priming
ATRA 2h
+ +
- +
+ -
- -
Staining control
APC-A
Proliferating
non-proliferating
slow-proliferating
F
Fold change of proliferating cells %
tIMEC
XD
MD
**
*
ATRA priming - - + + - - + + - - + +
ATRA short treatment - 2h - 2h - 2h - 2h - 2h - 2h

Figure 49 MD cells proliferation is inhibited by ATRA treatment. *FACS analysis of dye-retaining capacity in tIMEC, XD and MD cells. A) Histogram of FACS analysis of tIMEC, XD and MD cells showing CellVue Maroon staining intensity (APC-A) 96h after seeding with quantification of non-proliferating cell percentage over total in B). C) Histograms of FACS analysis in tIMEC, XD and MD cells showing CellVue Maroon staining intensity (APC-A) 96h after seeding (NT) in comparison with indicated ATRA treatment conditions. D) Quantification of non-proliferating cell percentage over total of the conditions in C) relative to the steady state. Gating for quantification of non-proliferating cells was set on cells analysed just after staining. E) Histogram of FACS analysis in tIMEC, XD and MD cells showing CellVue Maroon staining intensity (APC-A) versus SSC-A in staining control and steady state cells compared to cells primed only (priming +, 2h-), treated only with short treatment (priming -, 2h+), or primed and treated with short treatment (priming +, 2h +) after 8 days from seeding. F) Quantification of proliferating cell percentage over total of the conditions in D) relative to steady state. Gating for quantification of proliferating cells was set on cells analysed just after staining (as shown in panel E)). Average of 3 independent biological replicates +/- S.E.M.; Unpaired one-tailed student's t-test. Ns:P > 0.05; *:P ≤ 0.05; **:P ≤ 0.01;***:P ≤ 0.001;*

After ATRA treatment, all three cell populations showed an increase in the percentage of quiescent cells, although MD cells showed the most robust grade of responsiveness (**Figure 49**C and D). In order to assess whether the evaluation of MD cell response to ATRA treatment would be more sustained after longer time points, we analysed the percentages of proliferating cells eight days after seeding (**Figure 49**E). Indeed, the analysis of the percentage of cells capable of diluting the dye after eight days showed that the amount of proliferating MD cells decreased to almost half with respect to the steady-state after ATRA treatment: this effect resulted to be even more marked when the treatment was combined with the priming (**Figure 49**F). The same effect was not detectable in tIMEC and XD cells.

With this, we could conclude that the ATRA effect on proliferation was detectable both with long-term treatment and with short-term treatment combined with pre-treatment, and it translated in sustaining cell quiescence, especially in metastatic-derived cells.

At this point, we were interested in assessing whether the ATRA response in terms of proliferation inhibition may include a role for *SOX9*. Thus, we transduced tIMEC cells with a tet-on system containing vector to overexpress *SOX9* and transduced MD cells with one shRNA containing vector to knock-down *SOX9* in MD cells.

Western Blot analysis and relative quantification of SOX9 optical density on three biological replicates allowed us to assess that both overexpression and knock-down of SOX9 were efficient (**Figure 50**A). Transcript level augmentation in tIMEC overexpressing SOX9 and reduction in MD sh SOX9 were confirmed through RT-qPCR(**Figure 50**B): quantification of transcript levels indicates that SOX9 is decreased of 2 fold when

comparing SOX9 knock-down in MD cells to SOX9 overexpression in tIMEC; 2-fold increase is also visible in MD cells with respect to tIMEC, which confirms previous data

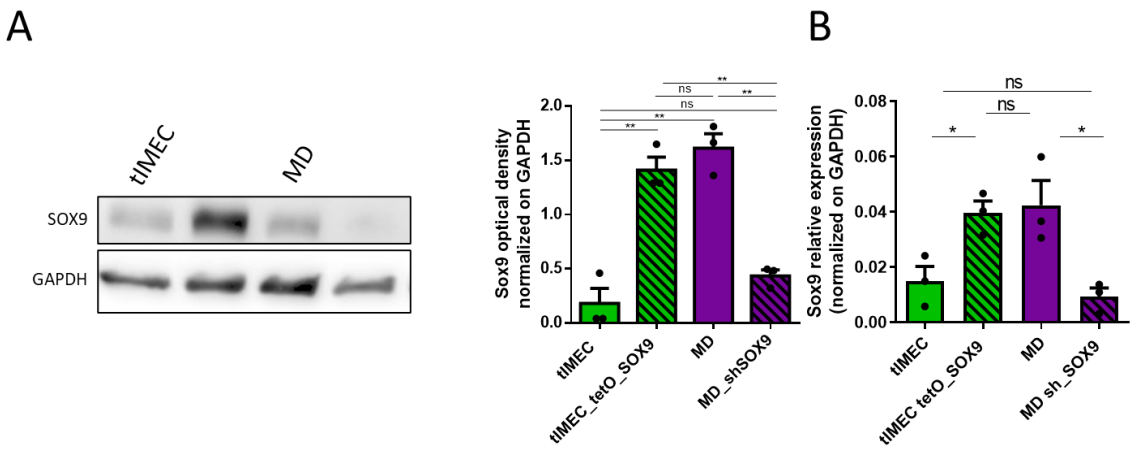


Figure 50 SOX9 protein and transcript levels in knocked-down and overexpressing cell types. *A) Representative Western Blot of SOX9 and GAPDH in tIMEC; tIMEC overexpressing SOX9, MD cells, and MD cells with SOX9 knock-down (left). SOX9 relative protein level quantification in the indicated cell types (right). B) RT-qPCR analysis of SOX9 transcript levels relative to GAPDH. Average of 3 independent biological replicates +/- S.E.M. Unpaired two-tailed student's t-test. Ns:P > 0.05; *:P ≤ 0.05; **:P ≤ 0.01.*

(**Figure 25**). Since the aim was to bring tIMEC SOX9 levels to MD levels and vice versa, leaky expression of the tetO-system in tIMEC without doxycycline induction was sufficient to achieve an increase of SOX9 protein to a level comparable to the MD cells.

The proliferation quantification of these two cell types compared to parental ones was then performed by analysing Ki67 levels. Quantification of SOX9 nuclear mean intensity and the percentage of Ki67 positive cells showed that SOX9 overexpression in tIMEC (**Figure 51** A and B) was concomitant to sustained quiescence (**Figure 51**C), whereas knock-down of SOX9 in MD (**Figure 51**A and B) was concurrent to proliferation increase (**Figure 51**C).

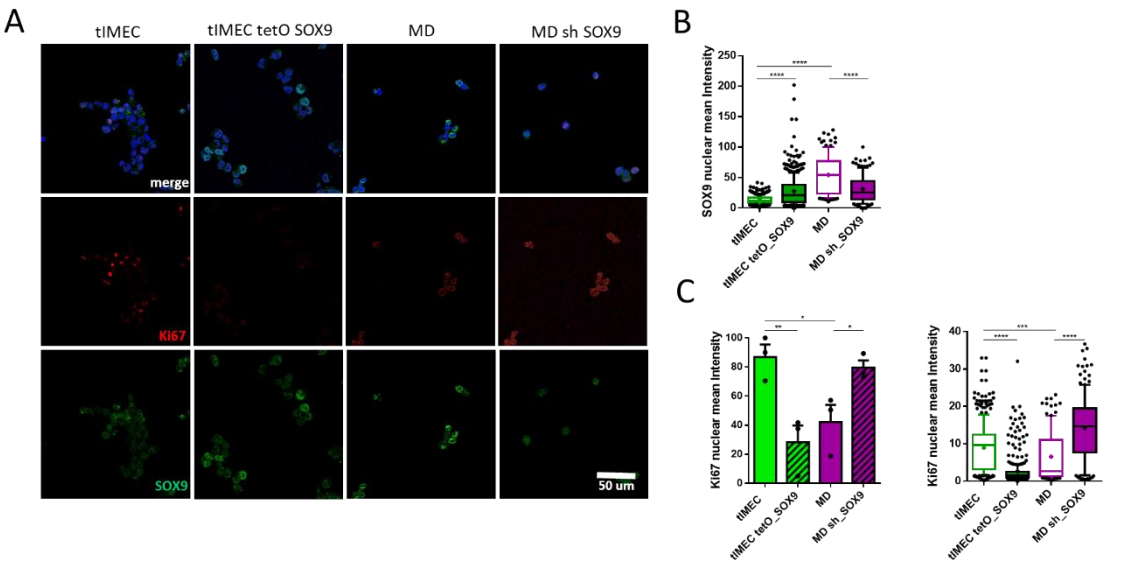


Figure 51 SOX9 is associated with tIMEC, XD and MD cell proliferation modulation. *A) Representative confocal images of DAPI (blue), SOX9 (green) and Ki67 (red) immunostaining in tIMEC, tIMEC tetO-Sox9, MD cells and MD sh-SOX9 cells. Images were acquired with a 63X objective. Scale bar = 50µm. B) Quantification of SOX9 nuclear mean intensity in the indicated cell types. N≥300 derived from the merge of 3 biological replicates; C) Quantification of percentage of KI67+ cells in the indicated cell types. Ki67 + cells were identified as mean intensity >10 nuclei. Average of 3 independent biological replicates. The box plots indicate median values (middle lines), first and third quartiles (box edges) and 10th and 90th percentiles (error bars) of the fluorescence intensity of analysed nuclei. Unpaired two-tailed student's t-test. Ns:P > 0.05; *:P ≤ 0.05; **:P ≤ 0.01;***:P ≤ 0.001;****:P ≤ 0.0001.*

In order to further corroborate the notion that ATRA treatment is sustaining quiescence in metastatic cells (**Figure 49**) and whether this effect might correlate with SOX9 levels, we analyzed the levels of the activated (phosphorylated) form of p-38 mitogen activated protein kinase (MAPK).

MAPK belong to a family of kinases that transduce various types of damage-related signals. The p38 family was firstly identified as regulators of inflammation, although it was also found to play a significant role in cell proliferation (Beamer and Corrêa, 2021; Khiem et al., 2008). In addition, activation of p38 can promote growth arrest through downregulation of cyclin D1 and the activation of p53-p21 or p16-Rb pathways (Gubern et al., 2016; Lavoie et al., 1996; Mikule et al., 2007). It was also shown that blockade of adhesion signalling resulted in the activation of p38, which favored survival and acquisition of a dormant phenotype by HEp3 (D-HEp3) cells through the initiation of a stress adaptive response known as the unfolded protein response (Ranganathan et al., 2006). Furthermore, p38α/β was shown to regulate a transcription factor network that

leads to quiescence involving proliferation/growth arrest and pluripotency/self-renewal genes (Adam et al., 2009).

p38 MAPK can be activated by phosphorylation: the activated form of p38 acts as an ERK inhibitor, preventing cell proliferation by inducing G0-G1 arrest or triggering senescence and apoptosis (Aguirre-Ghiso et al., 2003). We thus aimed at evaluating the levels of phospho-p38 in tIMEC, MD cells and respective SOX9-modulated derivatives before and after ATRA treatment (**Figure 52**A).

Quantification of the nuclear mean intensity (**Figure 52**B) showed that p-p38 levels, when comparing the steady state of the different cell populations, were slightly increasing in tIMEC after SOX9 overexpression and in MD cells with respect to tIMEC, and were instead decreased in MD cells after SOX9 knock-down to levels comparable to tIMEC. Immunostaining quantification also indicated that ATRA treatment did not have any effect on tIMEC, whereas p-p38 intensity increased in all the other three cell populations in concomitance to retinoic acid pathway activation. Nuclear mean intensity quantification also showed that treatment with RARα antagonist slightly decreased p-p38 signal intensity in all the cell lines.

Nevertheless, we noticed that, in some cases, differences between samples resulted to not be statistically significant because the mean fluorescence intensity was not changing significantly, although the distribution of the signal was. In most cases, indeed, small populations could be distinguished from the whole cell population because of a very bright nuclear p-p38 signal, and this was reflected in the distribution tails of high nuclear mean intensity present in many of the samples. Thus, we reasoned that changes in the mean fluorescence intensity may be due to difference in the background signal, and that it would be thus more accurate to analyze the percentage of p-p38 positive cells after the establishment of a threshold. Since the different cell populations were characterized by very different mean intensities, establishment of an intensity threshold on the basis of the distribution of the sample where the signal was the weakest would result in an overestimation of p-p38 positive cells in other samples. Subsequently, since we also reasoned that quiescent cells would probably represent a small percentage of the whole cell population anyway, we proceeded with analysis of p-p38 positive cells based on a threshold of mean nuclear mean intensity ≥50 (**Figure 52**C). This type of

quantification allowed us to assess that in tIMEC, approximately 3% of cells resulted to be p-p38^{+}, similarly to tIMEC overexpressing SOX9, whereas in MD cells (independently from SOX9 levels) the percentage was augmented to 5%, indicating that, at least in this experimental setting, quiescence may be induced independently from SOX9 but probably by the metastatic cellular context. Quantification of p-p38^{+} cells also showed that induction of the retinoic acid pathway did not have any significant effect in tIMEC or MD after SOX9 knock-down, whereas overexpression of SOX9 in tIMEC allowed activation of p38 similarly to MD cells. This indicated that activation of p-p38 by the retinoic acid pathway may act through SOX9. Inhibition of RARα through the treatment with its antagonist did not show any significant effect on the percentage of p-p38^{+} cells, indicating that this pathway is probably acting through other signal transducers.

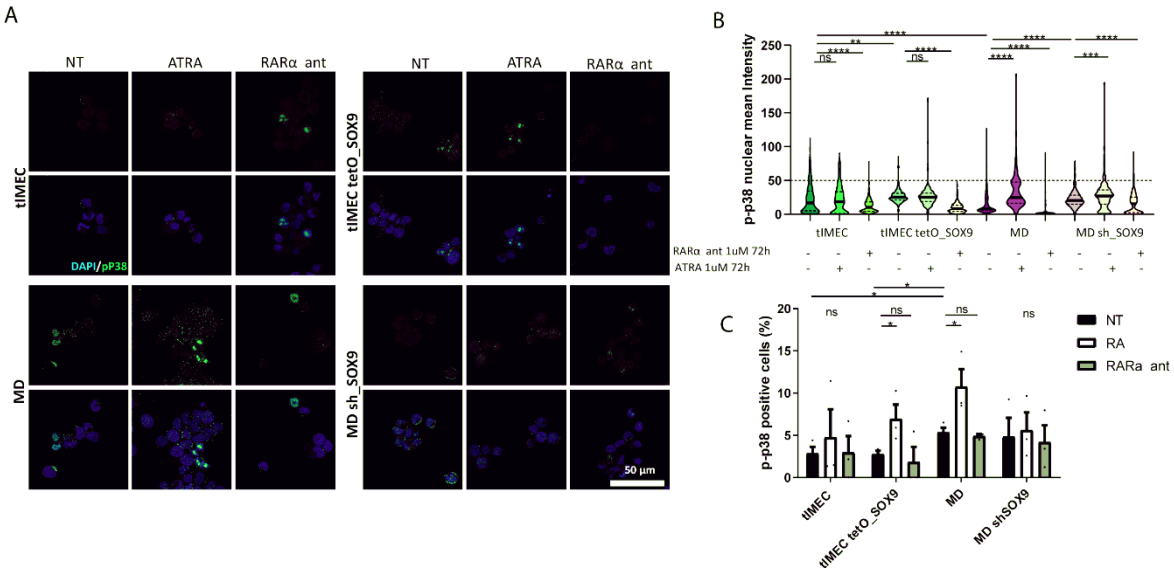


Figure 52 p-p38 immunostaining shows proliferation modulation by SOX9 and ATRA treatment in tIMEC and MD. *A) Representative confocal images of DAPI (blue), p-p38 (green) immunostaining in tIMEC, tIMEC tetO-SOX9, MD cells and MD sh-SOX9 cells with or without RARα antagonist or ATRA treatment (1 µM 72h). Images were acquired with a 63X objective. Scale bar = 50µm. B) Quantification of p-p38 nuclear mean intensity in the indicated cell types and conditions. C) Quantification of p-p38 positive cell percentage (threshold: mean intensity > 50); AVG +/- S.E.M. N≥1000, merge of 3 biological replicates. The violin plots indicate median values (middle lines), first and third quartiles (dashed lines) of the fluorescence intensity of analysed nuclei. Unpaired two-tailed student's t-test. Ns:P > 0.05; *:P ≤ 0.05; **:P ≤ 0.01;***:P ≤ 0.001;****:P ≤ 0.0001.*

These results suggested that ATRA treatment and SOX9 levels modulation are linked to proliferation regulation. However, further experiments have to be conducted to investigate how SOX9 and retinoic acid pathway activation are involved in proliferation regulation.

4.1.4.6. Retinoic acid, *SOX9* and invasion capacity

We further intended to investigate whether *SOX9* modulation in tIMEC and MD cells and whether induction of the retinoic acid pathway would affect the invasion capacity of these cell populations. To reach this goal, we performed invasion assays in tIMEC, tIMEC overexpressing *SOX9*, MD expressing shRNA targeting control (GFP) or SOX9 transcript, in untreated or ATRA treated (72 hours) cells (**Figure 53** A and C). Quantification of the spheroid area and perimeter showed that in tIMEC, spheroids showed similar shapes independently of *SOX9* levels or ATRA treatment, whereas in MD cells, knock-down of *SOX9* slightly decreased the spheroid area and perimeter, which was partially rescued by ATRA treatment (**Figure 53** B and D). Furthermore, quantification of the number of detached cells and of their migration distance from the original spheroid showed that in tIMEC, treatment with ATRA did not have any statistically significant effect, whereas *SOX9* overexpression caused an increment of both the number and the migration of detached cells. Also in MD cells, ATRA did not have any statistically significant effect on the number or migration of detached cells, while *SOX9* knock-down significantly decreased the invasion capacity of these cells.

