The Role of HOXA5 in Triple-negative Breast Cancer (TNBC)

Rinad Mahmoud

MBBS (Hons), MSc

School of Medicine

Translational Medical Sciences Unit

Biodiscovery Institute -3

Project Supervisor: Dr Paloma Ordonez-Moran

Second Supervisor: Dr Abdolrahman Shams Nateri

Abstract

Breast cancer (BC) accounts for about 15% of all cancer types in the UK and is considered the most common malignancy among females. Triple-negative breast cancer (TNBC) subtype accounts for 10-20% of all BC subtypes. TNBC is known for its aggressive clinical course and poor patient outcome compared to other BC types. There is an urge to further understand the pathology and identify novel effective treatments for TNBC. Recently, the enrichment of TNBC tumours with cancer stem cells has been recognized as a key for their oncogenic properties. However, the specific mechanisms that regulate TNBC self-renewal, cellular differentiation, and proliferation and consequently their progression and migration ability is still unclear. Dysregulations in HOX gene family expression including HOXA5 transcription factor has been associated with cancer in many ways due to their role in cellular differentiation. However, the role of HOXA5 in stemness and cellular differentiation in TNBC is not clear. Hence, in this study, we aim to define HOXA5 specific role in stemness, cellular differentiation and proliferation in the different TNBC subtypes.

We generated lentivirus modified HOXA5 expressing and knockdown TNBC cells to investigate HOXA5 elicited transcriptomic and protein expression changes. We tested the effects of HOXA5 on a selected set of markers (proliferation, stemness, differentiation markers) using RT-qPCR. We also tested changes in cellular survival using clonogenic assay and assessed HOXA5 effect on cellular migration using wound-healing assay. Moreover, we investigated HOXA5 effects on TNBC stemness capacity through mammosphere formation assays. Our study also aims to identify the ability of retinoids to modulate HOXA5 expression and its subsequent effect on TNBC cellular phenotype and proliferation. Therefore, we tested retinoid as short- and long-term treatment regimens and tested the treatment effects on HOXA5 levels and the selected proliferation marker using RT-qPCR. Our study identified HOXA5 as an important regulator of cancer stem cells in TNBC. Our data demonstrate that HOXA5 is relevant for survival and stemness in TNBC. This may also correspond to the different subtypes, where this effect is further evoked in the M and MSL TNBC subtypes. We also believe that HOXA5 stimulates cellular invasion and migration in TNBC. Our data suggest that HOXA5 elicit these effects through modulating the expression of a set of genes involved in cancer cells stemness and proliferation (ALDH1 family members ALDH1A1 and ALDH1A3), adhesion and phenotypic changes (CDH1 and KRT14), invasion and proliferation (YAP target genes; TACSTD2, CTGF, and CYR61). Our study also showed that retinoids induce stemness in TNBC and suggest that HOXA5 could play an important role in retinoid induced stemness traits and maintenance in TNBC. The study added to the understanding of the stemness and differentiation process in TNBC tissue. It also provided evidence of HOXA5 playing an oncogenic role in TNBC despite its beneficial effect in other breast cancer subtypes.

1 Table of Contents

2 Introduction

2.1 Breast Cancer

Breast cancer (BC) is the most common malignancy in woman. According to the global cancer statistics, in 2012 about 1.7 million women were diagnosed with BC and approximately 521,900 women deaths were reported to be caused by BC[1]. BC is also the most common malignancy in the UK, accounting for 15% of all new cancer cases. Between 2015-2017 around 55,200 newly diagnosed breast cancer cases were reported in the UK every year. This is around 150 cases every day[2]. In the United Kingdom, BC is also the most common cancer among females, with approximately 54,700 new cases recorded in 2017 only. The incidence of breast cancer was projected to rise by 2% in the UK between 2014 and 2035, to about 210 cases per 100,000 females by 2035[2]. Breast cancer comprises a heterogeneous group of diseases in terms of clinical presentation, molecular characteristics, and response to therapy. Molecular pathology, histopathology, and gene-expression profiling have been used to identify distinct biological and clinical breast cancer subtypes[3].

At the 2013 St. Gallen International Breast Cancer Conference, the new definition of breast cancer molecular subtypes was issued, classifying BC into five different classes, according to their expression status of the hormonal oestrogen receptor (ER), hormonal progesterone receptor (PR), and the expression of the human epidermal growth factor receptor 2 (HER2). The identified subtypes include the luminal A (ER/PR^+, $HER2^-$), luminal B ($ER/PR^+ < 20\%$, $HER2^-$); $HER2^+$ B2 (ER/PR^+, HER2 overexpression), HER2 overexpression (ER^-, PR^-, HER2 overexpression), and basal-like triple-negative breast cancer (TNBC) (ER^-, PR^-, $HER2^-$)[4]. The molecular subtyping helps to stratify BC patients, facilitating treatment personalization to improve patient therapeutic response[4].

It is important to understand and identify the links between basal-like breast cancer (BLBC) and TNBC. BLBC characteristically express genes that are usually expressed in the basal/myoepithelial cells of the normal breast[5]. Hierarchical clustering analysis has clustered tumours that are ER- and HER-2-negative under BLBC subtype[5]. Despite the variety of basal markers expressed by this subtype, including different cytokeratin's (e.g. CK5/6, CK17, and CK14), the ER/HER2 negativity of these tumours remain characteristic[6]. Studies within the literature have included PR negativity to define a new entity of these tumours, namely the triple-negative subtype (TN; ER^-, PR^-, and $HER-2^-$)[7]. It has been constantly argued whether all BLBCs are TNBC and/or whether all TNBCs are BLBCs. 50% to 80% of TN tumours were noted to express basal markers[8,9]. This led to suggestions that the BLBC subtype could be replaced by the TN phenotype[10]. However, in 2009, a study showed that TN tumours comprise distinct clinical subtypes that differ in prognostic value and clinical response. The study also highlighted the importance of assessing the expression of basal markers to identify BLBCs among these distinct clinical and biological TN subgroup[6].

2.2 Triple-Negative Breast Cancer (TNBC)

TNBC is a subset of BC tumours that characteristically lack the expression of ER, PR, and HER2[4]. From a clinical perspective, TN tumours are one of the most relevant BC subgroups. Epidemiological data indicate that TNBCs represent 10%–20% of all breast cancers. It is frequently seen in younger patients and is more common in African- American women[11]. From a clinical point, TN tumours are considered an important subgroup of breast cancer due to their aggressive clinical behaviour. Compared with other breast cancer subtypes, TNBC patients have a more aggressive clinical course. Generally, TNBC tumours are larger in size and of a higher grade. They usually show lymph node involvement at the time of diagnosis[12]. Regardless of having high clinical response rates to presurgical chemotherapy (neoadjuvant therapy), TNBC patients have a worse prognosis relative to other breast cancer subtypes patients[5,6]. Patients have an increased chance of developing distant recurrence peaking at three years and a mortality rate of 40% within five years of diagnosis[13]. TNBC patient's treatment has been extremely challenging. This is referred to the lack of molecular targets and the heterogeneity of the disease. TNBC is not sensitive to endocrine therapy. Therefore, chemotherapy is the main systemic course of treatment for TNBC; however, conventional post-operative adjuvant chemoradiotherapy efficiency is inadequate. Consequently, the residual metastatic lesions will eventually lead to tumour recurrence[14].

It is clear that a better understanding of the TNBC molecular basis is required, and there is an urge to develop effective treatments for this aggressive breast cancer subtype. Extensive TNBC genomic, molecular, and biological analysis can help to understand the disease complexity and distinguish new molecular drivers that can be targeted therapeutically[15].

2.2.1 TNBC Subtypes and treatment progress

In a study done by *Lehmann et.* accurate molecular subtyping of TNBC was done using cluster analysis of various gene expression levels in a significant number of TNBC patients' samples leading to the identification of 6 TNBC subtypes. Cluster analysis was applied on gene expression (GE) profiles from 587 TNBC cases and 21 breast cancer patient's datasets. Each subtype displays a unique ontology and GE. The distinguished subtypes include two basal-like (BL1 and BL2), a mesenchymal (M), a mesenchymal stem-like (MSL), an immunomodulatory (IM), a luminal androgen receptor (LAR) subtype, and an unstable subtype (US) comprising a combination of markers. The study demonstrates that with the availability of adequate sample size, TNBC GE analysis can determine distinct subtypes with putative molecular targets. This can provide potential predictive biomarkers to help to stratify patients and tailoring treatment and consequently improve patient response to therapy. These biomarkers can also be used for patient selection when designing TNBC clinical trials. Moreover, the study also identified TNBC cell-line models for each of these subtypes using TNBC patient tumours derived GE signature analysis. Identification of these cell lines provides key models for preclinical studies; this will facilitate the development of new targeted agents. Further TNBC subtypes and cell lines molecular characterization using DNA copy number genomic data analyses, epigenetic, microRNA, and whole genome sequencing will help

identify novel driver signalling pathways in each subtype, highlighting targets for future TNBC therapy[15].