In order to assess whether the conclusions that may be drawn from this experiment would be confirmed by inhibiting the retinoic acid receptor pathway, we also tested the effect of the treatment with the RARα antagonist BMS614 in the same cell populations (**Figure 54** A and C). Quantification of the spheroid area and perimeter shows that in tIMEC, spheroids showed similar shapes independently of *SOX9* levels or ATRA treatment, except for the combination of *SOX9* overexpression and RARα antagonist treatment, which showed a slight increase in the spheroid area when compared to the treatment alone. In MD cells, both RARα antagonist and knock-down of *SOX9* slightly decreased the spheroid area and perimeter, but we could not assess any effect of the antagonist on the cells where *SOX9* was knocked-down (**Figure 54** B and D). Quantification of the number of detached cells shows that in tIMEC, treatment with

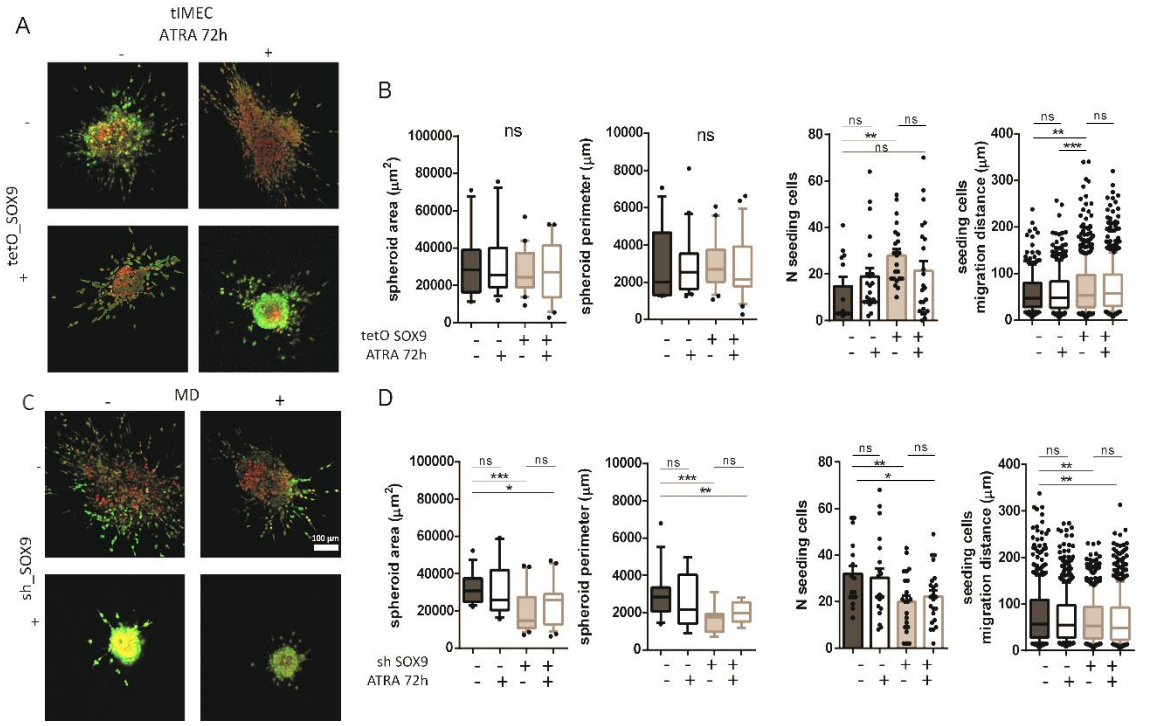


Figure 53 SOX9 and ATRA modulation of invasion capacity *A) Confocal 3D imaging of tIMEC and tIMEC tetO SOX9 spheroids treated for 72h with 1µM ATRA or vehicle and embedded in a collagen growth-factor depleted matrix. Cells are invading the spheroid-surrounding matrix to reach a chemo-attractant rich area. Representative images of 4 independent experiments. Maximum intensity projection of 50-100 2.6µm Z-stacks acquired with a 10X objective, 2x zoom on a SP8 confocal microscope. Calcein AM (green); H2B-mCherry (red) B) Quantification of spheroid area, perimeter, seeding cells number and distance (i.e. single cells detached from the originary spheroid mass and migrated towards chemoattractant) for the indicated conditions; C) Confocal 3D imaging of MD and MD shSOX9 spheroids treated for 72h with 1µM ATRA or vehicle and embedded in a collagen growth-factor depleted matrix. Cells are invading the spheroid-surrounding matrix to reach a chemo-attractant rich area. Representative images of 4 independent experiments. Maximum intensity projection of 50-100 2.6µm Z-stacks acquired with a 10X objective, 2x zoom on a SP8 confocal microscope. Calcein AM (green); H2B-mCherry (red) D) Quantification of spheroid area, perimeter, seeding cells number and distance (i.e. single cells detached from the originary spheroid mass and migrated towards chemoattractant) for the indicated conditions; merge of analysis on 24 spheroids per condition, retrieved from 4 biological replicates; AVG +/- S.E.M. student's t-test; Ns:P > 0.05; *:P ≤ 0.05; **:P ≤ 0.01;***:P ≤ 0.001;****:P ≤ 0.0001. The box plots in B) and D) indicate median values (middle lines), first and third quartiles (box edges) and 10th and 90th percentiles (error bars) of respectively area and perimeter of the analysed spheroids and distance of seeding cells from analysed spheroids.*

RARα antagonist did not have any statistically significant effect, whereas *SOX9* overexpression was increasing the number of detached cells. In addition, migration distance of detached cells was slightly increased by both *SOX9* overexpression and treatment with RARα antagonist. In MD cells, RARα antagonist did not have any statistically significant effect on the number of detached cells, but it seemed to increase their migration capacity in both steady-state and *SOX9* knock-down cells, while *SOX9* knock-down *per se* significantly decreased both the number and the migration distance of MD cells.

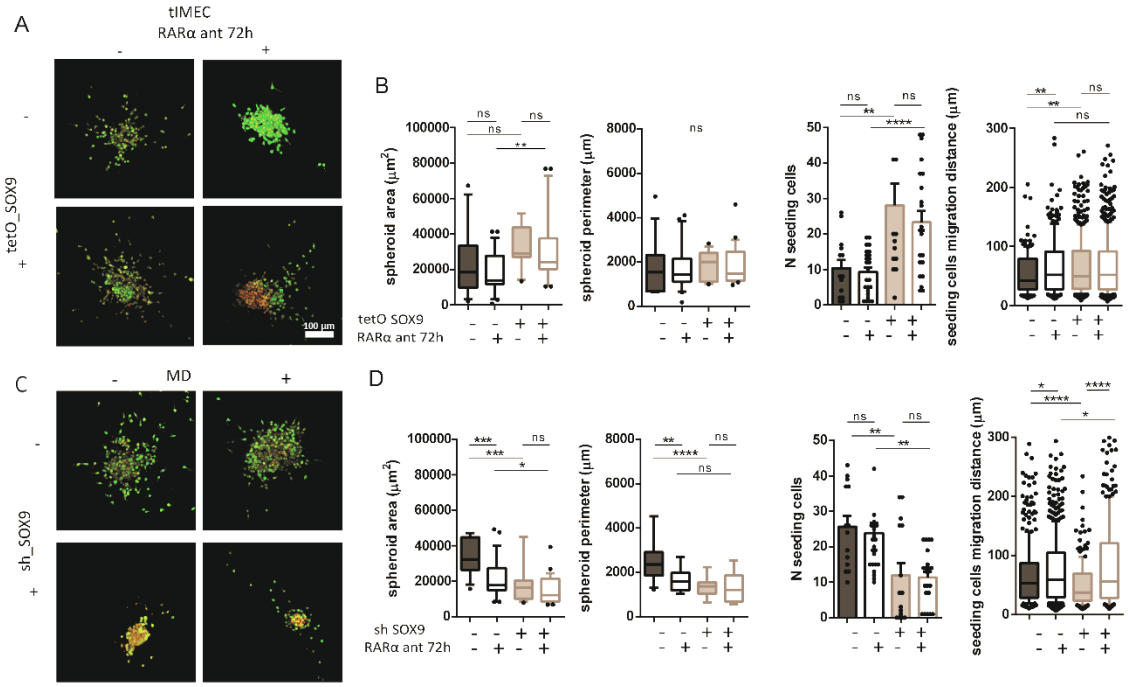


Figure 54 SOX9 and RA pathway inhibition modulation od invasion capacity *A) Confocal 3D imaging of tIMEC and tIMEC tetO SOX9 spheroids treated for 72h with 1µM RARα antagonist or vehicle and embedded in a collagen growth-factor depleted matrix. Cells are invading the spheroid-surrounding matrix to reach a chemo-attractant rich area. Representative images of 4 independent experiments. Maximum intensity projection of 50-100 2.6µm Z-stacks acquired with a 10X objective, 2x zoom on a SP8 confocal microscope. Calcein AM (green); H2B-mCherry (red) B) Quantification of spheroid area, perimeter, seeding cells number and distance (i.e. single cells detached from the originary spheroid mass and migrated towards chemoattractant) for the indicated conditions; C) Confocal 3D imaging of MD and MD shSOX9 spheroids treated for 72h with 1µM RARα antagonist or vehicle and embedded in a collagen growth-factor depleted matrix. Cells are invading the spheroid-surrounding matrix to reach a chemo-attractant rich area. Representative images of 4 independent experiments. Maximum intensity projection of 50-100 2.6µm Z-stacks acquired with a 10X objective, 2x zoom on a SP8 confocal microscope. Calcein AM (green); H2B-mCherry (red) D) Quantification of spheroid area, perimeter, seeding cells number and distance (i.e. single cells detached from the originary spheroid mass and migrated towards chemoattractant) for the indicated conditions; merge of analysis on 24 spheroids per condition, retrieved from 4 biological replicates; AVG +/- S.E.M. student's t-test; Ns:P > 0.05; *:P ≤ 0.05; **:P ≤ 0.01;***:P ≤ 0.001;****:P ≤ 0.0001. The box plots in B) and D) indicate median values (middle lines), first and third quartiles (box edges) and 10th and 90th percentiles (error bars) of respectively area and perimeter of the analysed spheroids and distance of seeding cells from analysed spheroids.*

With these results, we could conclude that invasion capacity is strongly modulated by SOX9 expression, but the retinoic acid pathway did not show any robust effect in affecting the capacity of these cells to invade a soft matrix. These results suggested that, contrary to the ATRA-induced cell quiescence, MD invasion capability is modulated by SOX9 independently form ATRA signalling.

5. Discussion

Breast cancer consists of highly heterogeneous tumours, whose cell of origin and driver oncogenes are challenging to be uniquely defined. Triple-negative breast cancer, in particular, is characterized by the most aggressive behaviour and is associated with the worst survival compared to other BC subtypes: in TNBC, 25% of patients will relapse with distant metastasis. While significant progress has been made to highlight the processes at the basis of tumour initiation, the molecular mechanisms and signalling pathways underlying the fatal late stages of metastatic dissemination is still unclear. A growing body of research has provided insights into the role of dormancy and reactivation in supporting cell plasticity and metastatic seeding. In this work, we provide data supporting the notion that changes of the chromatin landscape during tumor progression increases the responsiveness of cancer cells to environmental cues that they may encounter during dissemination and colonization of distant organs. Specifically, we showed that the augmented chromatin accessibility of MD-enhancers correlated with an increase response to ATRA signalling, which induces a slow-proliferative state, at least *in vitro*. Furthermore, this phenotypic response is mediated by the ATRA-responsive gene SOX9, whose transcription level correlates with the augmented chromatin accessibility and Promoter-Enhancer interactions retrieved in the metastatic cells.

5.1. Phenotypic and epigenetic landscape alteration in a triple-negative breast cancer model

We show that our xenograft-derived model resembles the steps of tumorigenesis, demonstrating that metastasis-derived cells are characterized by an intrinsic capability of forming spheroids when grown in a 3D setting and ultra-low attachment conditions, which better mimics the microenvironment of primary tumours with respect to 2D

culture (Huang et al., 2020). Metastatic derived cells also showed increased migration capacity with respect to both tumorigenic cells and primary tumour derived cells, demonstrating that MD cells resemble one of the typical traits that characterize metastatic cells (Koedoot et al., 2019)., When cultured in 3D systems, MD cells were able to invade their surroundings, detaching from the primary spheroid and moving as single cells to relatively long distances through soft matrices; this aggressive behaviour distinguished MD cells from tIMEC and XD cells, whose invading capability was significantly lower. This further indicated that MD cells resemble metastatic cells phenotypical traits even when cultured in vitro (Aw Yong et al., 2020). Nevertheless, *in-vivo* confirmation of MD cells dissemination capacity is required to further confirm their capability to travel to distant sites and establish micro-/macro-metastases. Also, we acknowledge that the lack of more than one representative cell population that would allow us to consider the intrinsic heterogeneity of metastasis derived cells will require confirmation of the obtained results in another model, such as the lung metastasis derived cells MDA-MB-231 (Cai et al., 2020).

Nevertheless, we were able to characterize the epigenetic landscape of the different cell types by analysing their gene expression profiling, chromatin accessibility and enhancer connectome. In the first place, ATAC-seq profiling showed that our TNBC model resembles patient tumours chromatin state, as their accessibility data cluster along with other cancer datasets when analysed through UMAP plot. Furthermore, Spearman pairwise distance indicated that MD cells show a high level of similarity with tIMEC and basal breast cancer samples (**Figure 17**).

At the same time, ATAC-seq and RNA-seq analysis show that tumorigenic cells, primary tumour derived cells and metastatic derived cells are different in terms of chromatin accessibility and gene expression. RNA-seq analysis highlights the differences between the primary tumour tissue and the primary tumour derived cells. GO analysis on the clusters that distinguish the PT sample from the other cell populations indicates that these differences can be considered as a limitation of the applied approach, as they are mainly linked to pathways related to tumour-microenvironment interactions and to metabolic processes which are known features of primary tumours, and that are not recapitulated in in-vitro cultured models. Nevertheless, MD cells display the most

distinctive accessibility and expression patterns, as shown by the differential accessibility and expression analyses. Nevertheless, integration of ATAC-seq and RNA-seq will be necessary to link the alteration of the epigenetic landscape to the expression pattern. MDs distinctive expression and chromatin accessibility pattern are also reflected by a different organization at the chromatin level, which was analysed by HiChIP. HiChIP suggested that chromatin rewiring characterizes the different tumorigenesis steps, altering physical interaction between regulatory elements and promoters. For example, *CXCR4* gains enhancer-promoter interactions specifically in MD cells, whereas no physical interactions are detected in tIMEC or XD cells. Looping anchors overlap with previously identified *CXCR4* breast-cancer specific enhancers, and this gain of enhancer-promoter interactions is in line with an increase in *CXCR4* expression, which was detectable in metastatic cells but not in tumorigenic cells (**Figure 55**) (Cai et al., 2020).

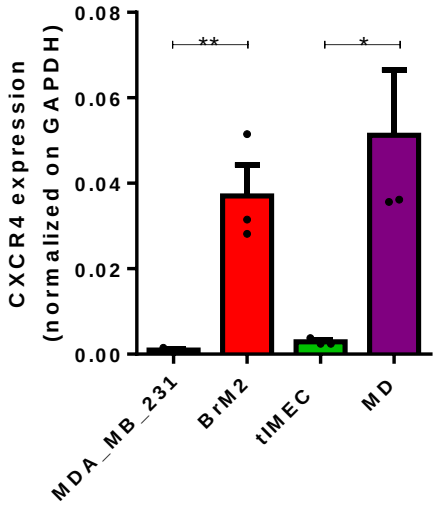


Figure 55 CXCR4 expression level in parental and metastatic MDA_MB_231, tIMEC and MD cells. *CXCR4 transcript levels in the parental MDA-MB-231 cells, Brain metastasis derived cells (BrM2), tIMEC and MD cells,, normalized on GAPDH transcript levels. 3 biological replicates, AVG +/- S.E.M. Unpaired two-tailed student's t-test. Ns:P > 0.05; *:P ≤ 0.05; **:P ≤ 0.01.*

C-Myc, a well-known benchmark oncogene of breast cancer, showed several enhancer-promoter interactions in all three cell types. Although there seemed to be common docking mechanisms between the cell types, most of the anchoring of loops was differentially placed in the different cell settings; in the metastatic cell line, long-range interactions were disrupted, and the number of 'local' interactions increased. As

recently reported, various enhancers at the *c-Myc* site ensure optimal levels of enhancer-promoter looping and mRNA expression (Schuijers et al., 2018).

MD cells unique enhancer accessibility and gene expression are also linked to the enrichment of specific transcription factor binding sites. With the IMAGE analysis, we found motifs for different TF families enriched specifically in metastatic-enriched enhancers. As expected, we found binding sites for EMT-involved transcription factors, lineage-specifying TFs and many signalling pathway responsive TFs. Among the transcription factors that emerged in the IMAGE analysis (**Figure 22**), we noticed the enrichment of retinoic-acid-responsive TFs, fostering our interest. SOX9, a transcription factor that already caught our attention because of the enrichment in its locus for features denoting activation of *de-novo* enhancers (**Figure 11**), emerged as another TF for which binding sites are enriched in MD cells enhancers. We also performed transcription factor network analysis, from which a central neighbourhood with SOX9 as central node was highlighted. Interestingly, in this analysis SOX9 results to be linked to one only causal TF activity, RXRα: this further stimulated our curiosity in investigating the retinoic acid pathway possible role in modulating SOX9 expression.

After characterization of the epigenetic landscape of our model, we can hypothesise that chromatin re-shaping may act as a significant regulator of cell fate during tumorigenesis and metastasis progression, although orthologous experiments and further investigation on the function of enhancer rewiring in metastatic cells are needed to confirm this hypothesis (Bi et al., 2020).

5.2. SOX9 enhancers interaction frequency alteration correlates with SOX9 overexpression in metastasis derived cells

In order to verify our hypothesis of enhancer de-regulation driving cell plasticity in metastasis, we focused on the locus of *SOX9* because of the reasons mentioned above:

SOX9 is a transcription factor that is overexpressed in triple-negative breast cancer, and we measured an increase of chromatin accessibility at distal sites in metastasis derived cells with respect to tumorigenic cells and primary tumour derived cells. Moreover, due to this absence of potential indirect effects with no surrounding genes being expressed, we considered *SOX9* as a suitable model locus to dissect the contributions of individual enhancers to target gene expression (Christin et al., 2020; Jana et al., 2020; Ma et al., 2020).

Firstly, we show that *SOX9* is expressed in MD cells at a higher level with respect to XD cells and tIMEC, therefore we assumed that changes in its expression might be linked to putative distal enhancers' measured differential chromatin accessibility. However, to confirm enhancer hyperactivation, assessment of histone modification enrichment concomitant to increased accessibility is essential: ChIP-seq analysis for detecting histone modification that identifies active enhancer regions (H3K4me1 and H3K27ac) is needed. Furthermore, the deletion or epigenetic repression through CRISPRi of SOX9 regulatory elements is necessary to confirm the causal link between enhancer hyperactivation and *SOX9* overexpression (Choi et al., 2021; Li et al., 2020).

Nevertheless, we are confident that regulatory elements identified through ATAC-seq are indeed acting on the *SOX9* promoter, as we were able to detect physical interaction between most of the putative enhancers and *SOX9*. Furthermore, we were able to assess that highly accessible chromatin regions proximal to *SOX9* promoter are interacting with a higher frequency with the promoter in metastatic cells with respect to XD cells and tIMEC. Of note, the augmented number of interactions observed might be due to higher H3K27ac deposition and not increased chromatin looping *per se*. Further experiments including DNA FISH or chromatin conformation capture techniques (3C or 4C) will be helpful to further confirm increased enhancer-promoter interaction and 3D genome re-organization during tumorigenesis (Dixon et al., 2015; Millau et al., 2016).

When analysing publicly available data of HMEC chromatin compartments, *SOX9* promoter is clearly positioned just at the edge between two TADs, but as indicated by HiChIP long-range contacts, we assume that regulatory elements that we identified reside in the same TAD as *SOX9* promoter (3D genome browser link). However, in order

to assess the 3D genome organization of the *SOX9* locus in the analysed cell types, we would need to perform Hi-C. The intersection of ATAC-seq peaks and HiChIP looping bins allowed us to identify those putative enhancers, out of the 20 regions characterized by increased accessibility we detected, that significantly interact with the *SOX9* promoter in MD cells. We analysed HiChIP data both at 10kb and 25kb resolution: we noticed a lower number of enhancer regions interacting with the *SOX9* promoter at higher resolution, and we were not able to detect interactions between *SOX9* promoter and the most proximal putative enhancer regions, indicating that sequencing depth is probably a limit for detecting weaker but still significant interactions at high resolution (Mumbach et al., 2016, 2017; Petrovic et al., 2019).

However, we were able to confirm the enhancer activity capacity of identified putative regulatory elements through luciferase assay and established an enhancer reporter system to visualise enhancer activation in specific tumorigenesis steps in vitro and in vivo. The results obtained in this experimental setting suggested that single enhancer elements are not sufficient to induce a robust expression of the reporter expression, indicating the possibility of the need for multiple regulatory regions to cooperate with each other to sustain enhancer activity, as reported in (Thomas et al., 2021). Enhancer super-additivity and redundancy were recently described as one of the mechanisms by which enhancers modulate gene expression in the cancer context (Bothma et al., 2015; Choi et al., 2021; Dukler et al., 2017; Thomas et al., 2021). Therefore, it will be necessary to test the activity of multiple regulatory elements combinations in the reporter system to evaluate the level of additivity and/or cooperation among these enhancers. After analysing IMAGE results, we hypothesised that induction of the single regulatory element activity could be achieved by inducing a specific signalling pathway, that through the recruitment of transcription factors to the enhancers would induce chromatin rewiring and increase the promoter activity. Again, chromosome conformation capture techniques combined with pathway activation are needed to confirm this hypothesis.

We tested the potential effect on *SOX9* expression by modulating different signalling pathways based on the transcription factors families that emerged from the IMAGE analysis, and we detected the most straightforward response by the induction of the

retinoic acid pathway. This was in line with GO results (**Figure 16**) that highlighted retinol-metabolism pathways to be enriched in MD cells. We thus proceeded with focusing on retinoic acid receptors and their activity. We confirmed that, as depicted in **Figure 56**, retinoic acid receptor binding sites are not exclusive of *SOX9* enhancers but also enriched in many other MD-specific enhancers, such as in *CXCR4* promoter-interacting enhancers. Although ChIP-seq analysis or orthologous experiments are necessary to prove RARs binding to *SOX9* enhancers, we hypothesised that the retinoic acid pathway might play a role in MD enhancer activity, thus in *SOX9* enhancer hyperactivation and in favouring the metastatic process.

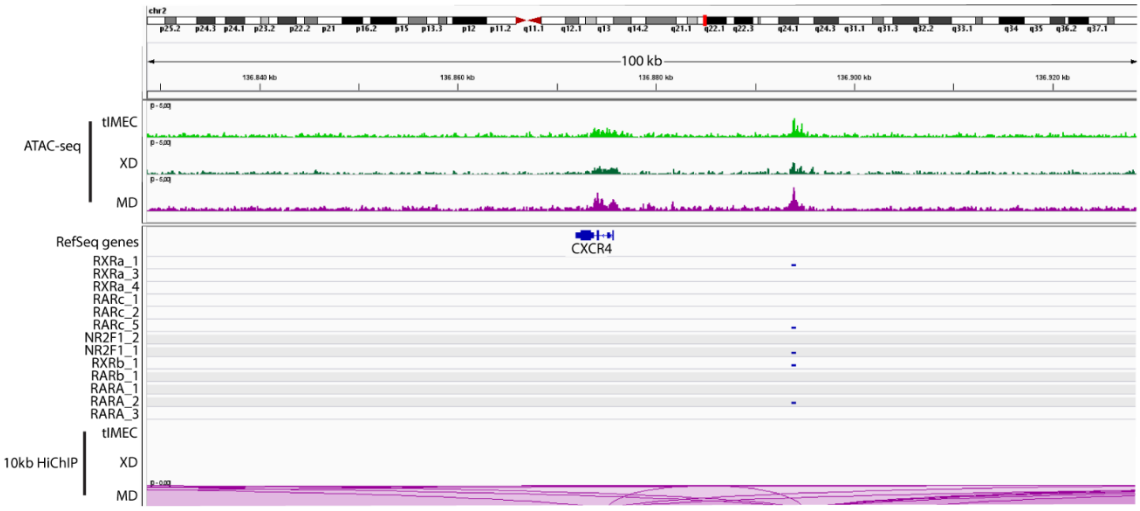


Figure 56 RARs binding sites in CXCR4 metastatic enhancers. *Integrative Genome Viewer view of the CXCR4 locus. From top: ATAC-seq tracks of tIMEC, XD cells and MD cell (data range =0-5); RefSeq gene annotation; RXRα, RARγ, NR2F1, RXRβ, RARβ, RARα binding sites retrieved from Kittler et al., Cell Rep 2013; HiChIP total significant (FDR>0.05) loops in tIMEC, XD cells, MD cells (resolution = 10kb).*

Since RARs binding sites are redundant in *SOX9* putative enhancers, we hypothesised that retinoic acid response might act through transcriptional memory. It is now apparent that epigenetic mechanisms provide this "priming" of the genome to respond more quickly to subsequent exposure to the signal (D'Urso and Brickner, 2014; Kerr and Corbett, 2010). We show indeed that SOX9 expression is induced by retinoic acid treatment, and a pre-treatment with ATRA leads to a more robust increase in mRNA expression, specifically in MD cells. We also show that SOX9 expression response to ATRA is dependent on the retinoic acid receptor RARα, since knock-down of this factor

abrogates the capacity of MD cells to respond to the retinoic acid pathway activation, and treatment with RARα antagonist leads to a strong decrease in *SOX9* expression, although this effect is not specific of MD cells but also detectable in XD cells. This may indicate that although activation of the retinoic acid pathway does not lead to a significant effect in terms of *SOX9* expression modulation in XD cells, abrogation of RARα activity is affecting the steady state maintenance of *SOX9* expression in these cells as in MD cells, suggesting that the levels of *SOX9* are already affected by the retinoic acid pathway activation in steady state conditions.

One might speculate that the short ATRA treatment drives RARs binding to *SOX9* enhancers, thus enabling a transient enhancer-promoter looping which translates in a slight response in terms of SOX9 expression. On the other hand, short treatment on ATRA pre-exposed cells is allowing steady and robust overexpression, possibly as a result of the rapid assembly of transcriptional condensates and stable bridging by TFs and co-activators between the regulatory element and the promoter, as a consequence of RARs already being bound to the regulatory element after the pre-exposure to the compound. Alternatively, retinoic acid pathway activation following pre-exposure may lead to the removal of repressive histone marks, thus allowing for faster H3K27 acetylation in response to the short treatment as suggested by previous work (Iberg-Badeaux et al., 2017; Tu et al., 2017). Further analyses would be required to distinguish between these two mechanisms. Nevertheless, the possibility that a priming event may facilitate the response of metastatic cells to a second cues is of potential interest as it may help in understanding the mechanisms governing the increment of cellular fitness of metastatic cells, and their capability to adapt to foreign environments.

5.3. Retinoic Acid enhances cell plasticity inducing quiescence in metastatic cells

We evaluated that ATRA receptors expression is not changing significantly in response to retinoic acid pathway levels. Indeed, RARα, RXRα, RARγ, and RXRβ nuclear protein levels are not significantly changing in MD, tIMEC or XD cells after ATRA

treatment, indicating that *SOX9* expression modulation has to be mediated by another mechanism. Also, we found RARα to be localized more in the cytosol than in the nucleus in most conditions, so we tested antibody specificity through immunostaining after knock-down, confirming the selectivity of the antibody and thus corroborating the notion indicating that RARα is subjected to shuttling from the cytosol to the nucleus, depending on the cell type and treatment (Han et al., 2009; Maruvada et al., 2003; Park et al., 2010). In order to assess the nuclear localization of RARα, we tested shorter ATRA treatment time points in combination with ATRA priming, and this indeed demonstrated RARα increased nuclear translocation in the three cell types in response to the treatment.

We hypothesise that ATRA receptors might cooperate with each other (heterodimerization) or with transcriptional cofactors (cofactor oligomerization) (**Figure 5**) in binding *SOX9* enhancers and acting as a 'bridge' to the *SOX9* promoter, facilitating chromatin looping and thus mediating the transcriptional response to ATRA. Although RXRβ and RARβ do not show clustering in cell nuclei, we show that RXRα and RARα do, and the number of clusters detected per nuclei is augmented after ATRA treatment. The involvement of retinoic acid receptors in activating *SOX9* enhancers and thus driving *SOX9* expression augmentation needs further investigation, focusing specifically on elucidating the clustering capacity and dynamics of the receptors and eventually their colocalization with *SOX9* enhancers. In this direction, optogenetics approaches might be crucial, and in situ hybridization approaches combined with immunostaining may be necessary to elucidate how retinoic acid receptors act to modulate *SOX9* expression (Fasciani et al., 2020; Fe Lanfranco et al., 2017).

However, we prove that following retinoic acid treatment, all the three cell types show a decrease of the Ki67 proliferation marker, which is most marked in MD cells, indicating a connection between ATRA and quiescence induction, in line with recent evidence (Khalil et al., 2021; Sosa et al., 2015). We also proved ATRA induction of quiescence through an orthologous approach based on dye retention assay, and the obtained results clearly show that a population of non-proliferating cells is more prominent in MD cells already in the steady state when compared to the other cell types, but even more significantly following ATRA treatment combined with the priming. This,

other than confirming retinoic acid pathway involvement in proliferation inhibition in this model, can also be interpreted as a confirmation of transcriptional memory being involved in the response to ATRA (Bevington et al., 2016; Woodworth and Holloway, 2017).

To link the decrease in proliferation levels to *SOX9* augmented expression, we show that *SOX9* levels anti-correlate with Ki67, suggesting a connection between *SOX9* expression level and proliferation modulation. Furthermore, we assessed that MD cells, independently of SOX9 levels, are also characterized by a more prominent p-p38^{+} population when compared to tIMEC and tIMEC overexpressing SOX9, indicating that the presence of a quiescent population may be a characteristic of metastatic cells which is not affected by *SOX9* expression. The quiescent population fraction is although enriched in response to ATRA specifically in MD cells and in tIMEC overexpressing SOX9, which leads us to conclude that retinoic acid pathway is inducing quiescence through SOX9, even though this mechanism has to be validated and further investigated. Nevertheless, we also assessed that inhibition of RARα does not show any significant effect on the percentage of p-p38^{+} cells, suggesting that other signal transducers may be involved in this response. In order to have a clearer and more precise description of the mechanism by which *SOX9* expression modulation is linked to ATRA response in terms of quiescence we will need to further investigate this, especially through *in vivo* validation.

Nevertheless, we were also interested in investigating SOX9 and retinoic acid pathway modulation effect on invasion capacity. Invasion assays on tIMEC and MD cells with *SOX9* levels modulation and following activation or inhibition of the retinoic acid pathway showed that *SOX9* expression correlates with the capacity of these cell populations to invade and migrate through soft matrices; those cell populations that express the highest levels of SOX9 manifest the highest capability of invading and migrating through collagen. Alteration of the retinoic acid pathway, however, did not allow us to drive any conclusion on the possible effect it may have on invasion capability, since we could not measure any major changes upon ATRA or RARα antagonist treatments. This may suggest that *SOX9* is independent from retinoic acid in stimulating cells to invade and migrate through the surroundings, and that SOX9 enhanced

expression following ATRA treatment may be counterbalanced by other processes happening in this context.

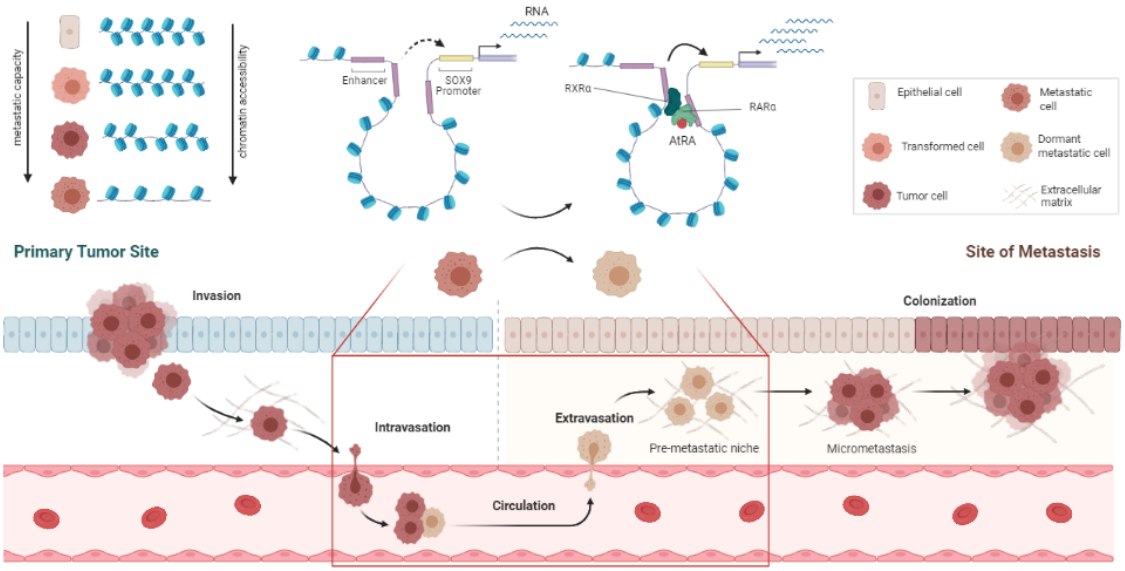
On the basis of the obtained results, we come to the conclusion that SOX9 provides a proof of concept of epigenetic alteration driving transcription modulation: SOX9 regulatory elements showed increased accessibility and interaction frequency with the promoter in MD cells, possibly supporting SOX9 expression. We showed that the increased SOX9 expression strengthened MD-specific migration and invasive capability with respect to XD and tIMEC. We further showed that MD cells response to ATRA in terms of quiescence induction was dependent on SOX9 levels, and that tIMEC were rendered responsive to ATRA after SOX9 overexpression, which indicates that in this context SOX9 expression is presumably the main factor that cooperates with ATRA signalling in inducing quiescence.