The Basal-like TNBC subtypes BL1 and BL2 are predominantly characterized by genomic instability where a number of basal-like TNBC subtypes correlated cell lines show nearly 2-fold the number of chromosome rearrangements compared to all other subtypes[16]. The basal-like subtypes were found to express high levels of proliferation and DNA damage response genes, and this suggests that basal-like TNBC tumours could profit from therapies targeting highly proliferative tumours such as DNA-damaging agents and anti-mitotic[17]. Studies supporting this finding showed that patients with basal-like tumours undergoing radiation-based and taxane-based treatment had about 4-fold higher pathologic Complete Response (pCR) relative to patients tumours with ML or LAR subtypes characteristics[18,19]. Studies suggest that developing markers to identify DNA damage response signalling defects and using proliferation biomarkers such as Ki67 can help to stratify patients for selective and tailored basal-like TNBC treatments[15].

Previously, about 50%–90% of TNBCs have been classified under the basal-like molecular subtype either by using IHC or by finding a correlation to the intrinsic molecular breast cancer subtypes[3]. TNBC and basal-like breast cancer relation remain controversial[6]. TNBCs with basal-like GE proportion in *Lehmann et.* study was 47%, showing a higher proportion of TNBCs that is correlated with the other molecular subtypes: normal breast-like (12%), luminal A (17%), luminal B (6%), HER2 (6%), or unclassified (12%). This study indicated that TNBC is not limited to basal-like phenotype comprising tumours; rather, it is a heterogeneous tumour collection with distinct phenotypes. This was supported by the diverse GE patterns identified by the study and the representative cell lines variation in sensitivity to the assessed targeted therapies in this study[15].

Looking at the M and MSL subtypes, the M subtype has a high active cell migration-associated signalling pathways regulated by actin protein, a high expression of extracellular matrix-receptor interacting pathways, and differentiation pathways, including the Wnt/β-catenin pathway, the transforming growth factor-β (TGF-β) signalling pathway, and the anaplastic lymphoma kinase pathway[15, 20]. Therefore, it was suggested that M subtype patients could be treated with epithelial-mesenchymal transition targeting agents[20]. On the other hand, the MSL subtype expresses lower cell proliferation-related genes related expression relative to the M subtype. MSL subtype highly expresses stemness-related genes such as ALDH1A1, ABCA8, PROCR, PER1, ENG, ABCB1, BCL2, BMP2, TERT2IP, and THY. The MSL subtypes also express high levels of several HOX genes, including HOXA5, HOXA10 involved in developmental processes. Many mesenchymal stem cell-specific markers were also noted in this subtype, like BMP2, ENG, KDR, ITGAV, NGFR, THY1, NT5E, PDGFR, and VCAM1[20]. A difference in clinical response was noted between the two subtypes, where M subtype patients present with shorter Relapse Free Survival (RFS). This was considered as a reflection of differences in proliferation between the two subtypes. The M subtype shows higher proliferation-associated genes expression, including Ki67[15]. Moreover, a decreased 5-year Distant Metastasis Free Survival (DMFS) was noted in

patients with M and MSL subtypes tumours. This was explained by the enrichment in epithelial-mesenchymal transition (EMT) and motility related pathways seen in tumours from M and MSL subtypes from *Lehmann et.* study[15]. It was previously shown that the EMT mediating gene Src has a prominent role in highly invasive cancer cells that have undergone EMT[21]. *Lehmann et.* data; suggested that the MSL subtype cell lines had differential sensitivity to dasatinib[15]. Dasatinib is a Src kinase family inhibitor, and it is a known potent suppressor of breast cancer stem cells in TNBC[22]. The study data proposed that EMT markers clinical value for preselection of patients in clinical trials using dasatinib[15].

Moreover, MSL subtype tumours GE profiles show similarities to mesenchymal cells and metaplastic breast cancers GE profile[23]. Metaplastic breast cancers are characterized by inherited plasticity[24]. Metaplastic breast cancers sequencing revealed that 47% of metaplastic cancers show high phosho-AKT expression and PIK3CA mutations[23]. This was consistent with *Lehmann et.* study findings, where TNBC MSL cell lines responded preferentially to NVP-BEZ235, the PI3K/mTOR inhibitor. Response to NVP-BEZ235 was seen in both xenografts with PIK3CA mutations and the MSL cell line (SUM159PT) that lacks PIK3CA mutation, highlighting the importance of the PI3K/mTOR pathway in the MSL TNBC subtype[15].

Furthermore, EMT can be regulated by the Wnt/β-catenin signalling pathway mediating tumour cell invasion. Wnt/β-catenin pathway (APC, CTNNB1, and WISP3) mutations occur in 52% of metaplastic breast cancer. This suggests that Wnt/β-catenin pathway dysregulation in these tumours can be used as a therapeutic target[25]. Wnt/β-catenin inhibitors are of great interest and are currently in clinical development[26]. This suggests that drugs targeting the Wnt/β-catenin pathway could be of great value for the MSL TNBC subtype treatment[15].

Within the IM subtype, a significant enrichment with immune cell-associated genes was noted. This was accompanied by enrichment in signal transduction pathways, including the NK cell pathway, the Th1/Th2 pathway, the dendritic cell (DC) pathway, the B cell receptor signalling pathway, the interleukin (IL)-12 pathway, the T cell receptor signalling, and the IL-7 pathway. Hence, a similarity between IM subtype and medullary carcinoma of the breast was proposed[20]. As a result of the identified GE in IM subtypes, the use of immune checkpoint inhibitors such as CTLA-4, PD1, and PDL1 is recommended for IM breast cancer subtype patients treatment[15].

The LAR subtype was subclassified according to AR gene signature and high expression of luminal cytokeratin[27]. The group distinguished 5 LAR subtype cell line models that showed high sensitivity to 17-DMAG and bicalutamide treatment, indicating that AR targeting therapies may be effective against AR tumours[15]. In addition to AR targeting agents, LAR cell lines were also sensitive to PI3K inhibition. The 5 LAR cells lines contain PIK3CA activating mutations and were sensitive to NVP-BEZ235, the PI3K inhibitor[15]. The study suggests that simultaneous targeting of PI3K/mTOR pathway and AR may be of clinical value for LAR TNBC patients[15].

Currently, most of TNBC molecular subtyping studies are based on the analysis of different genes mRNA levels. Nevertheless, the levels of mRNA expression do not accurately reflect the protein expression levels. Also, there are several protein translation modifications and regulatory steps, which can influence the prognostic prediction and targeted therapeutic effects in some patients. Concurrently, the accuracy in determining TNBC molecular subtype using immunohistochemical staining in clinical practice is still unclear, and its results are far from adequate. Hence, the identification of new TNBC subtype-specific therapeutic targets is required. Different TNBC molecular subtype associated biomarkers still await further study, and their clinical definitions should be determined to help improve patient therapeutic response[20].

2.3 Stemness and Cancer Stem cells (CSCs) in TNBC Biology and Therapeutics

Recently a new aspect in TNBC biology has been revealed. An enrichment of TNBC tumours with cancer stem cells (CSCs) was determined. This has been identified as another element that contributes to TNBC biology, tumorigenesis and resistance to therapy[28]. In normal breast tissue, the process of tissue maintenance and repair is dependent on the existence of tissue-resident stem cells, and this is defined by their multipotency and capacity for self-renewal. In 2006, *Shackleton et al.* discovered a mammary cells subpopulation within breast tissue, a population that does not express the endothelial marker CD31, neither the hematopoietic markers CD45 and TER119. This population was noted to highly express CD29 (β-1 integrin) and the human BC antigen CD24. *Shackleton et al.* reported the ability of these cells to reconstitute a complete mammary gland, where a single cell of this population can differentiate into complex alveolar-like structures that produces milk protein. This subpopulation was named mammary stem cells (MaSCs), a pool of cells that are able to generate neo-breast tissue and to maintain a stable pool of tissue-resident stem cell progenitors[29].