Disseminated tumour cells face multiple and diverse environmental stresses which may result in the activation of a state of dormancy, a phrase coined by Willis in the context of tumour progression as early as 1934 (Willis, 1934). Dormancy gives DTCs several advantages: prolonged survival in a foreign microenvironment, reversible growth arrest at these sites, and resistance to targeted and cytotoxic treatment (Ghajar, 2015). Multiple factors have been described to influence tumour cell dormancy: induction of dormancy results in many cases from the engagement of cells by the metastatic niche, from the interaction between cancer cells and stroma, or even from cancer treatment (Correia et al., 2021; Khoo et al., 2019; Phan and Croucher, 2020). The escape from immune surveillance has also been linked to dormancy: single dormant DTCs can survive in distant sites until they have the opportunity to re-awaken, getting away from immune clearance. This process is often rendered possible by the absence on dormant DTCs of surface markers such as the major histocompatibility complex (MHC-1) and cytokeratins, and by the decreased expression of Natural Killer (NK) cells activatory receptors or activating ligands (Arnaud et al., 2018; Correia et al., 2021).

We hypothesise that, in our model, *SOX9* might act as a “switch” between a proliferative and a quiescent state. *SOX9* is slightly overexpressed in the metastasis-derived cells with respect to the primary tumour xenograft cells, but this three-fold overexpression does not impede MD cells from proliferating. However, when the

retinoic acid pathway is activated, *SOX9* expression is further increased, concomitantly with a diminished cell division proficiency, resulting in an increment of slow-proliferating cells. We thus hypothesise that cells detaching from the primary breast tumour and engaging in the dissemination process get at some point exposed to retinoic acid: this induces an increase in *SOX9* expression and proliferation inhibition. Dormant, invasive *SOX9*^{high} cells have a higher probability of surviving the dissemination and seeding process and can thus establish in a metastatic niche; once there, following specific re-activation stimuli, cells can exit from dormancy, lower *SOX9* levels, and establish micro- and macro-metastases.

As depicted in **Figure 57**, we speculate that the augmented chromatin accessibility in metastatic cells is enabling enhancer activity to induce a slight *SOX9* overexpression in MD cells with respect to tIMEC, but given the heterogeneity of these cell populations, we hypothesise that the frequency of the interactions between *SOX9* promoter and its enhancers is relatively low. Activating the retinoic acid pathway might allow these interactions to occur more frequently, thus driving a robust response in terms of *SOX9* expression and increasing cell plasticity, translating into cell cycle progression inhibition and activation of the quiescence program on the one hand, and stimulating invasive capacity on the other.



Fi
a
th
a
reflecting cancer cells heterogeneity and plasticity. When the retinoic acid pathway is induced, retinoic acid receptors bind to SOX9 enhancers, and facilitate chromatin looping allowing for a stable interaction between enhancers and SOX9 promoter. This translates in a robust response, that translates in a strong SOX9 overexpression, and that involves inhibition of cell cycle progression, acquisition of stem cell traits, and thus enhanced metastatic capacity. Illustration created with BioRender.

As previously described (**Figure 2**), vitamin A is absorbed by cells mainly as *all-trans*-retinol (atROL) bound to RBPs in the plasma through the transporter STRA6. atROL oxidation to atRAL by RDH is the limiting step in the production of ATRA. In our model, RNAseq data indicates that both STRA6 and RDH5 are overexpressed in MD cells with respect to XD and tIMEC, thus suggesting that accumulation of ATRA may be due to enhanced import and oxidation of ATRA substrate (**Figure 58**). Furthermore, the CYP26B1 enzyme, responsible for ATRA degradation, shows a significantly lower RNA-seq signal in MD cells with respect to XD and tIMEC. This leads us to hypothesise that MD cells may accumulate ATRA both because of enhanced import and substrate availability and a lower level of degradation; this may induce an equilibrium shift towards a hyperactivation of the retinoic acid pathway.

We speculate that retinol exposure may occur during the dissemination or establishment of a pre-metastatic niche, especially considering that our metastatic model derives from metastasis to the liver, the site of retinol storage. This may facilitate those cells endowed with the metastatic potential to accumulate ATRA due to STRA6

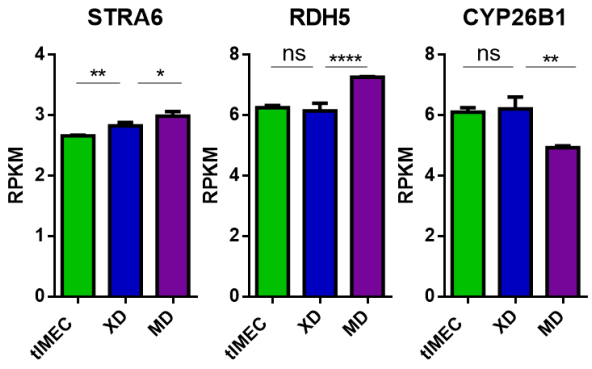


Figure 58 Gene expression of factors involved in the retinoic acid metabolism in tIMEC, XD and MD cells. *RNA-seq data (RPKM) of tIMEC, XD and MD cells for retinoic acid transporter STRA6, retinol dehydrogenase 5 (RDH5) and cytochrome P450 26B1 (CYP26B1). 3 biological replicates, AVG +/- S.E.M. Unpaired two-tailed student's t-test. Ns:P > 0.05; *:P ≤ 0.05; **:P ≤ 0.01; ***:P ≤ 0.005.*

and RDH overexpression and CYP26B1 low levels. ATRA accumulation would activate retinoic acid receptors, which through binding to regulatory elements thus would shift *SOX9* expression equilibrium towards strong enrichment for its expression, which in turn contributes to the activation of the quiescence program, thus facilitating the metastatic cascade through escaping surveillance of the immune system. Further investigation of

the mechanism by which ATRA-induced quiescence would advantage MD cells in immune surveillance escaping will be needed: testing the effects of co-culture of NK cells with the SOX9^{high}/p-P38^{high} fraction in comparison with the proliferative fraction of MD cell population may supply us with insights on the matter. Furthermore, characterization of NK cells activatory and inhibitory receptors and ligands in quiescent and proliferative sub-populations in MD cells will be needed in order to have indications on the mechanism by which metastatic cells would escape NK cytotoxicity. Finally, *in-vivo* experiments will be needed to confirm MD cells quiescence program activation in concomitance to *SOX9* overexpression and to demonstrate the advantage that this program would provide in the dissemination process.

In order to describe the mechanism by which ATRA signalling involves *SOX9* in the activation of the quiescence program, further investigation will be needed: *i*n-situ hybridization techniques and chromatin conformation capture experiments will be helpful to clarify *SOX9* enhancers remodelling in response to ATRA.

6. Conclusion and future perspectives

In this study, we provide data supporting the notion that changes of the chromatin landscape during tumour progression in TNBC increases the responsiveness of cancer cells to environmental cues that they may encounter during dissemination and colonization of distant organs. Specifically, we showed that *SOX9* serves as a proof of concept that augmented chromatin accessibility and interaction of regulatory elements correlates with transcription modulation, which results in an increased response to ATRA signalling, inducing a slow-proliferative state.

These findings reinforce the role of epigenetic alteration in modulating tumorigenicity and metastatic potential. Furthermore, they suggest SOX9 and the retinoic acid pathway as major players in the establishment of a quiescent phenotype, at least in our experimental setting.

Although several points in this work are still preliminary and will need to be confirmed by further investigation, the results described in this thesis potentially open new mechanistic and therapeutic avenues in TNBC. Mechanistically, it will be essential to explore the exact process by which retinoic acid triggers enhancer-promoter interaction, thus driving *SOX9* overexpression and inhibiting cell cycle progression, and to investigate whether other players are involved in the response to ATRA pathway activation. Therapeutically, if dormancy activation in response to retinoic acid receptors binding to SOX9 enhancers will be confirmed, the targeting of RARs would appear as a possibility to impede DTCs to enter this protective quiescence program.

7. References

Adam, R.C., and Fuchs, E. (2016). The Yin and Yang of Chromatin Dynamics In Stem Cell Fate Selection. Trends Genet. *32*, 89–100.

Adam, A.P., George, A., Schewe, D., Bragado, P., Iglesias, B. V, Ranganathan, A.C., Kourtidis, A., Conklin, D.S., and Aguirre-Ghiso, J.A. (2009). Computational Identification of a p38^{SAPK}-Regulated Transcription Factor Network Required for Tumor Cell Quiescence. Cancer Res. *69*, 5664 LP – 5672.

Adam, R.C., Yang, H., Rockowitz, S., Larsen, S.B., Nikolova, M., Oristian, D.S., Polak, L., Kadaja, M., Asare, A., Zheng, D., et al. (2015). Stem Cell Plasticity and Lineage Choice. Nature *521*, 366–370.

Afonja, O., Raaka, B.M., Huang, A., Das, S., Zhao, X., Helmer, E., Juste, D., and Samuels, H.H. (2002). RAR agonists stimulate SOX9 gene expression in breast cancer cell lines: Evidence for a role in retinoid-mediated growth inhibition. Oncogene *21*, 7850–7860.

Agostinetto, E., Eiger, D., Punie, K., and de Azambuja, E. (2021). Emerging Therapeutics for Patients with Triple-Negative Breast Cancer. Curr. Oncol. Rep. *23*, 57.

Agrawal, P., Blinka, S., Pulakanti, K., Reimer, M.H., and Rao, S. (2020). Defining a critical enhancer near Nanog using chromatin-focused approaches identifies RNA Pol II recruitment as required for expression. BioRxiv 2020.05.27.118612.

Aguirre-Ghiso, J.A. (2007). Models, mechanisms and clinical evidence for cancer dormancy. Nat. Rev. Cancer *7*, 834–846.

Aguirre-Ghiso, J.A., Estrada, Y., Liu, D., and Ossowski, L. (2003). ERK(MAPK) activity as a determinant of tumor growth and dormancy; regulation by p38(SAPK). Cancer Res. *63*, 1684–1695.

Aguirre-Ghiso, J.A., Bragado, P., and Sosa, M.S. (2013). Metastasis awakening: Targeting dormant cancer. Nat. Med. *19*, 276–277.

Ahmadiyeh, N., Pomerantz, M.M., Grisanzio, C., Herman, P., Jia, L., Almendro, V., He, H.H., Brown, M., Liu, X.S., Davis, M., et al. (2010). 8q24 prostate, breast, and colon cancer risk loci show tissue-specific long-range interaction with MYC. Proc. Natl. Acad. Sci. U. S. A. *107*, 9742–9746.

Al-Hajj, M., Wicha, M.S., Benito-Hernandez, A., Morrison, S.J., and Clarke, M.F. (2003). Prospective identification of tumorigenic breast cancer cells. Proc. Natl. Acad. Sci. U. S. A. *100*, 3983–3988.

Al-Mehdi, A.B., Tozawa, K., Fisher, A.B., Shientag, L., Lee, A., and Muschel, R.J. (2000). Intravascular origin of metastasis from the proliferation of endothelium-attached tumor cells: A new model for metastasis. Nat. Med. *6*, 100–102.

Aldaz, P., Otaegi-Ugartemendia, M., Saenz-Antoñanzas, A., Garcia-Puga, M., Moreno-Valladares, M., Flores, J.M., Gerovska, D., Arauzo-Bravo, M.J., Samprón, N., Matheu, A., et al. (2020). SOX9 promotes tumor progression through the axis BMI1-p21CIP. Sci. Rep. *10*, 357.

Alexander, J.M., Guan, J., Li, B., Maliskova, L., Song, M., Shen, Y., Huang, B., Lomvardas, S., and Weiner, O.D. (2019). Live-cell imaging reveals enhancer-dependent \textit{Sox2} transcription in the absence of enhancer proximity. Elife *8*, e41769.

Allshire, R.C., and Madhani, H.D. (2018). Ten principles of heterochromatin formation and function. Nat. Rev. Mol. Cell Biol. *19*, 229–244.

Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Chen, Y., Zhao, X., Schmidl, C., Suzuki, T., et al. (2014). An atlas of active enhancers across human cell types and tissues. Nature *507*, 455–461.

Arnaud, P., Naishitha, A., Nicoletta, M., Larkin, K.Z., Alizée, G., Ran, Y., Cédric, A., Jean, A., Mikala, E., A., I.-D.C., et al. (2018). Unresolved endoplasmic reticulum stress engenders immune-resistant, latent pancreatic cancer metastases. Science (80-.). *360*, eaao4908.

Atchley, D.P., Albarracin, C.T., Lopez, A., Valero, V., Amos, C.I., Gonzalez-Angulo, A.M., Hortobagyi, G.N., and Arun, B.K. (2008). Clinical and pathologic characteristics of patients with BRCA-positive and BRCA-negative breast cancer. J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol. *26*, 4282–4288.

Aw Yong, K.M., Ulintz, P.J., Caceres, S., Cheng, X., Bao, L., Wu, Z., Jiagge, E.M., and Merajver, S.D. (2020). Heterogeneity at the invasion front of triple negative breast cancer cells. Sci. Rep. *10*, 5781.

Bae, S., and Lesch, B.J. (2020). H3K4me1 Distribution Predicts Transcription State and Poising at Promoters. Front. Cell Dev. Biol. *8*, 289.

Balbín, M., Fueyo, A., Tester, A.M., Pendás, A.M., Pitiot, A.S., Astudillo, A., Overall, C.M., Shapiro, S.D., and López-Otín, C. (2003). Loss of collagenase-2 confers increased skin tumor susceptibility to male mice. Nat. Genet. *35*, 252–257.

Banani, S.F., Lee, H.O., Hyman, A.A., and Rosen, M.K. (2017). Biomolecular condensates: organizers of cellular biochemistry. Nat. Rev. Mol. Cell Biol. *18*, 285–298.

Batten, M.L., Imanishi, Y., Maeda, T., Tu, D.C., Moise, A.R., Bronson, D., Possin, D., Van Gelder, R.N., Baehr, W., and Palczewski, K. (2004). Lecithin-retinol acyltransferase is essential for accumulation of all-trans-retinyl esters in the eye and in the liver. J. Biol. Chem. *279*, 10422–10432.

Beamer, E., and Corrêa, S.A.L. (2021). The p38MAPK-MK2 Signaling Axis as a Critical Link Between Inflammation and Synaptic Transmission. Front. Cell Dev. Biol. *9*, 82.

Belorusova, A., Osz, J., Petoukhov, M. V, Peluso-Iltis, C., Kieffer, B., Svergun, D.I., and Rochel, N. (2016). Solution Behavior of the Intrinsically Disordered N-Terminal Domain of Retinoid X Receptor α in the Context of the Full-Length Protein. Biochemistry *55*, 1741–1748.

Bevington, S.L., Cauchy, P., Piper, J., Bertrand, E., Lalli, N., Jarvis, R.C., Gilding, L.N., Ott, S., Bonifer, C., and Cockerill, P.N. (2016). Inducible chromatin priming is associated with the establishment of immunological memory in T cells. EMBO J. *35*, 515–535.

Bhattacharyya, S., Chandra, V., Vijayanand, P., and Ay, F. (2019). Identification of significant chromatin contacts from HiChIP data by FitHiChIP. Nat. Commun. *10*, 4221.

Bi, M., Zhang, Z., Jiang, Y.Z., Xue, P., Wang, H., Lai, Z., Fu, X., De Angelis, C., Gong, Y., Gao, Z., et al. (2020). Enhancer reprogramming driven by high-order assemblies of transcription factors promotes phenotypic plasticity and breast cancer endocrine resistance. Nat. Cell Biol. *22*, 701–715.

Bleu, M., Mermet-Meillon, F., Apfel, V., Barys, L., Holzer, L., Bachmann Salvy, M., Lopes, R., Amorim Monteiro Barbosa, I., Delmas, C., Hinniger, A., et al. (2021). PAX8 and MECOM are interaction partners driving ovarian cancer. Nat. Commun. *12*, 2442.

Boija, A., Klein, I.A., Sabari, B.R., Dall'Agnese, A., Coffey, E.L., Zamudio, A. V, Li, C.H., Shrinivas, K., Manteiga, J.C., Hannett, N.M., et al. (2018). Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. Cell *175*, 1842-1855.e16.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120.

Bolstad, B. (2019). preprocessCore: A collection of pre-processing functions.

Borland, M.G., Khozoie, C., Albrecht, P.P., Zhu, B., Lee, C., Lahoti, T.S., Gonzalez, F.J., and Peters, J.M. (2011). Stable over-expression of PPARβ/δ and PPARγ to examine receptor signaling in human HaCaT keratinocytes. Cell. Signal. *23*, 2039–2050.

Bothma, J.P., Garcia, H.G., Ng, S., Perry, M.W., Gregor, T., and Levine, M. (2015). Enhancer additivity and non-additivity are determined by enhancer strength in the Drosophila embryo. Elife *4*, e07956.

Bourguet, W., Vivat, V., Wurtz, J.M., Chambon, P., Gronemeyer, H., and Moras, D. (2000). Crystal structure of a heterodimeric complex of RAR and RXR ligand-binding domains. Mol. Cell *5*, 289–298.