It was reported that the MaSCs capacity for self-renewal and multipotency is displayed in particular breast cancer subtypes, which share similar gene ontologies of MaSCs[15]. It has been hypothesized that TNBC growth originates from a CSCs population comprising oncogenic gene mutations that are critical to tumour growth and treatment resistance. This hypothesis was based on the identification of progenitor cells in TN tumours through their high expression of the CSCs markers CD44, CD24 and the Aldehyde Dehydrogenase 1 Family (ALDH1)[28]. A study that further supports this hypothesis identified enhanced stem cell gene expression signatures and high stem cell markers in the normal breast tissue of TNBC patients. This was considered as evidence that these cells might be the origin of the tumour or possibly a supportive microenvironment for tumorigenesis[30].

A great advancement in the management of non-TNBC breast cancer has been noted in the last decade, especially through using molecular characterization to guide for specific therapy. The growing perception of the importance of CSCs in TNBC, stem cell markers and perhaps the use of stem cell-targeting therapies may provide a major advance in TNBC treatment. However, the role of CSCs in TNBC still remains unclear. Also, stemness and self-renewal is a

complex process that is established and maintained by a molecular network. Identifying new contributors to TNBC stemness might assist in resolving this complexity and help identify new TNBC therapeutic strategies.

2.4 Homeobox-containing (HOX) Genes

Homeobox-containing (HOX) genes are master regulatory genes that encode the DNA-binding HOX proteins, which contribute to genomic regulation and control various morphogenesis characteristics and cell differentiation pathways[31]. The HOX proteins are transcription trans-regulators, meaning that they have the ability to repress or activate the expression of their target genes[31]. Murine and human samples genomic analysis revealed the existence of at least 39 HOX genes, organized in four clusters, namely, HOXA, HOXB, HOXC, and HOXD. Each of these clusters comprises between 9 and 11 genes that are localized in different chromosomes (in human chromosomes 7, 17, 12, and 2, respectively)[31].

The HOX genes earliest expression can be detected at the gastrulation phase of mammalian embryos. They are expressed within the three germinal layers of the embryo, where they are known to be involved in patterning the embryonic structures. The HOX genes help to determine the positioning of the axial skeleton, the genitalia, the upper and lower limbs, and the digestive tract. They are also involved in craniofacial morphogenesis and the nervous system development[31]. Homeobox genes dysregulation and mutational studies in mouse models have clearly supported their fundamental role in organogenesis. HOX genes mutations in humans were suggested to be responsible for some of the early spontaneous abortion cases. It was also proposed that HOX gene mutations are correlated with the increasing number of congenital syndromes[31].

2.4.1 HOX genes and Cancer

In cancer, HOX gene expression dysregulation has been recognized as a tumorigenesis driver[32]. HOX genes up-regulation or down-regulation have both been reported to be linked with cancer under different circumstances, where depending on cancer type, HOX genes can act as tumour suppressors or as proto-oncogenes[33]. For example, HOX gene activation was noted in murine and human leukemic cell lines. Particularly, HOXA cluster members expression was reported to be predominantly expressed within myelomonocytic cells. HOXA9 gene involvement in the chromosomal translocation t(7;11) (p15;p15) is recurrently reported in associated with AML, where it results in the production of the proto-oncogene NUP98/HOXA9, which may promote leukemogenesis by inhibiting HOXA9 mediated differentiation[34]. On the other hand, HOXA9 can act as a tumour suppressor, inhibiting tumour-phenotype in breast cancer through directly modulating the tumour suppressor gene BRCA1 expression[35].

Furthermore, many HOX genes deregulations have been noted in various cancer types, and their expression is often reported in association with increased cancer risk and poor cancer patients survival rates[32]. To further understand the nature of HOX genes related mutation, many studies have been developed, investigating different aspects of genomic regulation. Using integrated epigenomic scanning and gene expression profiling, aberrant methylation and epigenetic silencing of the HOXA gene cluster was noted in

human breast cancer[36]. The observed transcriptional repression was localized to the HOXA gene cluster and did not involve genes located upstream or downstream of the cluster. This study's findings were confirmed through chromatin immunoprecipitation, bisulfite sequencing, and RT-qPCR analysis. The loss of HOXA gene cluster expression in human breast cancer was closely linked to regional loss of permissive histone modifications and aberrant DNA methylation[36]. To support the study, pharmacologic manipulations of the breast cancer samples was done through in vitro treatment with 5-aza-dCyd the DNA methyltransferase inhibitor. Drug treatment resulted in reactivation of all silenced HOXA gene clusters confirming the previous results and proposing the significance of these aberrant epigenetic alterations in gene silencing in breast cancer. Overall, their data suggest that HOXA gene cluster inactivation in breast cancer represents a new type of genomic lesion-epigenetic microdeletion[36].

Moreover, evidence showed that the HOX genes contribute to seven different aspects of cancer development and progression, including angiogenesis, autophagy, differentiation, proliferation, invasion and metastasis, and apoptosis[33]. A vast amount of the literature is focused on HOX gene's role in cellular proliferation and differentiation. In the context of regeneration, aberrations in HOX genes expression can lead to increased cellular plasticity and loss of differentiation ability[37]. In tumours, cells can be derived to acquire stem cells phenotype and self-renewal characteristics. This can lead to the development of cancer stem cells and tumour initiation[38]. CSC are malignant stem/progenitor cells originating from either reprogramming of differentiated cells or transformation of normal stem/progenitor cells[38]. CSCs have been found in various tumour types[39]. Regardless of their origin, CSCs maintain a high degree of plasticity, and, in response to the tumour, microenvironment signals, are able to proliferate and give rise to a new heterogeneous malignant mass[40]. They are involved in tumour initiation, progression, invasion, and resistance to therapy and are linked directly to poor clinical outcome[40]. CSC formation requires attaining an epigenetic signature similar to that of normal stem cells, including certain chromatin modifications and DNA methylation that regulate stemness and differentiation genes[41]. Aberrations in HOX gene epigenetic regulation is commonly seen in cancer and can contribute to CSC plasticity[42]. One of the important examples is HOXA5 downregulation, a frequent event seen in breast cancer that is also associated with tumour aggressive, metastasis and poor prognosis[43].

2.5 HOXA5 in Breast Cancer

HOXA5 was identified as an important regulator of cancer growth and progression in various cancer types. It induces its effects through modulating different pathways affecting many of the Hallmarks of Cancer[33]. Loss of HOXA5 expression is frequently seen in breast cancer, and this is known to be correlated with higher pathological grade and poorer disease outcome[43]. A study by *Raman, V., Martensen, S., Reisman, D. et al.* identified HOXA5 loss as an important step in BC tumorigenesis. The study revealed that HOXA5 is a positive regulator of p53 transcription and function in luminal BC[44].

A study published by *Sukumar et.* in 2016 reported that HOXA5 expression in mammary epithelial cells stimulates transitioning of the progenitor population within the tumour into a more differentiated state. The study showed that the loss of HOXA5 in normal breast epithelial cells is accompanied by loss of the cell adhesion markers E-cadherin (CDH1) and CD24. Their data demonstrated that the loss of CDH1 and CD24 promotes the process of cellular dedifferentiation and transformation. This was directly linked to HOXA5, where HOXA5 was identified as a direct transcription regulator for CDH1 and CD24. This study identified a mechanism by which HOXA5 acts as a potential tumour suppressor through its direct action on lineage determination factors and cell adhesion in progenitor cells[43]. Their analysis of global genomic expression data from HOXA5-depleted breast non-malignant epithelial cell line MCF10A highlighted the role of HOXA5 in sustaining numerous molecular traits that are characteristic of the epithelial lineage, including tight junctions, cell-cell adhesion, and the expression of differentiation markers. In their work, HOXA5 depletion within the immortalized MCF10A cells reduced the stem cell $CD24^{+}$/CD44 expressing population, and this was accompanied by a reduction in CDH1 and CD24 expression and enhancement in the cells self-renewal capacity. Moreover, HOXA5 loss was noted to increase branching and protrusive morphology of these cells when suspended 3 dimensionally in Matrigel, suggesting the existence of epithelial to basal transition within these cells[43].