Bradner, J.E., Hnisz, D., and Young, R.A. (2017). Transcriptional Addiction in Cancer. Cell *168*, 629–643.

Braun, K.W., Tribley, W.A., Griswold, M.D., and Kim, K.H. (2000). Follicle-stimulating Hormone Inhibits All-trans-retinoic Acid-induced Retinoic Acid Receptor α Nuclear Localization and Transcriptional Activation in Mouse Sertoli Cell Lines*. J. Biol. Chem. *275*, 4145–4151.

Brickner, D.G., Randise-Hinchliff, C., Lebrun Corbin, M., Liang, J.M., Kim, S., Sump, B., D'Urso, A., Kim, S.H., Satomura, A., Schmit, H., et al. (2019). The Role of Transcription Factors and Nuclear Pore Proteins in Controlling the Spatial Organization of the Yeast Genome. Dev. Cell *49*, 936-947.e4.

Buenrostro, J.D., Wu, B., Chang, H.Y., and Greenleaf, W.J. (2015). ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. Curr. Protoc. Mol. Biol. *109*, 21.29.1-21.29.9.

Burton, A., and Torres-Padilla, M.-E. (2014). Chromatin dynamics in the regulation of cell fate allocation during early embryogenesis. Nat. Rev. Mol. Cell Biol. *15*, 723–734.

Cai, W.L., Greer, C.B., Chen, J.F., Arnal-Estapé, A., Cao, J., Yan, Q., and Nguyen, D.X. (2020). Specific chromatin landscapes and transcription factors couple breast cancer subtype with metastatic relapse to lung or brain. BMC Med. Genomics *13*, 33.

Calo, E., and Wysocka, J. (2013). Modification of Enhancer Chromatin: What, How, and Why? Mol. Cell *49*, 825–837.

Cameron, M.D., Schmidt, E.E., Kerkvliet, N., Nadkarni, K. V, Morris, V.L., Groom, A.C., Chambers, A.F., and MacDonald, I.C. (2000). Temporal progression of metastasis in lung: cell survival, dormancy, and location dependence of metastatic inefficiency. Cancer Res. *60*, 2541–2546.

Campbell, P.J., Getz, G., Korbel, J.O., Stuart, J.M., Jennings, J.L., Stein, L.D., Perry, M.D., Nahal-Bose, H.K., Ouellette, B.F.F., Li, C.H., et al. (2020). Pan-cancer analysis of whole genomes. Nature *578*, 82–93.

Carey, L.A., Dees, E.C., Sawyer, L., Gatti, L., Moore, D.T., Collichio, F., Ollila, D.W., Sartor, C.I., Graham, M.L., and Perou, C.M. (2007). The triple negative paradox: primary tumor chemosensitivity of breast cancer subtypes. Clin. Cancer Res. an Off. J. Am. Assoc. Cancer Res. *13*, 2329–2334.

Carter, D., Chakalova, L., Osborne, C.S., Dai, Y., and Fraser, P. (2002). Long-range chromatin regulatory interactions in vivo. Nat. Genet. *32*, 623–626.

Celià-Terrassa, T., and Kang, Y. (2016). Distinctive properties of metastasis-initiating cells. Genes Dev. *30*, 892–908.

Celià-Terrassa, T., and Kang, Y. (2018). Metastatic niche functions and therapeutic opportunities. Nat. Cell Biol. *20*, 868–877.

Centore, R.C., Sandoval, G.J., Soares, L.M.M., Kadoch, C., and Chan, H.M. (2020). Mammalian SWI/SNF Chromatin Remodeling Complexes: Emerging Mechanisms and Therapeutic Strategies. Trends Genet. *36*, 936–950.

Chakravarty, G., Moroz, K., Makridakis, N.M., Lloyd, S.A., Galvez, S.E., Canavello, P.R., Lacey, M.R., Agrawal, K., and Mondal, D. (2011). Prognostic significance of cytoplasmic SOX9 in invasive ductal carcinoma and metastatic breast cancer. Exp. Biol. Med. (Maywood). *236*, 145–155.

Chambers, A.F., Groom, A.C., and MacDonald, I.C. (2002). Dissemination and growth of cancer cells in metastatic sites. Nat. Rev. Cancer *2*, 563–572.

Charlton, J., Jung, E.J., Mattei, A.L., Bailly, N., Liao, J., Martin, E.J., Giesselmann, P., Brändl, B., Stamenova, E.K., Müller, F.J., et al. (2020). TETs compete with DNMT3 activity in pluripotent cells at thousands of methylated somatic enhancers. Nat. Genet. *52*, 819–827.

Chen, C.-H., Zheng, R., Tokheim, C., Dong, X., Fan, J., Wan, C., Tang, Q., Brown, M., Liu, J.S., Meyer, C.A., et al. (2020a). Determinants of transcription factor regulatory range. Nat. Commun. *11*, 2472.

Chen, H., Wu, J., Zhang, Z., Tang, Y., Li, X., Liu, S., Cao, S., and Li, X. (2018). Association Between BRCA Status and Triple-Negative Breast Cancer: A Meta-Analysis. Front. Pharmacol. *9*, 909.

Chen, S., Li, H., Li, X., Chen, W., Zhang, X., Yang, Z., Chen, Z., Chen, J., Zhang, Y., Shi, D., et al. (2020b). High SOX8 expression promotes tumor growth and predicts poor prognosis through GOLPH3 signaling in tongue squamous cell carcinoma. Cancer Med.

9, 4274–4289.

Chen, Y., Clarke, O.B., Kim, J., Stowe, S., Kim, Y.-K., Assur, Z., Cavalier, M., Godoy-Ruiz, R., von Alpen, D.C., Manzini, C., et al. (2016). Structure of the STRA6 receptor for retinol uptake. Science *353*, aad8266.

Cho, W.-K., Spille, J.-H., Hecht, M., Lee, C., Li, C., Grube, V., and Cisse, I.I. (2018). Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. Science *361*, 412–415.

Choi, J., Lysakovskaia, K., Stik, G., Demel, C., Söding, J., Tian, T. V, Graf, T., and Cramer, P. (2021). Evidence for additive and synergistic action of mammalian enhancers during cell fate determination. Elife *10*, e65381.

Chong, S., Dugast-Darzacq, C., Liu, Z., Dong, P., Dailey, G.M., Cattoglio, C., Heckert, A., Banala, S., Lavis, L., Darzacq, X., et al. (2018). Imaging dynamic and selective low-complexity domain interactions that control gene transcription. Science *361*.

Christin, J.R., Wang, C., Chung, C.Y., Liu, Y., Dravis, C., Tang, W., Oktay, M.H., Wahl, G.M., and Guo, W. (2020). Stem Cell Determinant SOX9 Promotes Lineage Plasticity and Progression in Basal-like Breast Cancer. Cell Rep. *31*, 107742.

Clevers, H., and Nusse, R. (2012). Wnt/β-catenin signaling and disease. Cell *149*, 1192–1205.

Corces, M.R., Trevino, A.E., Hamilton, E.G., Greenside, P.G., Sinnott-Armstrong, N.A., Vesuna, S., Satpathy, A.T., Rubin, A.J., Montine, K.S., Wu, B., et al. (2017). An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. Nat. Methods *14*, 959–962.

Corces, M.R., Granja, J.M., Shams, S., Louie, B.H., Seoane, J.A., Zhou, W., Silva, T.C., Groeneveld, C., Wong, C.K., Cho, S.W., et al. (2018). The chromatin accessibility landscape of primary human cancers. Science (80-.). *362*, eaav1898.

Cordeiro, T.N., Sibille, N., Germain, P., Barthe, P., Boulahtouf, A., Allemand, F., Bailly, R., Vivat, V., Ebel, C., Barducci, A., et al. (2019). Interplay of Protein Disorder in Retinoic Acid Receptor Heterodimer and Its Corepressor Regulates Gene Expression. Structure *27*, 1270-1285.e6.

Correia, A.L., Guimaraes, J.C., Auf der Maur, P., De Silva, D., Trefny, M.P., Okamoto, R., Bruno, S., Schmidt, A., Mertz, K., Volkmann, K., et al. (2021). Hepatic stellate cells suppress NK cell-sustained breast cancer dormancy. Nature *594*, 566–571.

D'Urso, A., and Brickner, J.H. (2014). Mechanisms of epigenetic memory. Trends Genet. *30*, 230–236.

Daniel, C.W., and Smith, G.H. (1999). The mammary gland: a model for development. J. Mammary Gland Biol. Neoplasia *4*, 3–8.

Dasgupta, A., Lim, A.R., and Ghajar, C.M. (2017). Circulating and disseminated tumor cells: harbingers or initiators of metastasis? Mol. Oncol. *11*, 40–61.

Dawson, H.D., Yamamoto, Y., Zolfaghari, R., Rosales, F.J., Dietz, J., Shimada, T., Li, N., and Ross, A.C. (2000). Regulation of Hepatic Vitamin A Storage in a Rat Model of Controlled Vitamin A Status during Aging. J. Nutr. *130*, 1280–1286.

Dekker, J., and Misteli, T. (2015). Long-Range Chromatin Interactions. Cold Spring Harb. Perspect. Biol. *7*.

Demicheli, R., Terenziani, M., and Bonadonna, G. (1998). Estimate of tumor growth time for breast cancer local recurrences: rapid growth after wake-up? Breast Cancer Res. Treat. *51*, 133–137.

Demicheli, R., Retsky, M.W., Hrushesky, W.J.M., and Baum, M. (2007). Tumor dormancy and surgery-driven interruption of dormancy in breast cancer: learning from failures. Nat. Clin. Pract. Oncol. *4*, 699–710.

Dent, R., Trudeau, M., Pritchard, K.I., Hanna, W.M., Kahn, H.K., Sawka, C.A., Lickley, L.A., Rawlinson, E., Sun, P., and Narod, S.A. (2007). Triple-negative breast cancer: Clinical features and patterns of recurrence. Clin. Cancer Res. *13*, 4429–4434.

Deryugina, E.I., and Quigley, J.P. (2006). Matrix metalloproteinases and tumor metastasis. Cancer Metastasis Rev. *25*, 9–34.

Dixon, J.R., Jung, I., Selvaraj, S., Shen, Y., Antosiewicz-Bourget, J.E., Lee, A.Y., Ye, Z., Kim, A., Rajagopal, N., Xie, W., et al. (2015). Chromatin architecture reorganization during stem cell differentiation. Nature *518*, 331–336.

Dukler, N., Gulko, B., Huang, Y.-F., and Siepel, A. (2017). Is a super-enhancer greater than the sum of its parts? Nat. Genet. *49*, 2–3.

Er, E.E., Valiente, M., Ganesh, K., Zou, Y., Agrawal, S., Hu, J., Griscom, B., Rosenblum, M., Boire, A., Brogi, E., et al. (2018). Pericyte-like spreading by disseminated cancer cells activates YAP and MRTF for metastatic colonization. Nat. Cell Biol. *20*, 966–978.

Fagnocchi, L., Mazzoleni, S., and Zippo, A. (2016). Integration of Signaling Pathways with the Epigenetic Machinery in the Maintenance of Stem Cells. Stem Cells Int. *2016*, 8652748.

Fagnocchi, L., Poli, V., and Zippo, A. (2018). Enhancer reprogramming in tumor progression: a new route towards cancer cell plasticity. Cell. Mol. Life Sci. *75*, 2537–2555.

Fang, X., Cai, Y., Liu, J., Wang, Z., Wu, Q., Zhang, Z., Yang, C.J., Yuan, L., and Ouyang, G. (2011). Twist2 contributes to breast cancer progression by promoting an epithelial–mesenchymal transition and cancer stem-like cell self-renewal. Oncogene *30*, 4707–4720.

Farrell, C.M., O'Leary, N.A., Harte, R.A., Loveland, J.E., Wilming, L.G., Wallin, C., Diekhans, M., Barrell, D., Searle, S.M.J., Aken, B., et al. (2014). Current status and new features of the Consensus Coding Sequence database. Nucleic Acids Res. *42*, D865-72.

Fasciani, A., D'Annunzio, S., Poli, V., Fagnocchi, L., Beyes, S., Michelatti, D., Corazza, F., Antonelli, L., Gregoretti, F., Oliva, G., et al. (2020). MLL4-associated condensates counterbalance Polycomb-mediated nuclear mechanical stress in Kabuki syndrome. Nat. Genet. *52*, 1397–1411.

Fe Lanfranco, M., Loane, D.J., Mocchetti, I., Burns, M.P., and Villapol, S. (2017). Combination of Fluorescent in situ Hybridization (FISH) and Immunofluorescence Imaging for Detection of Cytokine Expression in Microglia/Macrophage Cells. Bio-

Protocol *7*.

Feng, Y., Liu, X., and Pauklin, S. (2021). 3D chromatin architecture and epigenetic regulation in cancer stem cells. Protein Cell.

Fidler, I.J. (2003). The pathogenesis of cancer metastasis: the “seed and soil” hypothesis revisited. Nat. Rev. Cancer *3*, 453–458.

Fleischer, T., Tekpli, X., Mathelier, A., Wang, S., Nebdal, D., Dhakal, H.P., Sahlberg, K.K., Schlichting, E., Sauer, T., Geisler, J., et al. (2017). DNA methylation at enhancers identifies distinct breast cancer lineages. Nat. Commun. *8*, 1379.

Fluegen, G., Avivar-Valderas, A., Wang, Y., Padgen, M.R., Williams, J.K., Nobre, A.R., Calvo, V., Cheung, J.F., Bravo-Cordero, J.J., Entenberg, D., et al. (2017). Phenotypic heterogeneity of disseminated tumour cells is preset by primary tumour hypoxic microenvironments. Nat. Cell Biol. *19*, 120–132.

Foulkes, W.D., Stefansson, I.M., Chappuis, P.O., Bégin, L.R., Goffin, J.R., Wong, N., Trudel, M., and Akslen, L.A. (2003). Germline BRCA1 mutations and a basal epithelial phenotype in breast cancer. J. Natl. Cancer Inst. *95*, 1482–1485.

Friedmann-Morvinski, D., Bushong, E.A., Ke, E., Soda, Y., Marumoto, T., Singer, O., Ellisman, M.H., and Verma, I.M. (2012). Dedifferentiation of neurons and astrocytes by oncogenes can induce gliomas in mice. Science *338*, 1080–1084.

Fudenberg, G., Imakaev, M., Lu, C., Goloborodko, A., Abdennur, N., and Mirny, L.A. (2016). Formation of Chromosomal Domains by Loop Extrusion. Cell Rep. *15*, 2038–2049.

Fukaya, T., Lim, B., and Levine, M. (2016). Enhancer Control of Transcriptional Bursting. Cell *166*, 358–368.

Fumagalli, A., Oost, K.C., Kester, L., Morgner, J., Bornes, L., Bruens, L., Spaargaren, L., Azkanaz, M., Schelfhorst, T., Beerling, E., et al. (2020). Plasticity of Lgr5-Negative Cancer Cells Drives Metastasis in Colorectal Cancer. Cell Stem Cell *26*, 569-578.e7.

Furlong, E.E.M., and Levine, M. (2018). Developmental enhancers and chromosome topology. Science (80-.). *361*, 1341 LP – 1345.

Ganesh, K., and Massagué, J. (2021). Targeting metastatic cancer. Nat. Med. *27*, 34–44.

Ganesh, K., Basnet, H., Kaygusuz, Y., Laughney, A.M., He, L., Sharma, R., O’Rourke, K.P., Reuter, V.P., Huang, Y.H., Turkekul, M., et al. (2020). L1CAM defines the regenerative origin of metastasis-initiating cells in colorectal cancer. Nat. Cancer *1*, 28–45.

Ganji, M., Shaltiel, I.A., Bisht, S., Kim, E., Kalichava, A., Haering, C.H., and Dekker, C. (2018). Real-time imaging of DNA loop extrusion by condensin. Science (80-.). *360*, 102–105.

Germain, P., Chambon, P., Eichele, G., Evans, R.M., Lazar, M.A., Leid, M., De Lera, A.R., Lotan, R., Mangelsdorf, D.J., and Gronemeyer, H. (2006). International Union of Pharmacology. LX. Retinoic Acid Receptors. Pharmacol. Rev. *58*, 712 LP – 725.

Ghajar, C.M. (2015). Metastasis prevention by targeting the dormant niche. Nat. Rev.

Cancer *15*, 238–247.

Giancotti, F.G. (2013). XMechanisms governing metastatic dormancy and reactivation. Cell *155*, 750.

Gkountela, S., Castro-Giner, F., Szczerba, B.M., Vetter, M., Landin, J., Scherrer, R., Krol, I., Scheidmann, M.C., Beisel, C., Stirnimann, C.U., et al. (2019). Circulating Tumor Cell Clustering Shapes DNA Methylation to Enable Metastasis Seeding. Cell *176*, 98-112.e14.

Gonzalez-Avila, G., Sommer, B., Mendoza-Posada, D.A., Ramos, C., Garcia-Hernandez, A.A., and Falfan-Valencia, R. (2019). Matrix metalloproteinases participation in the metastatic process and their diagnostic and therapeutic applications in cancer. Crit. Rev. Oncol. Hematol. *137*, 57–83.

Greenwald, W.W., Li, H., Smith, E.N., Benaglio, P., Nariai, N., and Frazer, K.A. (2017). Pgltools: a genomic arithmetic tool suite for manipulation of Hi-C peak and other chromatin interaction data. BMC Bioinformatics *18*, 207.