Interestingly, HOXA5 expression can be induced by drug treatment. A significant amount of the literature has addressed the use of retinoic acid (RA) analogues to induce HOXA5 expression in different types of tumours. In *Sukumar et.* work, retinal, a retinoic acid analogue was used to induce endogenous HOXA5 expression in non-malignant breast cells. To investigate HOXA5 contribution to loss and/or acquisition of epithelial traits in breast cancer tissue the group engineered non-malignant MCF10A-HOXA5 breast cells, which resulted in slowing the transition of these cells from the less differentiated $CD24^{-}$/$CD44^{+}$ state to a more differentiated $CD24^{+}$/$CD44^{+}$ status. This transition was promoted by retinal treatment, which stimulated HOXA5 expression, leading to re-expression of the tight junction integral plasma-membrane proteins occludin and claudin-7 (CLDN7) and epithelial markers CDH1 and CD24[43]. Moreover, in the luminal SUM149 breast cancer cells, HOXA5 overexpression was found to induce cellular apoptosis[45]. In fact, a study published in 2007 showed that HOXA5 is also involved in RA-mediated apoptosis and cell growth inhibition. The study proposed that the efficiency of retinoids as chemo preventive agents in breast cancer is based on their ability to provoke cellular differentiation and apoptosis through retinoid-binding and activation of the retinoic acid receptor β (RARβ)[45]. The result from these studies reveals a new concept that could possibly be tested in TNBC. The effects of HOXA5 in TNBC tissue remains an unclear area; this also applies to the effects of retinoids treatment on TNBC.

2.6 Retinoids

2.6.1 Origin, biological effects, derivatives

Retinoids are natural and synthetic derivatives of vitamin-A (retinol)[46]. A large body of literature documented retinoids involvement in various biological processes, including vision, embryologic development, reproduction, and immunologic regulation. They are also known to be an important physiological regulator of cellular growth and differentiation[47]. Plant carotenoids such as retinyl esters and β-carotene found in animal tissue are the natural dietary sources of retinoids[48]. Within the intestinal lumen, retinyl esters are hydrolysed into retinol then absorbed by the enterocytes, while β-carotene is converted into retinol in the intestinal mucosal cells. At high intake, retinol is transported by the circulation and stored in the liver. Many tissues can uptake and store retinol, and most can efficiently oxidize retinol to retinoic acid (RA)[46,49]. Despite retinol and its metabolite retinaldehyde being the functional retinoids involved in vision, the RA isomers, *9-cis-RA* (9cRA), *13-cis-RA* (13cRA) and *all-trans-RA* (atRA), are considered retinoid actions primary mediators in other physiological systems[46,49].

2.6.2 Retinoids in cancer treatment and chemoprevention

In the past decades, attention has been given to the role of retinoid signalling pathways in oncogenesis[46]. Epidemiological, preclinical, and clinical findings supported the use of retinoids for cancer therapy and chemoprevention[50]. Wolbach and Howe were the first to report vitamin-A related preclinical studies in 1925. The studies confirmed the development of squamous metaplasia in vitamin-A deficient rodents. In contrast, pulmonary metaplastic changes found in smokers were reversed by vitamin-A repletion[51]. Finding documented and reviewed throughout the past three decades provided a foundation for retinoids use in cancer chemoprevention clinical practice. In the nineties, successful treatment with retinoids has been recorded in many preneoplastic conditions, including cervical dysplasia, oral leucoplakia, and xeroderma pigmentosum. Retinoids were also found to reduce tumour secondaries in the liver and in the breast tissue[52,50].

Also, the remarkable clinical responses attained in acute promyelocytic leukaemia (APL) management using atRA acid and the effects of combined 13cRA and interferon therapy for cervical and cutaneous squamous cell carcinoma have reawakened interest in retinoids as chemotherapeutic agents[46,53]. Studies using animal models uncovered the chemo-preventive effects of retinoids on chemical mutagens exposed epithelial tissues[54]. Epidemiological observations previously reviewed by Hong and Itri in 1994 showed that individuals with lower intake of vitamin-A are more susceptible to the development of cancer, indicating an inverse relationship between cancer incidence and serum vitamin-A (β-carotene) levels[54,55].

In the 70s and the 80s, a new branch of cancer treatment called ‘’Differentiation Therapy’’ was introduced. Differentiation therapy targets and reactivates the cancer cells endogenous cellular differentiation program aiming to modest tumour phenotype through restoring the mutation process. This therapeutic approach offered a less aggressive treatment that reduced normal cell

damage[56]. Retinoids are key biological differentiation therapy compounds. This is referred to their important contribution to differentiation, growth, and apoptosis regulation[56]. Clinical studies have been conducted to test retinoids use for cancer therapy.

In contrast to the reported beneficial clinical effects, evidence of primary or acquired resistance to retinoids therapy has limited their clinical activity. It is crucial for their rational use in the clinic that the mechanisms of intrinsic or acquired resistance to retinoids are well understood to be successfully overcome (Figure. 1). This understanding will help predict patients that most likely will benefit from retinoid therapy and help identify strategies to optimize combination retinoid regimens to overcome therapeutic resistance[50].

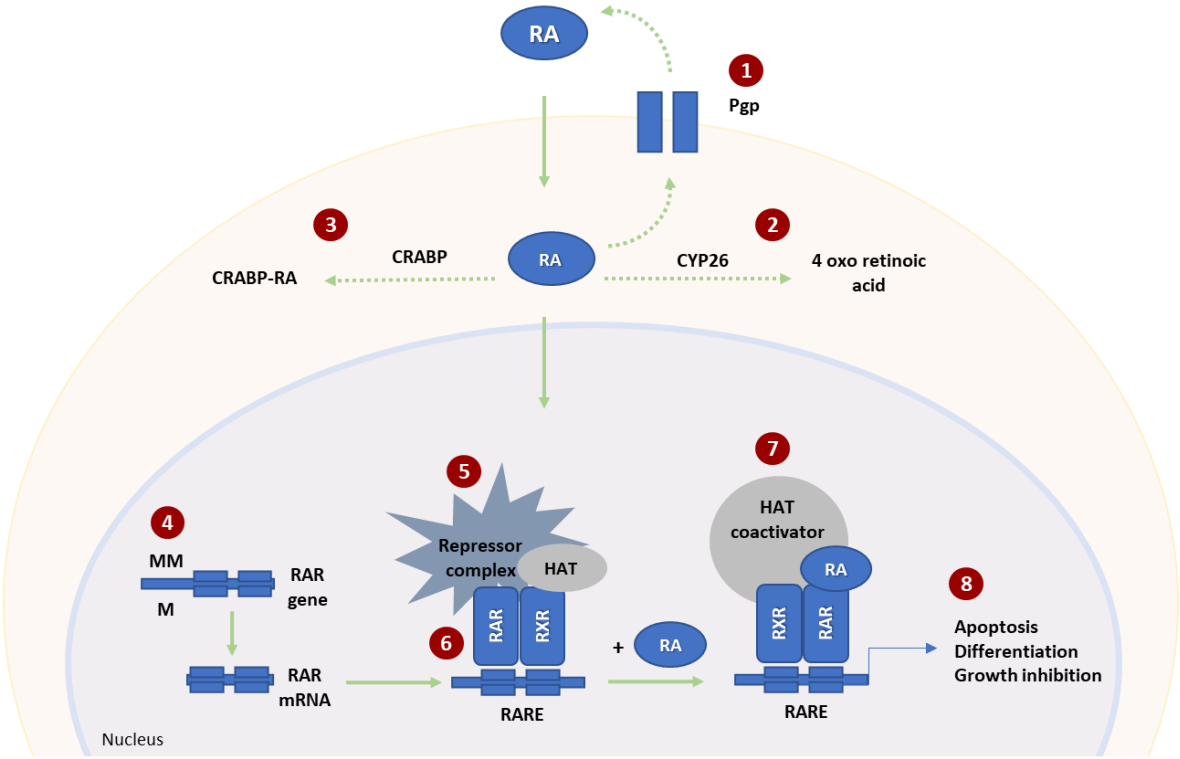

Figure 1: Retinoic acid (RA) potential therapeutic resistance mechanisms. *Cellular retinoid resistance may occur through (**1**) High P450 catabolism, (**2**) Increased drug export, (**3**) protein-mediated retinoids sequestration, (**4**) retinoic acid receptor (RARs) silencing through its gene promoter methylation, (**5**) persistent histone deacetylation, (**6**) Retinoic acid receptor (RAR) ligand-binding domain mutations and rearrangements. (**7**) coactivator factor alteration, or (**8**) altered expression of target gene expression downstream elements[50]. Pgp (P-glycoprotein), CRABP (cellular retinoic acid-binding protein), RXR (retinoids acid nuclear receptor X), HAT (Histone acetyltransferases), and RARE (Retinoic acid response elements).*

2.6.3 Retinoids Receptors and Pathways and their significance in Cancer

Retinoid's chemo-preventive and therapeutic effects are mediated through a number of cytoplasmic and nuclear receptors activating various intracellular pathways. However, despite the data obtained regarding this, the exact pathways involved remain unclear[57].