Grimm, D., Bauer, J., Wise, P., Krüger, M., Simonsen, U., Wehland, M., Infanger, M., and Corydon, T.J. (2020). The role of SOX family members in solid tumours and metastasis. Semin. Cancer Biol. *67*, 122–153.

Guan, X., Jin, J., Tang, L., Wang, X., Wang, Y., and Shu, Y. (2019). Association of SOX9 expression with sensitivity to CDK7 inhibition and overall survival in triple-negative breast cancer. J. Clin. Oncol. *37*, 1065.

Gubern, A., Joaquin, M., Marquès, M., Maseres, P., Garcia-Garcia, J., Amat, R., González-Nuñez, D., Oliva, B., Real, F.X., de Nadal, E., et al. (2016). The N-Terminal Phosphorylation of RB by p38 Bypasses Its Inactivation by CDKs and Prevents Proliferation in Cancer Cells. Mol. Cell *64*, 25–36.

Gupta, P.B., Fillmore, C.M., Jiang, G., Shapira, S.D., Tao, K., Kuperwasser, C., and Lander, E.S. (2011). Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells. Cell *146*, 633–644.

Han, F., Liu, W., Jiang, X., Shi, X., Yin, L., Ao, L., Cui, Z., Li, Y., Huang, C., Cao, J., et al. (2015). SOX30, a novel epigenetic silenced tumor suppressor, promotes tumor cell apoptosis by transcriptional activating p53 in lung cancer. Oncogene *34*, 4391–4402.

Han, Y.H., Zhou, H., Kim, J.H., Yan, T.D., Lee, K.H., Wu, H., Lin, F., Lu, N., Liu, J., Zeng, J.Z., et al. (2009). A unique cytoplasmic localization of retinoic acid receptor-γ and its regulations. J. Biol. Chem. *284*, 18503–18514.

Harper, K.L., Sosa, M.S., Entenberg, D., Hosseini, H., Cheung, J.F., Nobre, R., Avivar-Valderas, A., Nagi, C., Girnius, N., Davis, R.J., et al. (2016). Mechanism of early dissemination and metastasis in Her2+ mammary cancer. Nature *540*, 588–592.

Harris, L.N., Broadwater, G., Lin, N.U., Miron, A., Schnitt, S.J., Cowan, D., Lara, J., Bleiweiss, I., Berry, D., Ellis, M., et al. (2006). Molecular subtypes of breast cancer in relation to paclitaxel response and outcomes in women with metastatic disease: results from CALGB 9342. Breast Cancer Res. *8*, R66.

Harrow, J., Frankish, A., Gonzalez, J.M., Tapanari, E., Diekhans, M., Kokocinski, F.,

Aken, B.L., Barrell, D., Zadissa, A., Searle, S., et al. (2012). GENCODE: the reference human genome annotation for The ENCODE Project. Genome Res. *22*, 1760–1774.

Hashimoto, Y., and Shudo, K. (1991). Retinoids and their nuclear receptors. Cell Biol. Rev. *25*, 209-230,233-235.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell *38*, 576–589.

Hiratsuka, S., Watanabe, A., Sakurai, Y., Akashi-Takamura, S., Ishibashi, S., Miyake, K., Shibuya, M., Akira, S., Aburatani, H., and Maru, Y. (2008). The S100A8-serum amyloid A3-TLR4 paracrine cascade establishes a pre-metastatic phase. Nat. Cell Biol. *10*, 1349–1355.

Hnisz, D., Abraham, B.J., Lee, T.I., Lau, A., Saint-André, V., Sigova, A.A., Hoke, H.A., and Young, R.A. (2013). Super-enhancers in the control of cell identity and disease. Cell *155*, 934–947.

Hnisz, D., Schuijers, J., Li, C.H., and Young, R.A. (2018). Regulation and Dysregulation of Chromosome Structure in Cancer. Annu. Rev. Cancer Biol. *2*, 21–40.

Hosseini, H., Obradovic, M.M.S., Hoffmann, M., Harper, K.L., Sosa, M.S., Werner-Klein, M., Nanduri, L.K., Werno, C., Ehrl, C., Maneck, M., et al. (2016). Early dissemination seeds metastasis in breast cancer. Nature *540*, 552–558.

Howlader, N., Noone, A.M., Krapcho, M., Garshell, J., Miller, D., Altekruse, S.F., Kosary, C.L., Yu, M., Ruhl, J., Tatalovich, Z., et al. (2014). SEER Cancer Statistics Review, 1975-2011, National Cancer Institute. Based on November 2013 SEER data submission, posted to the SEER web site, April 2014. Seercancergov/Csr.

Hu, Z., and Curtis, C. (2020). Looking backward in time to define the chronology of metastasis. Nat. Commun. *11*, 4–7.

Hu, M., Yao, J., Carroll, D.K., Weremowicz, S., Chen, H., Carrasco, D., Richardson, A., Violette, S., Nikolskaya, T., Nikolsky, Y., et al. (2008). Regulation of In Situ to Invasive Breast Carcinoma Transition. Cancer Cell *13*, 394–406.

Huang, H., Hu, J., Maryam, A., Huang, Q., Zhang, Y., Ramakrishnan, S., Li, J., Ma, H., Ma, V.W.S., Cheuk, W., et al. De fi ning super-enhancer landscape in triple-negative breast cancer by multiomic pro fi ling. 1–16.

Huang, Z., Yu, P., and Tang, J. (2020). Characterization of Triple-Negative Breast Cancer MDA-MB-231 Cell Spheroid Model. Onco. Targets. Ther. *13*, 5395–5405.

Iberg-Badeaux, A., Collombet, S., Laurent, B., van Oevelen, C., Chin, K.-K., Thieffry, D., Graf, T., and Shi, Y. (2017). A Transcription Factor Pulse Can Prime Chromatin for Heritable Transcriptional Memory. Mol. Cell. Biol. *37*.

Ishihara, K., Nakamoto, M., and Nakao, M. (2016). DNA methylation-independent removable insulator controls chromatin remodeling at the HOXA locus via retinoic acid signaling. Hum. Mol. Genet. *25*, 5383–5394.

Jana, S., Madhu Krishna, B., Singhal, J., Horne, D., Awasthi, S., Salgia, R., and Singhal,

S.S. (2020). SOX9: The master regulator of cell fate in breast cancer. Biochem. Pharmacol. *174*, 113789.

Jansen, A., van der Zande, E., Meert, W., Fink, G.R., and Verstrepen, K.J. (2012). Distal chromatin structure influences local nucleosome positions and gene expression. Nucleic Acids Res. *40*, 3870–3885.

Justilien, V., Regala, R.P., Tseng, I.-C., Walsh, M.P., Batra, J., Radisky, E.S., Murray, N.R., and Fields, A.P. (2012). Matrix metalloproteinase-10 is required for lung cancer stem cell maintenance, tumor initiation and metastatic potential. PLoS One *7*, e35040.

Kagey, M.H., Newman, J.J., Bilodeau, S., Zhan, Y., Orlando, D.A., van Berkum, N.L., Ebmeier, C.C., Goossens, J., Rahl, P.B., Levine, S.S., et al. (2010). Mediator and cohesin connect gene expression and chromatin architecture. Nature *467*, 430–435.

Kandaswamy, R., Sava, G.P., Speedy, H.E., Beà, S., Martín-Subero, J.I., Studd, J.B., Migliorini, G., Law, P.J., Puente, X.S., Martín-García, D., et al. (2016). Genetic Predisposition to Chronic Lymphocytic Leukemia Is Mediated by a BMF Super-Enhancer Polymorphism. Cell Rep. *16*, 2061–2067.

Kaplan, R.N., Riba, R.D., Zacharoulis, S., Bramley, A.H., Vincent, L., Costa, C., MacDonald, D.D., Jin, D.K., Shido, K., Kerns, S.A., et al. (2005). VEGFR1-positive haematopoietic bone marrow progenitors initiate the pre-metastatic niche. Nature *438*, 820–827.

Karagiannis, G.S., Pastoriza, J.M., Wang, Y., Harney, A.S., Entenberg, D., Pignatelli, J., Sharma, V.P., Xue, E.A., Cheng, E., D'Alfonso, T.M., et al. (2017). Neoadjuvant chemotherapy induces breast cancer metastasis through a TMEM-mediated mechanism. Sci. Transl. Med. *9*, eaan0026.

Kawai, T., Yasuchika, K., Ishii, T., Miyauchi, Y., Kojima, H., Yamaoka, R., Katayama, H., Yoshitoshi, E.Y., Ogiso, S., Kita, S., et al. (2016). SOX9 is a novel cancer stem cell marker surrogated by osteopontin in human hepatocellular carcinoma. Sci. Rep. *6*, 30489.

Kerr, S.C., and Corbett, A.H. (2010). Should INO Stay or Should INO Go: A DNA "Zip Code" Mediates Gene Retention at the Nuclear Pore. Mol. Cell *40*, 3–5.

Khalil, B.D., Sanchez, R., Rahman, T., Rodriguez-Tirado, C., Moritsch, S., Martinez, A.R., Miles, B., Farias, E., Mezei, M., Nobre, A.R., et al. (2021). An NR2F1-specific agonist suppresses metastasis by inducing cancer cell dormancy. J. Exp. Med. *219*, e20210836.

Khan, M.A., Jain, V.K., Rizwanullah, M., Ahmad, J., and Jain, K. (2019). PI3K/AKT/mTOR pathway inhibitors in triple-negative breast cancer: a review on drug discovery and future challenges. Drug Discov. Today *24*, 2181–2191.

Khan, M.I., Zamzami, M.A., Ahmad, A., and Choudhry, H. (2021). Molecular profiling of epigenetic landscape of cancer cells during extracellular matrix detachment. Sci. Rep. *11*, 2784.

Khiem, D., Cyster, J.G., Schwarz, J.J., and Black, B.L. (2008). A p38 MAPK-MEF2C pathway regulates B-cell proliferation. Proc. Natl. Acad. Sci. *105*, 17067 LP – 17072.

Khoo, W.H., Ledergor, G., Weiner, A., Roden, D.L., Terry, R.L., McDonald, M.M., Chai, R.C., De Veirman, K., Owen, K.L., Opperman, K.S., et al. (2019). A niche-dependent

myeloid transcriptome signature defines dormant myeloma cells. Blood *134*, 30–43.

Khurana, E., Fu, Y., Chakravarty, D., Demichelis, F., Rubin, M.A., and Gerstein, M. (2016). Role of non-coding sequence variants in cancer. Nat. Rev. Genet. *17*, 93–108.

Kiel, M.J., Yilmaz, O.H., Iwashita, T., Yilmaz, O.H., Terhorst, C., and Morrison, S.J. (2005). SLAM family receptors distinguish hematopoietic stem and progenitor cells and reveal endothelial niches for stem cells. Cell *121*, 1109–1121.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. *14*, R36.

Kim, T.-K., Hemberg, M., Gray, J.M., Costa, A.M., Bear, D.M., Wu, J., Harmin, D.A., Laptewicz, M., Barbara-Haley, K., Kuersten, S., et al. (2010). Widespread transcription at neuronal activity-regulated enhancers. Nature *465*, 182–187.

Klemm, S.L., Shipony, Z., and Greenleaf, W.J. (2019). Chromatin accessibility and the regulatory epigenome. Nat. Rev. Genet. *20*, 207–220.

Kocal, G.C., Güven, S., Foygel, K., Goldman, A., Chen, P., Sengupta, S., Paulmurugan, R., Baskin, Y., and Demirci, U. (2016). Dynamic Microenvironment Induces Phenotypic Plasticity of Esophageal Cancer Cells under Flow. Sci. Rep. *6*, 1–11.

Koedoot, E., Fokkelman, M., Rogkoti, V.-M., Smid, M., van de Sandt, I., de Bont, H., Pont, C., Klip, J.E., Wink, S., Timmermans, M.A., et al. (2019). Uncovering the signaling landscape controlling breast cancer cell migration identifies novel metastasis driver genes. Nat. Commun. *10*, 2983.

Konopka, T. (2020). umap: Uniform Manifold Approximation and Projection.

Kung, J.T., Kesner, B., An, J.Y., Ahn, J.Y., Cifuentes-Rojas, C., Colognori, D., Jeon, Y., Szanto, A., del Rosario, B.C., Pinter, S.F., et al. (2015). Locus-Specific Targeting to the X Chromosome Revealed by the RNA Interactome of CTCF. Mol. Cell *57*, 361–375.

Lambert, S.A., Jolma, A., Campitelli, L.F., Das, P.K., Yin, Y., Albu, M., Chen, X., Taipale, J., Hughes, T.R., and Weirauch, M.T. (2018). The Human Transcription Factors. Cell *175*, 598–599.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

Larsen, S.B., Cowley, C.J., Sajjath, S.M., Barrows, D., Yang, Y., Carroll, T.S., and Fuchs, E. (2021). Establishment, maintenance, and recall of inflammatory memory. Cell Stem Cell *28*, 1758-1774.e8.

Laughney, A.M., Hu, J., Campbell, N.R., Bakhoum, S.F., Setty, M., Lavallée, V.-P., Xie, Y., Masilionis, I., Carr, A.J., Kottapalli, S., et al. (2020). Regenerative lineages and immune-mediated pruning in lung cancer metastasis. Nat. Med. *26*, 259–269.

Lavoie, J.N., L'Allemain, G., Brunet, A., Müller, R., and Pouysségur, J. (1996). Cyclin D1 expression is regulated positively by the p42/p44MAPK and negatively by the p38/HOGMAPK pathway. J. Biol. Chem. *271*, 20608–20616.

Lawrence, M., Huber, W., Pagès, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M.T., and Carey, V.J. (2013). Software for Computing and Annotating Genomic

Ranges. PLOS Comput. Biol. *9*, e1003118.

Lee, J., Yesilkanal, A.E., Wynne, J.P., Frankenberger, C., Liu, J., Yan, J., Elbaz, M., Rabe, D.C., Rustandy, F.D., Tiwari, P., et al. (2019). Effective breast cancer combination therapy targeting BACH1 and mitochondrial metabolism. Nature *568*, 254–258.

Lefebvre, V., Dumitriu, B., Penzo-Méndez, A., Han, Y., and Pallavi, B. (2007). Control of cell fate and differentiation by Sry-related high-mobility-group box (Sox) transcription factors. Int. J. Biochem. Cell Biol. *39*, 2195–2214.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

Li, K., Liu, Y., Cao, H., Zhang, Y., Gu, Z., Liu, X., Yu, A., Kaphle, P., Dickerson, K.E., Ni, M., et al. (2020). Interrogation of enhancer function by enhancer-targeting CRISPR epigenetic editing. Nat. Commun. *11*, 485.

Lian, G.-Y., Wang, Q.-M., Mak, T.S.-K., Huang, X.-R., Yu, X.-Q., and Lan, H.-Y. (2021). Inhibition of tumor invasion and metastasis by targeting TGF-β-Smad-MMP2 pathway with Asiatic acid and Naringenin. Mol. Ther. - Oncolytics *20*, 277–289.

Liang, Y., Wu, H., Lei, R., Chong, R.A., Wei, Y., Lu, X., Tagkopoulos, I., Kung, S.-Y., Yang, Q., Hu, G., et al. (2012). Transcriptional Network Analysis Identifies BACH1 as a Master Regulator of Breast Cancer Bone Metastasis. J. Biol. Chem. *287*, 33533–33544.

Light, W.H., Freaney, J., Sood, V., Thompson, A., D'Urso, A., Horvath, C.M., and Brickner, J.H. (2013). A Conserved Role for Human Nup98 in Altering Chromatin Structure and Promoting Epigenetic Transcriptional Memory. PLOS Biol. *11*, e1001524.

Lin, L., Holmes, B., Shen, M.W., Kammeron, D., Geijsen, N., Gifford, D.K., and Sherwood, R.I. (2020). Comprehensive Mapping of Key Regulatory Networks that Drive Oncogene Expression. Cell Rep. *33*, 108426.

Liu, Z., Zhang, X.S., and Zhang, S. (2014). Breast tumor subgroups reveal diverse clinical prognostic power. Sci. Rep. *4*, 1–9.

Long, H.K., Prescott, S.L., and Wysocka, J. (2016). Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution. Cell *167*, 1170–1187.

Losada, A. (2014). Cohesin in cancer: chromosome segregation and beyond. Nat. Rev. Cancer *14*, 389–393.

Lu, J.-W., Hsieh, M.-S., Hou, H.-A., Chen, C.-Y., Tien, H.-F., and Lin, L.-I. (2017). Overexpression of SOX4 correlates with poor prognosis of acute myeloid leukemia and is leukemogenic in zebrafish. Blood Cancer J. *7*, e593–e593.

Luzzi, K.J., MacDonald, I.C., Schmidt, E.E., Kerkvliet, N., Morris, V.L., Chambers, A.F., and Groom, A.C. (1998). Multistep nature of metastatic inefficiency: dormancy of solitary cells after successful extravasation and limited survival of early micrometastases. Am. J. Pathol. *153*, 865–873.

Ma, Y., Shepherd, J., Zhao, D., Bollu, L.R., Tahaney, W.M., Hill, J., Zhang, Y., Mazumdar, A., and Brown, P.H. (2020). SOX9 is essential for triple-negative breast cancer cell survival and metastasis. Mol. Cancer Res. *18*, 1825–1838.

Madsen, J.G.S., Rauch, A., Van Hauwaert, E.L., Schmidt, S.F., Winnefeld, M., and Mandrup, S. (2018). Integrated analysis of motif activity and gene expression changes of transcription factors. Genome Res. *28*, 243–255.

Malik, V., Glaser, L. V, Zimmer, D., Velychko, S., Weng, M., Holzner, M., Arend, M., Chen, Y., Srivastava, Y., Veerapandian, V., et al. (2019). Pluripotency reprogramming by competent and incompetent POU factors uncovers temporal dependency for Oct4 and Sox2. Nat. Commun. *10*, 3477.

Malki, S., Boizet-Bonhoure, B., and Poulat, F. (2010). Shuttling of SOX proteins. Int. J. Biochem. Cell Biol. *42*, 411–416.

Malladi, S., Macalinao, D.G., Jin, X., He, L., Basnet, H., Zou, Y., de Stanchina, E., and Massagué, J. (2016). Metastatic Latency and Immune Evasion through Autocrine Inhibition of WNT. Cell *165*, 45–60.

Mangelsdorf, D.J., Ong, E.S., Dyck, J.A., and Evans, R.M. (1990). Nuclear receptor that identifies a novel retinoic acid response pathway. Nature *345*, 224–229.

Mangelsdorf, D.J., Umesono, K., Kliewer, S.A., Borgmeyer, U., Ong, E.S., and Evans, R.M. (1991). A direct repeat in the cellular retinol-binding protein type II gene confers differential regulation by RXR and RAR. Cell *66*, 555–561.

Mansour, M.R., Abraham, B.J., Anders, L., Berezovskaya, A., Gutierrez, A., Durbin, A.D., Etchin, J., Lawton, L., Sallan, S.E., Silverman, L.B., et al. (2014). An oncogenic super-enhancer formed through somatic mutation of a noncoding intergenic element. Science (80-.). *346*, 1373 LP – 1377.

Marcato, P., Dean, C.A., Liu, R.-Z., Coyle, K.M., Bydoun, M., Wallace, M., Clements, D., Turner, C., Mathenge, E.G., Gujar, S.A., et al. (2015). Aldehyde dehydrogenase 1A3 influences breast cancer progression via differential retinoic acid signaling. Mol. Oncol. *9*, 17–31.

Marches, R., Hsueh, R., and Uhr, J.W. (1999). Cancer dormancy and cell signaling: induction of p21(waf1) initiated by membrane IgM engagement increases survival of B lymphoma cells. Proc. Natl. Acad. Sci. U. S. A. *96*, 8711–8715.

Maruvada, P., Baumann, C.T., Hager, G.L., and Yen, P.M. (2003). Dynamic shuttling and intranuclear mobility of nuclear hormone receptors. J. Biol. Chem. *278*, 12425–12432.

McCarthy, D.J., Chen, Y., and Smyth, G.K. (2012). Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucleic Acids Res. *40*, 4288–4297.

Meacham, C.E., and Morrison, S.J. (2013). Tumour heterogeneity and cancer cell plasticity. Nature *501*, 328–337.

Mehta, G.A., Khanna, P., and Gatza, M.L. (2019). Emerging Role of SOX Proteins in Breast Cancer Development and Maintenance. J. Mammary Gland Biol. Neoplasia *24*, 213–230.

Mezquita, B., Mezquita, P., Pau, M., Gasa, L., Navarro, L., Samitier, M., Pons, M., and Mezquita, C. (2018). All-trans-retinoic acid activates the pro-invasive Src-YAP-Interleukin

6 axis in triple-negative MDA-MB-231 breast cancer cells while cerivastatin reverses this action. Sci. Rep. *8*, 7047.

Miao, Z., Deng, X., Shuai, P., and Zeng, J. (2018). Upregulation of SOX18 in colorectal cancer cells promotes proliferation and correlates with colorectal cancer risk. Onco. Targets. Ther. *11*, 8481–8490.

Mikule, K., Delaval, B., Kaldis, P., Jurcyzk, A., Hergert, P., and Doxsey, S. (2007). Loss of centrosome integrity induces p38-p53-p21-dependent G1-S arrest. Nat. Cell Biol. *9*, 160–170.

Millau, J.-F., Wijchers, P., and Gaudreau, L. (2016). High-Resolution 4C Reveals Rapid p53-Dependent Chromatin Reorganization of the CDKN1A Locus in Response to Stress. PLoS One *11*, e0163885.

Morgan, M.A., and Shilatifard, A. (2015). Chromatin signatures of cancer. Genes Dev. *29*, 238–249.

Morgan M, Obenchain V, Hester J, P.H. (2020). SummarizedExperiment: SummarizedExperiment container.

Morris, G.J., Naidu, S., Topham, A.K., Guiles, F., Xu, Y., McCue, P., Schwartz, G.F., Park, P.K., Rosenberg, A.L., Brill, K., et al. (2007). Differences in breast carcinoma characteristics in newly diagnosed African-American and Caucasian patients: a single-institution compilation compared with the National Cancer Institute's Surveillance, Epidemiology, and End Results database. Cancer *110*, 876–884.

Mukherjee, P., Gupta, A., Chattopadhyay, D., and Chatterji, U. (2017). Modulation of SOX2 expression delineates an end-point for paclitaxel-effectiveness in breast cancer stem cells. Sci. Rep. *7*, 9170.

Mumbach, M.R., Rubin, A.J., Flynn, R.A., Dai, C., Khavari, P.A., Greenleaf, W.J., and Chang, H.Y. (2016). HiChIP: efficient and sensitive analysis of protein-directed genome architecture. Nat. Methods *13*, 919–922.

Mumbach, M.R., Satpathy, A.T., Boyle, E.A., Dai, C., Gowen, B.G., Cho, S.W., Nguyen, M.L., Rubin, A.J., Granja, J.M., Kazane, K.R., et al. (2017). Enhancer connectome in primary human cells identifies target genes of disease-associated DNA elements. Nat. Genet. *49*, 1602–1612.

N., S., R., A., A.D., J., R.D., E., T.C., G., P., C., A.M., G., C., A., Saad, N., Alberio, R., et al. (2018). Cancer reversion with oocyte extracts is mediated by cell cycle arrest and induction of tumour dormancy. Oncotarget *9*, 16008–16027.

Nagao, A., Noie, T., Horiuch, H., Yamada, H., Momiyama, M., Nakajima, K., Satou, S., Satodate, H., Nara, S., and Harihara, Y. (2021). Long-term survival after pancreatic metastasis resection from breast cancer: a systematic literature review. Surg. Case Reports *7*.

Nagarajan, S., Hossan, T., Alawi, M., Najafova, Z., Indenbirken, D., Bedi, U., Taipaleenmäki, H., Ben-Batalla, I., Scheller, M., Loges, S., et al. (2014). Bromodomain Protein BRD4 Is Required for Estrogen Receptor-Dependent Enhancer Activation and Gene Transcription. Cell Rep. *8*, 460–469.

Nagelkerke, A., Bussink, J., Mujcic, H., Wouters, B.G., Lehmann, S., Sweep, F.C.G.J., and Span, P.N. (2013). Hypoxia stimulates migration of breast cancer cells via the PERK/ATF4/LAMP3-arm of the unfolded protein response. Breast Cancer Res. *15*, R2.

Naumov, G.N., MacDonald, I.C., Weinmeister, P.M., Kerkvliet, N., Nadkarni, K. V, Wilson, S.M., Morris, V.L., Groom, A.C., and Chambers, A.F. (2002). Persistence of solitary mammary carcinoma cells in a secondary site: a possible contributor to dormancy. Cancer Res. *62*, 2162–2168.

Nik-Zainal, S., Van Loo, P., Wedge, D.C., Alexandrov, L.B., Greenman, C.D., Lau, K.W., Raine, K., Jones, D., Marshall, J., Ramakrishna, M., et al. (2012). The life history of 21 breast cancers. Cell *149*, 994–1007.

O'Leary, N.A., Wright, M.W., Brister, J.R., Ciufo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D., et al. (2016). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res. *44*, D733-45.

O'Malley, R.C., Huang, S.C., Song, L., Lewsey, M.G., Bartlett, A., Nery, J.R., Galli, M., Gallavotti, A., and Ecker, J.R. (2016). Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape. Cell *166*, 1598.

Oldridge, D.A., Wood, A.C., Weichert-Leahey, N., Crimmins, I., Sussman, R., Winter, C., McDaniel, L.D., Diamond, M., Hart, L.S., Zhu, S., et al. (2015). Genetic predisposition to neuroblastoma mediated by a LMO1 super-enhancer polymorphism. Nature *528*, 418–421.

Oskarsson, T., Batlle, E., and Massagué, J. (2014). Metastatic stem cells: Sources, niches, and vital pathways. Cell Stem Cell *14*, 306–321.

Panda, M., Tripathi, S.K., and Biswal, B.K. (2021). SOX9: An emerging driving factor from cancer progression to drug resistance. Biochim. Biophys. Acta - Rev. Cancer *1875*, 188517.

Pantel, K., Brakenhoff, R.H., and Brandt, B. (2008). Detection, clinical relevance and specific biological properties of disseminating tumour cells. Nat. Rev. Cancer *8*, 329–340.

Parise, C.A., and Caggiano, V. (2017). Risk of mortality of node-negative, ER/PR/HER2 breast cancer subtypes in T1, T2, and T3 tumors. Breast Cancer Res. Treat. *165*, 743–750.

Park, S.Y., and Nam, J.S. (2020). The force awakens: metastatic dormant cancer cells. Exp. Mol. Med. *52*, 569–581.

Park, U.-H., Kim, E.-J., and Um, S.-J. (2010). A novel cytoplasmic adaptor for retinoic acid receptor (RAR) and thyroid receptor functions as a Derepressor of RAR in the absence of retinoic acid. J. Biol. Chem. *285*, 34269–34278.

Pennacchio, L.A., Bickmore, W., Dean, A., Nobrega, M.A., and Bejerano, G. (2013). Enhancers: five essential questions. Nat. Rev. Genet. *14*, 288–295.

Perou, C.M., Sùrlie, T., Eisen, M.B., Rijn, M. Van De, Jeffrey, S.S., Rees, C.A., Pollack, J.R., Ross, D.T., Johnsen, H., Akslen, L.A., et al. (2000). Molecular portraits of human

breast tumours. letters to nature 748. Nature *533*, 747–752.

Petkovich, M., Brand, N.J., Krust, A., and Chambon, P. (1987). A human retinoic acid receptor which belongs to the family of nuclear receptors. Nature *330*, 444–450.

Petrovic, J., Zhou, Y., Fasolino, M., Goldman, N., Schwartz, G.W., Mumbach, M.R., Nguyen, S.C., Rome, K.S., Sela, Y., Zapataro, Z., et al. (2019). Oncogenic Notch Promotes Long-Range Regulatory Interactions within Hyperconnected 3D Cliques. Mol. Cell *73*, 1174-1190.e12.

Phan, T.G., and Croucher, P.I. (2020). The dormant cancer cell life cycle. Nat. Rev. Cancer *20*, 398–411.

Phanstiel, D.H. (2019). Sushi: Tools for visualizing genomics data. R package version 1.22.0.

Philip, M., Fairchild, L., Sun, L., Horste, E.L., Camara, S., Shakiba, M., Scott, A.C., Viale, A., Lauer, P., Merghoub, T., et al. (2017). Chromatin states define tumour-specific T cell dysfunction and reprogramming. Nature *545*, 452–456.

Phillips, N.E., Mandic, A., Omidi, S., Naef, F., and Suter, D.M. (2019). Memory and relatedness of transcriptional activity in mammalian cell lineages. Nat. Commun. *10*, 1208.

Piva, M., Domenici, G., Iriondo, O., Rábano, M., Simões, B.M., Comaills, V., Barredo, I., López-Ruiz, J.A., Zabalza, I., Kypta, R., et al. (2014). Sox2 promotes tamoxifen resistance in breast cancer cells. EMBO Mol. Med. *6*, 66–79.

Plevritis, S.K., Munoz, D., Kurian, A.W., Stout, N.K., Alagoz, O., Near, A.M., Lee, S.J., Van Den Broek, J.J., Huang, X., Schechter, C.B., et al. (2018). Association of screening and treatment with breast cancer mortality by molecular subtype in US women, 2000-2012. JAMA - J. Am. Med. Assoc. *319*, 154–164.

Poli, V., Fagnocchi, L., Fasciani, A., Cherubini, A., Mazzoleni, S., Ferrillo, S., Miluzio, A., Gaudioso, G., Vaira, V., Turdo, A., et al. (2018). MYC-driven epigenetic reprogramming favors the onset of tumorigenesis by inducing a stem cell-like state. Nat. Commun. *9*.

Pomerantz, M.M., Qiu, X., Zhu, Y., Takeda, D.Y., Pan, W., Baca, S.C., Gusev, A., Korthauer, K.D., Severson, T.M., Ha, G., et al. (2020). Prostate cancer reactivates developmental epigenomic programs during metastatic progression (Springer US).

Pomp, V., Leo, C., Mauracher, A., Korol, D., Guo, W., and Varga, Z. (2015). Differential expression of epithelial–mesenchymal transition and stem cell markers in intrinsic subtypes of breast cancer. Breast Cancer Res. Treat. *154*, 45–55.

Psaila, B., and Lyden, D. (2009). The metastatic niche: Adapting the foreign soil. Nat. Rev. Cancer *9*, 285–293.

Pugacheva, E.M., Kubo, N., Loukinov, D., Tajmul, M., Kang, S., Kovalchuk, A.L., Strunnikov, A. V, Zentner, G.E., Ren, B., and Lobanenkov, V. V (2020). CTCF mediates chromatin looping via N-terminal domain-dependent cohesin retention. Proc. Natl. Acad. Sci. *117*, 2020 LP – 2031.

Quevedo, M., Meert, L., Dekker, M.R., Dekkers, D.H.W., Brandsma, J.H., van den Berg, D.L.C., Ozgür, Z., van IJcken, W.F.J., Demmers, J., Fornerod, M., et al. (2019). Mediator

complex interaction partners organize the transcriptional network that defines neural stem cells. Nat. Commun. *10*, 2669.

Quintanal-Villalonga, Á., Chan, J.M., Yu, H.A., Pe'er, D., Sawyers, C.L., Sen, T., and Rudin, C.M. (2020). Lineage plasticity in cancer: a shared pathway of therapeutic resistance. Nat. Rev. Clin. Oncol. *17*, 360–371.

Ranganathan, A.C., Adam, A.P., and Aguirre-Ghiso, J.A. (2006). Opposing roles of mitogenic and stress signaling pathways in the induction of cancer dormancy. Cell Cycle *5*, 1799–1807.

Rao, E., Hou, Y., Huang, X., Wang, L., Wang, J., Zheng, W., Yang, H., Yu, X., Yang, K., Bugno, J., et al. (2021). All-trans retinoic acid overcomes solid tumor radioresistance by inducing inflammatory macrophages. Sci. Immunol. *6*, eaba8426.

Rao, S.S.P., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., et al. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell *159*, 1665–1680.

Rauch, A., Haakonsson, A.K., Madsen, J.G.S., Larsen, M., Forss, I., Madsen, M.R., Van Hauwaert, E.L., Wiwie, C., Jespersen, N.Z., Tencerova, M., et al. (2019). Osteogenesis depends on commissioning of a network of stem cell transcription factors that act as repressors of adipogenesis. Nat. Genet. *51*, 716–727.

Raverdeau, M., and Mills, K.H.G. (2014). Modulation of T Cell and Innate Immune Responses by Retinoic Acid. J. Immunol. *192*, 2953 LP – 2958.

Rhie, S.K., Perez, A.A., Lay, F.D., Schreiner, S., Shi, J., Polin, J., and Farnham, P.J. (2019). A high-resolution 3D epigenomic map reveals insights into the creation of the prostate cancer transcriptome. Nat. Commun. *10*, 4154.

Rios, A.C., Capaldo, B.D., Vaillant, F., Pal, B., van Ineveld, R., Dawson, C.A., Chen, Y., Nolan, E., Fu, N.Y., Jackling, F.C., et al. (2019). Intraclonal Plasticity in Mammary Tumors Revealed through Large-Scale Single-Cell Resolution 3D Imaging. Cancer Cell *35*, 618-632.e6.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Research 43(7), e47.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. *43*, e47.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140.

Robson, M.I., de Las Heras, J.I., Czapiewski, R., Sivakumar, A., Kerr, A.R.W., and Schirmer, E.C. (2017). Constrained release of lamina-associated enhancers and genes from the nuclear envelope during T-cell activation facilitates their association in chromosome compartments. Genome Res. *27*, 1126–1138.

Rossetti, S., Ren, M., Visconti, N., Corlazzoli, F., Gagliostro, V., Somenzi, G., Yao, J., Sun, Y., and Sacchi, N. (2016). Tracing anti-cancer and cancer-promoting actions of all-trans retinoic acid in breast cancer to a RARα epigenetic mechanism of mammary epithelial cell fate. Oncotarget *7*, 87064–87080.

Sabari, B.R., Dall'Agnese, A., Boija, A., Klein, I.A., Coffey, E.L., Shrinivas, K., Abraham, B.J., Hannett, N.M., Zamudio, A. V., Manteiga, J.C., et al. (2018). Coactivator condensation at super-enhancers links phase separation and gene control. Science (80-.). *361*, eaar3958.