2.6.3.1 Cytoplasmic retinoid receptors

There are four cytoplasmic binding proteins for retinoids. Two of them were identified as retinol cytoplasmic binding proteins, namely, CRBP-I and CRBP-II. The other two are the retinoic acid cytoplasmic binding proteins CRABP-I and CRABP-II. The cytoplasmic retinoic acid-binding proteins function as retinoids intracellular regulators. They regulate retinoids intracellular concentration and act as a buffer through sequestering retinoic acid in the cytoplasm and preventing excess ligand from reaching the nucleus. They also carry and present retinoic acid to different enzyme systems for metabolism. The cytoplasmic receptors shuttle retinoic acid to the nuclear receptors[57] (Figure. 1).

Studies have revealed CRBP and CRABP expression and/or functional alterations in different tumours. CRBP expression dysregulation was linked to human breast cancers, where CRBP ectopic expression in SV40 T-antigen transformed human breast epithelial cell line significantly reduced cellular survival[58]. Moreover, CRBP hypermethylation was commonly noted in many tumours, including breast cancer tumours and cell lines. Further investigation of this area revealed that CRBP methylation occurred in premalignant lesions and frequently co-existed with RARβ2 hypermethylation. Therefore, it was suggested that the frequent cytoplasmic retinoids receptor epigenetic disruption in human cancer may have important implications for cancer prevention and treatment with retinoids[59].

2.6.3.2 Nuclear receptors

Retinoic acid action is known to be mediated through inducing gene transcription modifications. In 1987 Chambon's lab and Evans' lab reported the discovery of retinoic acid nuclear receptors[62,63]. These receptors shared structural modules with steroid hormone family receptors. cDNA libraries screening uncovered the existence of a retinoic acid nuclear receptors family. The family of retinoic acid receptors includes the *all-trans* and *9-cis* retinoic acid-binding RAR α, β and γ receptors[62]. They also include RXR α, β and γ receptors, which solely bind to *9-cis* retinoic acid but with high affinity[63]. Studies showed that RXR has the ability to heterodimers with a number of nuclear receptors, including RAR, Thyroid Hormone receptor, Vitamin D3 receptor, Peroxisomal Proliferator–Activator Receptor and numerous orphan receptors[57]. Under physiological conditions, RXR/RAR heterodimer is considered the only productive DNA binding form[57]. Usually, for the RXR heterodimers, RXR partners ligands binding to receptors can activate target gene transcription. Unlike in RAR/RXR heterodimer, where the ligand cannot bind to RXR due to the interference of steric effects[64]. Nevertheless, retinoic acid induced RAR activation results in a conformational change that allows RXR binding to the ligand. Ligand binding to both receptors within the heterodimer exerts a synergistic increase in target gene transcription[65]. Phenotypic analysis of RAR genes knock-out transgenic mice suggests that the RAR subtypes are functionally redundant[66]. Similarly, functional redundancy was attributed to RXR following its inactivation. This does not include RXRα, where its gene disruption in both alleles was found to be lethal for the embryo[57,67]. However, findings regarding receptor redundancy have been questioned, and it was

suggested that experimental conditions could be responsible for these findings[67].

Studies using F9 teratocarcinoma cells showed that RAR genes deletion could lead to different results suggesting that the different receptors control specific RA-induced differentiation pathways[68]. RAR genes can generate RNA isoforms. This is through using different promoters and splicing sites[69]. Despite the isoform's expression in a tissue-specific manner, it is not proved whether they regulate different genes sets or not[57].

2.6.3.3 Nuclear retinoid receptors and cancer

Alteration in RXR expression results in multitudes effects. This is referred to its ability to form heterodimers with a number of different nuclear receptors. Low RXRs expression has also been noted in different tumours, including non-small lung cancer and squamous cell carcinoma[70,71]. It was previously demonstrated that mechanistically disruption of the RXRα gene in mouse prostatic epithelia cells leads to prostate intraepithelial neoplasia, which can progress into high-grade neoplasms[72].

RAR expression and/or functional alterations have been noted in many cancers. In acute promyelocytic leukaemia, it was proved that genetic translocation resulting in the production of the oncogenic RARα containing fusion protein is the major cause of this disease[73]. Moreover, In the luminal MCF-7 human breast carcinoma cell line RARα2 was found to be silenced through promoter methylation. Re-expression of RARα2 was established using combined DNA methyl transferase inhibitor and histone deacetylase inhibitor. However, the response was correlated with ligand-dependent growth inhibition[74]. A significant amount of data from different studies showed that RARβ expression is lost in various tumours, including breast carcer[75]. RARβ is silenced either through promoter sequence methylation, deletion or mutation. Accordingly, retinoic acid administration can induce RARβ expression in some tumours, and this is accompanied by growth suppression[76,77].

RAR gene expression is frequently either lost or reduced in many cancer types due to hypermethylation or aberrant histone acetylation. RARβ2 promoter contains a RA response element (RARE) that is normally transactivated by RXR/RAR heterodimerization in response to RA binding[78]. Studies confirmed RARβ2 epigenetic silencing in breast[79,80]. A growing body of evidence supports the hypothesis that the retinoic acid receptor beta2 (RAR β-2) gene is a tumour suppressor gene that induces apoptosis and the chemo-preventive and therapeutic effects of retinoids are due to induction of RAR-beta2. During breast cancer progression, RAR β-2 is reduced or even lost. It is known from studies of other tumour-suppressor genes that methylation of the 5'-region is the cause of loss of expression. Several groups demonstrated that this is also true for the RAR β-2 in breast cancer by treating breast cancer cell lines with a demethylating agent and examining the expression of the RAR β-2 in response to a challenge with retinoic acid. Studies using sodium bisulfite genomic sequencing as well as methylation-specific PCR showed that a number of breast cancer cell lines as well as breast cancer tissue showed signs of methylation. The RAR β-2 gene was unmethylated in non-neoplastic breast tissue as well as in other normal tissues. A combination of retinoic acid with

demethylating agents as well as with histone deacetylase inhibitors acts synergistically to inhibit growth. This review presents data that suggest that treatment of cancer patients with demethylating agents followed by retinoic acid may offer a new therapeutic modality. Both the time of commencement of chemoprevention and the choice of substances that are able either to prevent de novo methylation or to reverse methylation-caused gene silencing may be important considerations[81].

2.7 HOXA5 and Retinoic acid

Retinoid's chemo preventive treatment efficiency in breast cancer depends on their ability to induce cellular differentiation and provoke cellular apoptosis. This is mediated through retinoids binding to RAR β. In a study done by *Sukumar et.,* a direct link between HOXA5 and retinoids was revealed in which HOXA5 was identified as a direct downstream effector of RAR β, and it was suggested that HOXA5 might contribute to the anticancer and chemo preventive effects induced by retinoids[45]. The study showed that HOXA5 expression could only be induced by RA in RARβ expressing breast cancer tissue. They were able to identify for the first time the RA response element in HOXA5 located in the 3' end of the HOXA5 gene. Using Chip assay on their in-vivo model's sample, they showed that RAR β binds directly to this region. Therefore, RAR β overexpression was found to strongly enhance cellular response to RA in breast cancer cells, while its knockdown did result in abolishing RA mediated HOXA5 expression. Moreover, the study highlighted the existence of a coordinated loss of both RAR β and HOXA5 expression within their tested MCF10A breast epithelial model during its transformation and progression stages. HOXA5 loss resulted in a partially abolished retinoid-induced apoptosis in these cells resulting in an increase of cellular survival upon RA treatment. This supported the intersection of RA/ RAR β and HOXA5 actions to induce apoptosis and differentiation in breast cancer cells in which HOXA5 functions directly downstream to RAR β[45].

Ordonez-Moran et al. have previously revealed a novel mechanism by which the induction of intestinal stem cell differentiation can impact colorectal tumour progression. The group showed that HOXA5 transcription factor expression is being repressed by the Wnt pathway as a strategy to enhance stem cells self-renewal potential. It was suggested that HOXA5 re-expression could induce the loss of cancer stem cell phenotype, leading to suppression of tumour progression and metastasis. *Ordonez-Moran et al. in vivo* work demonstrated that HOXA5 expression reduces CSCs and abrogates metastasis formation. The work showed that retinoids could reverse the Wnt-mediated HOXA5 inhibition and consequently suppress cancer growth *in vivo*[82]. Importantly, this study showed that HOXA5 induction is essential for retinoid-induced differentiation therapy in colon cancer, suggesting that the selection of colorectal cancer patients who may benefit from this therapy should be identified according to their HOX expression pattern[82]. Targeting the CSCs population within the tumour through HOXA5 induction is a very promising therapeutic strategy, where CSCs have been proven to be resistant to conventional therapy and are known to be responsible for tumour relapse.

The lab extended their studies to HOXA5 role in breast cancer. They tested HOXA5 over-expression (HOXA5 OE) in the low endogenous HOXA5 expressing TNBC cell line MDA-MB-231 xenograft tumours (Figure 2A). Results showed a significant reduction in size compared to control tumours (Figure. 2B). Nevertheless, further understating of HOXA5 role in stemness induction and maintenance, cellular differentiation and plasticity in breast cancer were required, especially in TNBC. Exploring the subsequent HOXA5 expression regulatory role in different TNBC subtypes is a new aspect that has not been previously tested. Also, drug-induced HOXA5 mediated differentiation should also be tested to present solid evidence for retinoids uses in combined TNBC therapy. Identifying HOXA5 previously as a direct downstream effector of RAR β in breast cancer provides a strong rationale for testing retinoids in TNBC[45].

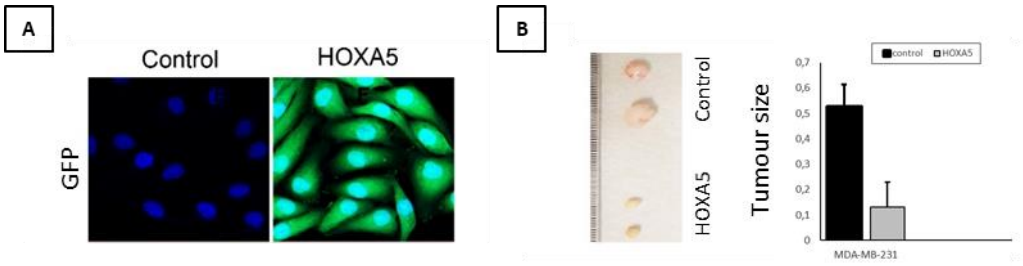

Figure 2: HOXA5 over-expression in the TNBC cell line MDA-MB-231. A. *Fluorescence microscopy images of MDA-MB-231 cells. Nuclei are stained by DAPI (blue) and the green colour is showing HOXA5-GFP in control and HOXA5 MDA-MB-231 cells.* ***B.*** *Left, image of xenograft tumours of control (top) and HOXA5 (below) MDA-MB-231 cells; Right, graph showing the tumours size, control (black) and HOXA5 (grey) MDA-MB-231 cells (n=3).*

3 Study Aims and objectives

Taking all together, we hypothesized that HOXA5 transcription factor plays an important regulatory role in TNBC subtype. Using retinoids to induce HOXA5 expression in TNBC can probably modulate the cancer stem cell population, cellular differentiation and proliferation, and the cancer cells migration ability (Figure. 4). Therefore, in this study, we aim to:

1. Define the importance of HOXA5 in Breast Cancer and different TNBC subtypes through investigating HOXA5 expression significance in breast cancer cell lines and patients' samples information within the public database and collaborators patients' data.

2. Define the role of HOXA5 in different TNBC subtypes through:

a. Generate lentivirus modified HOXA5 expressing and knockdown from different TNBC cell lines representing the basal-like (BL) and mesenchymal (M), and mesenchymal-like (MSL) TNBC subtypes (Table. 1).
b. Investigate the elicited transcriptomic changes from HOXA5 expression and knockdown on various stemness (ALDH1A1, ALDH1A3), differentiation (CDH1 and CK14) and proliferation (KI67) markers using qPCR analysis and immunofluorescence imaging.
c. Investigate the expression of other gene markers involved in different signalling pathways in HOXA5 overexpression or knockdown cells, including the Wnt

pathway (AXIN2 and β catenin) and the YAP pathway (CTGF, TACSTD2, and CYR61) using qPCR analysis and immunofluorescence of the MDA-MB-231-HOXA5, MDA-MB-468-HOXA5, and BT549-HOXA5 KD cells.

d. Assess the effects of HOXA5 on stemness capacity through testing the ability of HOXA5 modified cell lines to form 3D-mammospheres.

3. Test the effects of retinoids treatment as monotherapy on HOXA5 expression and its consequent impact on stemness, and proliferation markers in TNBC subtypes through:

a. Testing retinoids short treatment (96 hours) and long treatment (2 weeks) effects on different TNBC cell lines phenotype and stemness capacity through imaging, markers quantification using qPCR and mammosphere formation assay.

b. Compare both treatment effects on HOXA5 expression induction using qPCR.

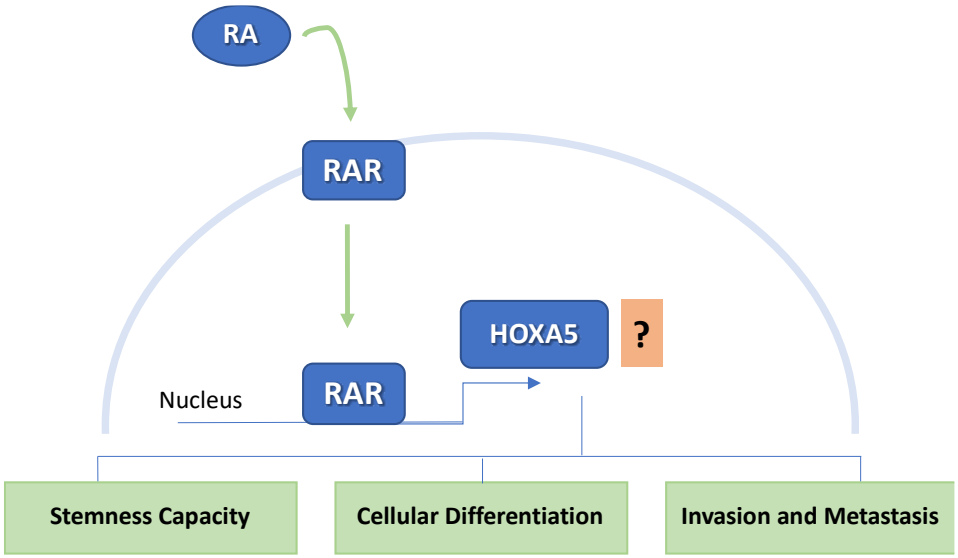

Figure 3: HOXA5 expression in TNBC modulate cancer stem cells population, cellular differentiation and proliferation and cancer cells invasive and metastatic properties. *Using retinoids in TNBC induce HOXA5 expression levels in response to their activation of Retinoic acid receptor (RAR). The consequent induction of HOXA5 in TNBC will modulate important cellular proliferation within the cells including the stem cells capacity, cellular differentiation and proliferation, and cellular invasion and migration ability.*

4 Materials and Methods

4.1 Cell culture and Cryopreservation

The human breast carcinoma-derived cells (see Table. 1) were provided by collaborating groups within Translational Medical Sciences Unit, BDI, University of Nottingham. Cells were previously authenticated by the collaborators using Short Tandem Repeats profiling. Cells were regularly tested for mycoplasma contamination. Cultivation was done in Class II Microbiological cabinets under sterile conditions. Cells were cultured in high glucose Dulbecco's Modified Eagle Medium (DMEM) (Gibco, ThermoFisher Scientific) with 10% foetal bovine serum (FBS, Invitrogen) and maintained at 37 °C with 5% CO_2. Liquid nitrogen frozen cells were manually defrosted. 1 mL of cells was transferred into a centrifuge tube in 3 mL of media. Samples were centrifuged (300g / 5 min) to remove the supernatant (90% DMEM and 10% DMSO), and the cell pellets were resuspended in 1 mL of pre-warmed medium. Next, cell-media solution was transferred into cell culture flasks and maintained at 37 °C with 5% CO_2 in DMEM with 10% FBS. Cells were regularly examined under the microscope to assess viability and confluence. On 80-90% confluence, splitting was done by removing old media and washing using 10 mL PBS. Cells were trypsinized and incubated at 37°C / 10 min. Next, Trypsin was deactivated using media, flask content was centrifuged (300g / 5 min), the supernatant was disposed, and the cell pellet was resuspended in DMEM with 10% FBS.

4.1.1 Cells cryopreservation

To avoid genetic changes and cellular loss through contamination, excess cells were cryopreserved in DMEM 10% FBS with 10% DMSO. Vials were stored at - 80°C until required.