Sajnani, K., Islam, F., Smith, R.A., Gopalan, V., and Lam, A.K.-Y. (2017). Genetic alterations in Krebs cycle and its impact on cancer pathogenesis. Biochimie *135*, 164–172.

Samarut, E., and Rochette-Egly, C. (2012). Nuclear retinoic acid receptors: Conductors of the retinoic acid symphony during development. Mol. Cell. Endocrinol. *348*, 348–360.

Santos, N.C., and Kim, K.H. (2010). Activity of retinoic acid receptor-alpha is directly regulated at its protein kinase A sites in response to follicle-stimulating hormone signaling. Endocrinology *151*, 2361–2372.

Sarkar, A., and Hochedlinger, K. (2013). The Sox family of transcription factors: Versatile regulators of stem and progenitor cell fate. Cell Stem Cell *12*, 15–30.

Sartorelli, V., and Puri, P.L. (2018). Shaping Gene Expression by Landscaping Chromatin Architecture: Lessons from a Master. Mol. Cell *71*, 375–388.

Schuijers, J., Manteiga, J.C., Weintraub, A.S., Day, D.S., Zamudio, A.V., Hnisz, D., Lee, T.I., and Young, R.A. (2018). Transcriptional Dysregulation of MYC Reveals Common Enhancer-Docking Mechanism. Cell Rep. *23*, 349–360.

Servant, N., Varoquaux, N., Lajoie, B.R., Viara, E., Chen, C.-J., Vert, J.-P., Heard, E., Dekker, J., and Barillot, E. (2015). HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. Genome Biol. *16*, 259.

Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A., and Cavalli, G. (2012). Three-Dimensional Folding and Functional Organization Principles of the Drosophila Genome. Cell *148*, 458–472.

Shah, S.P., Roth, A., Goya, R., Oloumi, A., Ha, G., Zhao, Y., Turashvili, G., Ding, J., Tse, K., Haffari, G., et al. (2012). The clonal and mutational evolution spectrum of primary triple-negative breast cancers. Nature *486*, 395–399.

Shekar, N., Mallya, P., Gowda, D. V., and Jain, V. (2020). Triple-negative breast cancer: Challenges and treatment options. Int. J. Res. Pharm. Sci. *11*, 1977–1986.

Sitbon, D., Podsypanina, K., Yadav, T., and Almouzni, G. (2017). Shaping Chromatin in the Nucleus: The Bricks and the Architects. Cold Spring Harb. Symp. Quant. Biol. *82*, 1–14.

Siwek, W., Tehrani, S.S.H., Mata, J.F., and Jansen, L.E.T. (2020). Activation of Clustered IFNγ Target Genes Drives Cohesin-Controlled Transcriptional Memory. Mol. Cell *80*, 396-409.e6.

Sørlie, T., Perou, C.M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen,

M.B., Van De Rijn, M., Jeffrey, S.S., et al. (2001). Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc. Natl. Acad. Sci. U. S. A. *98*, 10869–10874.

Sosa, M.S., Avivar-Valderas, A., Bragado, P., Wen, H.-C., and Aguirre-Ghiso, J.A. (2011). Molecular Pathways ERK1/2 and p38a/b Signaling in Tumor Cell Quiescence: Opportunities to Control Dormant Residual Disease.

Sosa, M.S., Bragado, P., and Aguirre-Ghiso, J.A. (2014). Mechanisms of disseminated cancer cell dormancy: An awakening field. Nat. Rev. Cancer *14*, 611–622.

Sosa, M.S., Parikh, F., Maia, A.G., Estrada, Y., Bosch, A., Bragado, P., Ekpin, E., George, A., Zheng, Y., Lam, H.M., et al. (2015). NR2F1 controls tumour cell dormancy via SOX9- and RARβ-driven quiescence programmes. Nat. Commun. *6*.

Soutourina, J. (2018). Transcription regulation by the Mediator complex. Nat. Rev. Mol. Cell Biol. *19*, 262–274.

Spielmann, M., Lupiáñez, D.G., and Mundlos, S. (2018). Structural variation in the 3D genome. Nat. Rev. Genet. *19*, 453–467.

Sporikova, Z., Koudelakova, V., Trojanec, R., and Hajduch, M. (2018). Genetic Markers in Triple-Negative Breast Cancer. Clin. Breast Cancer *18*, e841–e850.

Srinivas, H., Juroske, D.M., Kalyankrishna, S., Cody, D.D., Price, R.E., Xu, X.-C., Narayanan, R., Weigel, N.L., and Kurie, J.M. (2005). c-Jun N-Terminal Kinase Contributes to Aberrant Retinoid Signaling in Lung Cancer Cells by Phosphorylating and Inducing Proteasomal Degradation of Retinoic Acid Receptor α. Mol. Cell. Biol. *25*, 1054–1069.

St Pierre, R., and Kadoch, C. (2017). Mammalian SWI/SNF complexes in cancer: emerging therapeutic opportunities. Curr. Opin. Genet. Dev. *42*, 56–67.

van Steensel, B., and Belmont, A.S. (2017). Lamina-Associated Domains: Links with Chromosome Architecture, Heterochromatin, and Gene Repression. Cell *169*, 780–791.

Stevison, F., Hogarth, C., Tripathy, S., Kent, T., and Isoherranen, N. (2017). Inhibition of the all-trans Retinoic Acid (atRA) Hydroxylases CYP26A1 and CYP26B1 Results in Dynamic, Tissue-Specific Changes in Endogenous atRA Signaling. Drug Metab. Dispos. *45*, 846–854.

Stros, M., Launholt, D., and Grasser, K.D. (2007). The HMG-box: a versatile protein domain occurring in a wide variety of DNA-binding proteins. Cell. Mol. Life Sci. *64*, 2590–2606.

Sze, C.C., and Shilatifard, A. (2016). MLL3/MLL4/COMPASS Family on Epigenetic Regulation of Enhancer Function and Cancer. Cold Spring Harb. Perspect. Med. *6*.

Tang, L., Jin, J., Xu, K., Wang, X., Tang, J., and Guan, X. (2020). SOX9 interacts with FOXC1 to activate MYC and regulate CDK7 inhibitor sensitivity in triple-negative breast cancer. Oncogenesis *9*, 47.

Tang, Z., Luo, O.J., Li, X., Zheng, M., Zhu, J.J., Szalaj, P., Trzaskoma, P., Magalska, A., Wlodarczyk, J., Ruszczycki, B., et al. (2015). CTCF-Mediated Human 3D Genome Architecture Reveals Chromatin Topology for Transcription. Cell *163*, 1611–1627.

Tariq, A., Hao, Q., Sun, Q., Singh, D.K., Jadaliha, M., Zhang, Y., Chetlangia, N., Ma, J.,

Holton, S.E., Bhargava, R., et al. (2020). LncRNA-mediated regulation of SOX9 expression in basal subtype breast cancer cells. RNA *26*, 175–185.

The Arabidopsis Genome Initiative, T.A.G. (2000). Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature *408*, 796–815.

Thomas, H.F., Kotova, E., Jayaram, S., Pilz, A., Romeike, M., Lackner, A., Penz, T., Bock, C., Leeb, M., Halbritter, F., et al. (2021). Temporal dissection of an enhancer cluster reveals distinct temporal and functional contributions of individual elements. Mol. Cell *81*, 969-982.e13.

Tischkowitz, M., Brunet, J.-S., Bégin, L.R., Huntsman, D.G., Cheang, M.C.U., Akslen, L.A., Nielsen, T.O., and Foulkes, W.D. (2007). Use of immunohistochemical markers can refine prognosis in triple negative breast cancer. BMC Cancer *7*, 134.

Tu, W.J., Hardy, K., Sutton, C.R., McCuaig, R., Li, J., Dunn, J., Tan, A., Brezar, V., Morris, M., Denyer, G., et al. (2017). Priming of transcriptional memory responses via the chromatin accessibility landscape in T cells. Sci. Rep. *7*, 44825.

Uy, G.K.A., Lisa, P., Irina, M., T., T.M., Anita, G., Ellen, W., R., I.N., Maria, N., Shijing, L., Siqi, L., et al. (2021). Stem cells expand potency and alter tissue fitness by accumulating diverse epigenetic memories. Science (80-.). *374*, eabh2444.

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

Vidal, E., Sayols, S., Moran, S., Guillaumet-Adkins, A., Schroeder, M.P., Royo, R., Orozco, M., Gut, M., Gut, I., Lopez-Bigas, N., et al. (2017). A DNA methylation map of human cancer at single base-pair resolution. Oncogene *36*, 5648–5657.

Vinci, M., Box, C., and Eccles, S.A. (2015). Three-dimensional (3D) tumor spheroid invasion assay. J. Vis. Exp. e52686.

Vinckier, N.K., Patel, N.A., Geusz, R.J., Wang, A., Wang, J., Matta, I., Harrington, A.R., Wortham, M., Wetton, N., Wang, J., et al. (2020). LSD1-mediated enhancer silencing attenuates retinoic acid signalling during pancreatic endocrine cell development. Nat. Commun. *11*, 2082.

Wainwright, E.N., and Scaffidi, P. (2017). Epigenetics and Cancer Stem Cells: Unleashing, Hijacking, and Restricting Cellular Plasticity. Trends in Cancer *3*, 372–386.

Wang, S.-P., Tang, Z., Chen, C.-W., Shimada, M., Koche, R.P., Wang, L.-H., Nakadai, T., Chramiec, A., Krivtsov, A. V, Armstrong, S.A., et al. (2017). A UTX-MLL4-p300 Transcriptional Regulatory Network Coordinately Shapes Active Enhancer Landscapes for Eliciting Transcription. Mol. Cell *67*, 308-321.e6.

Wang, Y.-F., Dang, H.-F., Luo, X., Wang, Q.-Q., Gao, C., and Tian, Y.-X. (2021a). Downregulation of SOX9 suppresses breast cancer cell proliferation and migration by regulating apoptosis and cell cycle arrest. Oncol Lett *22*, 517.

Wang, Y., Liu, J., Ying, X., Lin, P.C., and Zhou, B.P. (2016). Twist-mediated Epithelial-mesenchymal Transition Promotes Breast Tumor Cell Invasion via Inhibition of Hippo Pathway. Sci. Rep. *6*, 24606.

Wang, Z., Kang, L., Zhang, H., Huang, Y., Fang, L., Li, M., Brown, P.J., Arrowsmith, C.H.,

Li, J., and Wong, J. (2019). AKT drives SOX2 overexpression and cancer cell stemness in esophageal cancer by protecting SOX2 from UBR5-mediated degradation. Oncogene *38*, 5250–5264.

Wang, Z., Wang, P., Li, Y., Peng, H., Zhu, Y., Mohandas, N., and Liu, J. (2021b). Interplay between cofactors and transcription factors in hematopoiesis and hematological malignancies. Signal Transduct. Target. Ther. *6*, 24.

Weintraub, A.S., Li, C.H., Zamudio, A. V, Sigova, A.A., Hannett, N.M., Day, D.S., Abraham, B.J., Cohen, M.A., Nabet, B., Buckley, D.L., et al. (2017). YY1 Is a Structural Regulator of Enhancer-Promoter Loops. Cell *171*, 1573-1588.e28.

Welch, D.R., and Hurst, D.R. (2019). Defining the Hallmarks of Metastasis. Cancer Res. *79*, 3011–3027.

Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., and Young, R.A. (2013). Master Transcription Factors and Mediator Establish Super-Enhancers at Key Cell Identity Genes. Cell *153*, 307–319.

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.

Willis, R.A. (1934). The spread of tumours in the human body (J. & A. Churchill).

Wingelhofer, B., Neubauer, H.A., Valent, P., Han, X., Constantinescu, S.N., Gunning, P.T., Müller, M., and Moriggl, R. (2018). Implications of STAT3 and STAT5 signaling on gene regulation and chromatin remodeling in hematopoietic cancer. Leukemia *32*, 1713–1726.

Woodworth, A.M., and Holloway, A.F. (2017). The Role of Epigenetic Regulation in Transcriptional Memory in the Immune System (Elsevier Inc.).

Xu, S., Liu, Y., Zhang, T., Zheng, J., Lin, W., Cai, J., Zou, J., Chen, Y., Xie, Y., Chen, Y., et al. (2021). The Global, Regional, and National Burden and Trends of Breast Cancer From 1990 to 2019: Results From the Global Burden of Disease Study 2019. Front. Oncol. *11*, 689562.

Yachida, S., Jones, S., Bozic, I., Antal, T., Leary, R., Fu, B., Kamiyama, M., Hruban, R.H., Eshleman, J.R., Nowak, M.A., et al. (2010). Distant metastasis occurs late during the genetic evolution of pancreatic cancer. Nature *467*, 1114–1117.

Yang, S.X., and Polley, E.C. (2019). Systemic treatment and radiotherapy, breast cancer subtypes, and survival after long-term clinical follow-up. Breast Cancer Res. Treat. *175*, 287–295.

Yao, Y., Chu, Y., Xu, B., Hu, Q., and Song, Q. (2019). Risk factors for distant metastasis of patients with primary triple-negative breast cancer. Biosci. Rep. *39*, 1–10.

Yao, Z., Sun, B., Hong, Q., Yan, J., Mu, D., Li, J., Sheng, H., and Guo, H. (2015). The role of tumor suppressor gene SOX11 in prostate cancer. Tumour Biol. J. Int. Soc. Oncodevelopmental Biol. Med. *36*, 6133–6138.

Yeh, A.C., and Ramaswamy, S. (2015). Mechanisms of Cancer Cell Dormancy--Another Hallmark of Cancer? Cancer Res. *75*, 5014–5022.

Yoon, H.-G., Chan, D.W., Huang, Z.-Q., Li, J., Fondell, J.D., Qin, J., and Wong, J. (2003).

Purification and functional characterization of the human N-CoR complex: the roles of HDAC3, TBL1 and TBLR1. EMBO J. *22*, 1336–1346.

Yu, M., Bardia, A., Wittner, B.S., Stott, S.L., Smas, M.E., Ting, D.T., Isakoff, S.J., Ciciliano, J.C., Wells, M.N., Shah, A.M., et al. (2013). Circulating breast tumor cells exhibit dynamic changes in epithelial and mesenchymal composition. Science (80-.). *339*, 580–584.

Zaret, K.S., and Mango, S.E. (2016). Pioneer transcription factors, chromatin dynamics, and cell fate control. Curr. Opin. Genet. Dev. *37*, 76–81.

Zhang, K., Hocker, J.D., Miller, M., Hou, X., Chiou, J., Poirion, O.B., Qiu, Y., Li, Y.E., Gaulton, K.J., Wang, A., et al. (2021). A cell atlas of chromatin accessibility across 25 adult human tissues. BioRxiv 2021.02.17.431699.

Zhang, X., Choi, P.S., Francis, J.M., Imielinski, M., Watanabe, H., Cherniack, A.D., and Meyerson, M. (2016). Identification of focally amplified lineage-specific super-enhancers in human epithelial cancers. Nat. Genet. *48*, 176–182.

Zhang, X., Zhang, Z., Zhang, Q., Zhang, Q., Sun, P., Xiang, R., Ren, G., and Yang, S. (2018). ZEB1 confers chemotherapeutic resistance to breast cancer by activating ATM. Cell Death Dis. *9*, 57.

Zhao, Z., Zhang, Z., Li, J., Dong, Q., Xiong, J., Li, Y., Lan, M., Li, G., and Zhu, B. (2020). Sustained TNF-α stimulation leads to transcriptional memory that greatly enhances signal sensitivity and robustness. Elife *9*, e61965.

Zhou, Y., Gerrard, D.L., Wang, J., Li, T., Yang, Y., Fritz, A.J., Rajendran, M., Fu, X., Stein, G., Schiff, R., et al. (2019). Temporal dynamic reorganization of 3D chromatin architecture in hormone-induced breast cancer and endocrine resistance. Nat. Commun. *10*, 1522.

Zhu, F., Farnung, L., Kaasinen, E., Sahu, B., Yin, Y., Wei, B., Dodonova, S.O., Nitta, K.R., Morgunova, E., Taipale, M., et al. (2018). The interaction landscape between transcription factors and the nucleosome. Nature *562*, 76–81.

Zhu, L.J., Gazin, C., Lawson, N.D., Pagès, H., Lin, S.M., Lapointe, D.S., and Green, M.R. (2010). ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. BMC Bioinformatics *11*, 237.

Zhu, X., Chen, L., Huang, B., Wang, Y., Ji, L., Wu, J., Di, G., Liu, G., Yu, K., Shao, Z., et al. (2020). The prognostic and predictive potential of Ki-67 in triple-negative breast cancer. Sci. Rep. *10*, 225.

Zilli, F., Marques Ramos, P., Auf der Maur, P., Jehanno, C., Sethi, A., Coissieux, M.-M., Eichlisberger, T., Sauteur, L., Rouchon, A., Bonapace, L., et al. (2021). The NFIB-ERO1A axis promotes breast cancer metastatic colonization of disseminated tumour cells. EMBO Mol. Med. *13*, e13162.

Zou, H., Wang, S., Wang, S., Wu, H., Yu, J., Chen, Q., Cui, W., Yuan, Y., Wen, X., He, J., et al. (2018). SOX5 interacts with YAP1 to drive malignant potential of non-small cell lung cancer cells. Am. J. Cancer Res. *8*, 866–878.

(2019). Bioconductor Package Maintainer. liftOver: Changing genomic coordinate

systems with rtracklayer::liftOver.