Table 1: TNBC cell lines Subtyping

Cell Line	TNBC Subtype[A]	Intrinsic Subtype[B]	Basal Subtype[C]
MDA-MB-231	MSL	Unclassified	Basal B
MDA-MB-157	MSL	Unclassified	Basal B
MDA-MB-468	BL-1	Basal	Basal A
BT549	M	Unclassified	Basal B
HCC1806	BL-2	Unclassified	Basal A

Source: **A.** TNBC subtype Lehmann et.[15] **B.** Intrinsic Molecular subtype determined by correlation with UNC/intrinsic breast centroids[83] **C.** Basal subtype taken from Neve RM et al.[84]

4.2 Drug Treatment

Cells were plated in six-well plates at a density of $4x10^4$ cells per well 24 hours prior to drug treatment to allow cell attachment. Drugs were dissolved in dimethyl sulfoxide (DMSO)

(ThermoFisher Scientific, UK) at a final concentration of 1 µM all-trans-Retinoic acid (Retinoic Acid, *All Trans* Isomer, MP Biomedicals). Drug concentration was obtained from reviewing previously published work using the compound to induce HOXA5 levels in BC[85]. Concentrations were further optimised on our tested cell lines to reduce toxicity and maximise drug efficiency in inducing HOXA5 levels. Short treatment experiments were done in 6 days period. Long term treatment experiments were run for a period between 18 days during which cells were split according to cellular confluence at 1:10 ratio for MDA-MB-231, MDA-MB-157, BT549, and HCC1806, and at 1:5 ratio for the MDA-MB-468 cell line due to its lower doubling time. Cells were treated with drug dissolved in fresh for every dose. Two control wells were set per plate, one maintained treatment free and the other treated with DMSO to ensure results obtained are due to drug effects and not the dissolving reagent.

4.3 Mammospheres

24 well plates (ThermoFisher Scientific, UK) were coated with Poly 2-hydroxyethyl methacrylate (PolyHEMA) (Sigma-Aldrich) to create ultra-low adherent plates. Single cells were plated in the coated plates at $2x10^4$ cells/ mL density. Cells were grown in mammosphere medium consisting of DMEM Nutrient Mixture F-12 (DMEM/F12) supplemented with 20 ng/mL epidermal growth factor (Recombinant Human EGF, gibco, ThermoFisher Scientific) and B27. EGF was reconstituted in distilled water. Both EGF and B27 supplements are known to inhibit cellular differentiation[86]. Cells were incubated at 37 °C and 5% CO_2 for 7-10 days to attain first-generation mammospheres. Formed mammospheres were collected using a 1000 µl pipette, cells were gently centrifuged (200g / 5 min), and the supernatant was disposed. Cells were dissociated mechanically by pipetting and counted for re-plating at $5x10^3$ cells/ml density to retrieve the next generation. Serial passaging to attain consequent generations of mammospheres (up to three) was done with each generation cultivated in mammosphere medium for 7-10 days. Images were taken for the third-generation spheres using Nikon Eclipse Ti2 Widefield Fluorescent Microscope for representation.

4.4 Viral infection

HOXA5-GFP and HOXA5-TurboRFP cells were generated by lentiviral infection. Lentivirus particles and plasmids were previously generated and provided by Dr Ordonez-Moran. For this end, cells were placed in six-well plates as controls and infected cells at a density of $5x10^4$ cells per well 24 hours prior to viral infection. Cells were infected with a lentiviral vector and left overnight. Viral media was removed and replaced with DMEM with 10% FBS to cultivate cells. Successful infection was confirmed using fluorescence imaging (Nikon Eclipse Ti2 Widefield Fluorescent Microscope). Positive cells were isolated using Astrios Cell Sorter depending on GFP/RFP expression. The positive and negative populations were collected in 6 well plates. Media was replaced o the same day, and cells were allowed to expand.

4.5 Real-time (RT)-qPCR

4.5.1 RNA Extraction

RNA was prepared using miRNeasy Mini Kit (Qiagen, UK) from single cells or mammospheres. The media was removed. In the fume-hood, 200 µl of QIA-lysis buffer (Qiagen) was added per well, and cells were homogenized using a scraper. Samples were incubated for 5 minutes at room temperature. For phase separation of the samples, 40 µl of chloroform was added per sample, samples were vortexed for 15 seconds and centrifuged (8000g / 15 min / 4°C). The RNA containing aqueous phase was transferred into a new eppendorfs and 1.5 the sample volume of 75% ethanol was added. The mixture was transferred into a mini-column and centrifuged (12000g / 1 min) to discard ethanol. As the RNA is bound to the membrane, samples were washed with 500 µL of RWT and RPE buffer in a row and centrifuged (12000g / 1 min) to wash away the phenol and any other contaminants. Next, RNA was eluted in 20 µl of RNAse-free water. RNA samples were quantified using Nanodrop-2000c (Thermo Scientific. UK) and stored at - 80°C.

4.5.2 cDNA Synthesis

GoScript™ Reverse Transcription System (Invitrogen) was used for cDNA synthesis. For primers annealing, in a sterile eppendorf, 12 µl of RNA oligodT primers (d18T and d24T) and nucleotides (dNTPs) was prepared and incubated on a heating plate at 65°C for 5 minutes to melt template secondary structure and provide optimum temperature for oligodT primers attachment. Next, samples were held on ice for 5 minutes to prevent secondary structure reforming (see Table. 2 for volumes details). For reverse transcription reaction, Reverse Transcriptase Master-mix was prepared using Superscript II, 5x reaction buffer, DTT 0.01 M and ribonuclease inhibitor (RNasin) by multiplication of individual reagent volumes requirement per reaction (see Table. 3 for volumes details). For each sample, 8 µl of master-mix was added and samples were centrifuged (12000g / 1 min) and incubated in a thermal cycler at 42°C / 50 min and then at 70°C / 10 min to inactivate the enzyme. cDNAs were diluted using RNase free water and stored at - 20°C.

Table 2: Primer annealing mix for cDNA Synthesis

	Samples (RT)	**Neg. Control (RT)**
RNA (0.5-2 µg)	Accord to control	Accord to control
Primers (oligo dT)	1 µl	1 µl
dNTPs	1 µl	1 µl
Nuclease-free water	Accord to control	Accord to control
Total volume	12 µl	12 µl

Table 3: Reverse Transcription reaction Master-mix

Mix	**Sample**
Annealed primer/RNA/sample	12 µl
DTT, 0.1M	2 µl
5X Reaction Buffer	4 µl
Ribonuclease inhibitor (RNasin)	1 µl
RT enzyme Superscript	0.5 µl
Total volume	20 µl

4.5.3 RT-qPCR

4.5.3.1 Primers design:

To design primers for analysis of mRNA levels by quantitative PCR. Target gene mRNA full sequence as obtained from the National Centre for Biotechnology Information (NCBI) website (https://www.ncbi.nlm.nih.gov/nuccore/?term=HOXA5+homo+sapiens). Gene mRNA sequence was exported to the primer design online tool Primer3web 4.1.0 (https://primer3.ut.ee/). The tool was used to design the best possible forward and reverse primer sequence considering that each primers size is between 23 and 25 nucleotides, the selected primers lie approximately 100–150 nucleotides apart, and both primers are designed to have a similar melting time (T_m) (approximately 57°C). To reduces the possible background SYBR signal from contaminating genomic DNA within the sample primers were designed to span any intron/exon junction were possible. Next, the selected set of primers biological sequences were compared and assessed for their resemblance to genomic sequences at ‘’Human Genomic and transcript databases” using the online alignment tool NCBI Basic Local alignment tool (BLAST).

Note: Primers designed to assess HOXA5 Knockdown mRNA levels were designed to target the same splice forms that the shRNAs target. The specifications previously described for primers was applied.

(https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome). Primers were purchased from ThermoFisher Scientific, UK (see primer list including primers sequence below, Table. 4).

Before using a new set of primers in a quantitative experiment, primers were tested by adding a dissociation step to the PCR profile and the dissociation curve was assessed after real-time PCR is performed to be sure that a distinct peak of SYBR activity is apparent. Also, a titration of three 10-fold dilutions of a positive control cDNA sample was run to ensure that the SYBR activity accurately reflects the dilutions and select the optimal cDNA dilution for the experiment.

4.5.3.2 Performing qPCR

For each primer-set, a separate SYBR green Master-mix (Applied Biosystems, UK) was prepared by multiplying reagent volumes by the required number of samples to reduce pipetting errors (see Table. 5 for volumes details). A 10% extra master mix was prepared so that there will be enough to account for any pipetting errors. Samples were prepared in MicroAmp optical 96 well plates (Applied Biosystems, UK) and run in triplicate. Samples were probed for the housekeeping genes GAPDH and/or TBP as an internal control. Amplification was done by Applied Biosystems Viia7™ Real-time PCR system. Upon programme completion, ΔΔCt values and melting curves were generated through the built-in software. ΔΔCt values were exported to Microsoft Excel for further analysis.

Table 4: qPCR Primer Sequence

Gene	Forward Primer (5'-3')	Reverse Primer (5'-3')
GAPDH	GAGTCAACGGATTTGGTCGT	TTGATTTTGGAGGGATCTCG
TBP	CGAAACGCCGAATATAATCCCA	GACTGTTCTTCACTCTTGGCTC
RAR β	AGCTGGCTTGTCATAATTC	CATAGAAAAGTCCACCCAACTCC
HOXA5	TGAGCTTTAAAGTACTGAGCAG	TGGCAATAAACAGGCTCATGATT
HOXA5 KD	GCGAGCCACAAATCAAGCA	CCCTGAATTGCTCGCTCAC
KI67	ATGCAGAATCAGAAAGGGAAAGG	TTGTCTTTCTTGATCTCAGGCAC
CDH1	CACTTTTCATAGTTCCGCTCTGT	CACTTTTCATAGTTCCGCACAGT
KRT14	CCAATGTCCTTCTGCAGATTGAC	CAACTCTGTCTCATACTTGGTGC
ALDH1A1	AAAGCCATAACAATCTCCTCTGC	CCAGTTCTCTTCCATTTCCAGAC
ALDH1A3	AATAAAATGAGGGCCCGTAACAG	TATCTCTGACAAGGGTGATAGGC
AXIN2	AATTCCATACCGGAGGATGC	TCCACAGGCAAACTCATCG
TACSTD2	GAATCCATTGCGACATTGTTGTGAAG	CATAGGCCCAGTTAACAAACTCC
CTGF	AGATTCCCACCCAATTCAAAACA	CATTCTGGTGCTGTGTACTAATGT
CYR61	CCACACGAGTTACCAATGACAA	AATCCGGGTTTCTTTCACAAGG

Table 5: qPCR master mix volumes used to perform qPCR analysis of the tested samples

Master mix	Volume in µl
***Gene* primers mix**	1.5
SYBR green	7.5
H_2O	3
Total volume	12
Sample	3

4.6 Clonogenic assay

Following optimisation of clonogenic cell density of the tested cell lines. Cells were seeded at a concentration of 750 cells for the MDA-MB-231 and BT549 cells and at 1000 cells for the MDA-MB-468 per 10 cm petri dish in 15 mL of DMEM with 10% FBS to obtain average colony counts. Cells were incubated at 37 °C with 5% CO2. Over two weeks, cells were allowed to form colonies. Methylene blue solution was prepared by dissolving 10 g of methylene blue powder in 1 L of PBS diluted 70% ethanol. Media was removed from the dishes, and colonies were fixed and stained with 10 mL of 1% methylene blue (Sigma/Aldrich) at room temperature for 60 minutes. Methylene blue was aspirated from the dishes, and using tap water, flasks were rinsed three times. Flasks were set to dry overnight, and colonies were counted three times manually using a grid overlay to avoid double counting.

4.7 Migration Assay (Wound healing Assay)

Wound-healing assays are standard and commonly used methods for the investigation of cell migration. This method was implemented to evaluate the migration activity rate of MDA-MB-231-HOXA5 and MDA-MB-468-HOXA5 cells.

Cells were seeded in DMEM with 10% FBS within Culture-Insert 2 Well in µ-Dish 35 mm, high ibi treated (Thistle Scientific, UK) at 5×10^4 cells / well density for MDA-MB-231-HOXA5 cells and control, and at 75×10^3 cells / well density for MDA-MB-468-HOXA5 cells and the controls. After 24 h scratch was created by lifting the insert rubber partition with a sterile tweezer. Thereafter, the debris was removed, and cells were washed with 1 mL of DMEM with 1% FBS to assure the edges of the scratch were smoothed by washing. Utmost care was taken to create wounds of the same dimensions for both the experimental and control cells to minimize any possible variety resulting from a difference in scratch width.

The cells were then incubated with DMEM medium containing 1% FBS to stop cellular proliferation and induce cell migration. Images were taken using Nikon Eclipse Ti2 Widefield Fluorescent Microscope. Cell migration was assessed by ImageJ software (version 1.50i, National Institute of Health, Bethesda, MD, USA). The area of the initial wound was measured at 0 h, followed by gap area measurements after 8 h, 16 h, and 24 h for MDA-MB-231 and at 24 h and 48 hr for MDA-MB-468. Relative migration distance was assessed by the remaining area uncovered by the cells measured by ImageJ software. Values were plotted and compared by One-Way Analysis of variance (ANOVA), using GraphPad Prism.8.

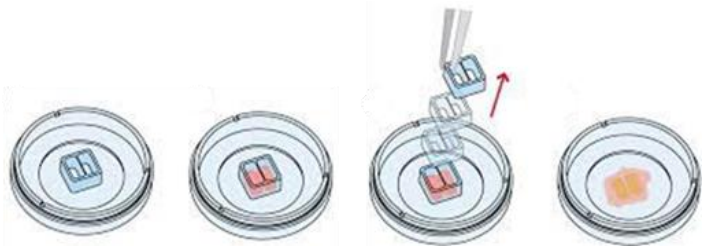

Figure 4: Schematic representation of steps to perform Wound Healing (Scratch test) using Culture-Insert. *Cells are seeded in DMEM with 10% FBS within Culture-Insert 2 wells. Cells are alloed to attach overnight. After 24 h scratch is created by lifting insert rubber partition with a sterile tweezer. Thereafter, the debris is removed, and cells should be washed gently with 1 mL of DMEM with 1% FBS to assure the edges of the scratch stay smooth by washing.*

4.8 Immunofluorescence (IF)

16 mm round cover slips were placed in a 24 well plate. Around 2×10^4 cell was seeded per well and allowed to attach and reach 60-80% confluence over 24-48 hrs depending on each cell line doubling time. On required confluence, media was removed, and cells were washed

once with PBS. Following that, cells were fixed using 250 µl of ice cold 4% paraformaldehyde (PFA) (ThermoFisher Scientific, UK) per well for 10 minutes. PFA was discarded, and cells were washed three times using PBS with 1 in 1000 tween®20 (Tween®20, Promega, UK) (5 min per wash). Cell permeabilization was done by adding 500 µl of PBS with 1% tween®20 (permeabilizing solution). Cells were incubated in permeabilizing solution for 10 minutes on the rocking plate at room temperature. The permeabilizing solution was removed, and cells were washed three times using PBS with 1 in 1000 tween®20 (5 min per wash). Next, 500 µl/well of blocking solution (50 mL PBS with 1 in 1000 tween®20 with 1.5-gram Bovine Serum Albumin (BSA) was added and incubated for 10 minutes on a rocker at room temperature. The blocking solution was removed, and cells were washed three times with PBS with 1 in 1000 tween®20 (5 min per wash). Primary antibodies (Mouse anti-Human*, Santa Cruz Biotechnology, Inc. USA*) were prepared at 1 in 100 concentration in blocking solution. 60-100 µl of the primary antibody solution was added as drops on a strip of parafilm (antibodies were labelled to avoid confusion). Using tweezers, cover slips were flipped with cell side on antibody drops. To maximise antibody affinity to its target antigen and maintain the overall solution reaction, cover slips were incubated overnight at 4°C and were kept under damp conditions using PBS wet towels. Cover slips were flipped back in a 24 well plate and washed three times using PBS with 1 in 1000 tween®20 (5 min per wash). Secondary antibody (IgG (H+L) Highly Cross-Adsorbed Goat anti-Mouse, Alexa Fluor® 488, Invitrogen) was prepared at 1 in 100 concentration with DAPI 1 in 500 concentration in the previously described blocking solution. 60-100 µl of the secondary antibody solution was added as drops on a strip of parafilm. Cover slips were flipped with cell side on antibody drops and incubated in the dark for 1 hour at room temperature. Cover slips were flipped back in 24 well plates and washed three times using PBS with 1 in 1000 tween®20 (5 min per wash). Cover slips were mounted onto a glass slide using 5 µl of PBS with Glycerol (50:50). Slides were stored in a dark box at 4°C. IF images were taken using Zeiss 780 confocal microscope.

4.9 Tumour Tissue sections Immunofluorescence (IF)

MDA-MB-231 control, MDA-MB-231 HOXA5 expression, MCF-7 wild-type, and MCF-7 HOXA5 expression xenograft tumour blocks were generated previously and provided by Dr Ordonez-Moran. From the tumour blocks, 4 mm tissue sections were cut using Leica RM2235 Manual Rotary Microtome for Routine Sectioning (Leica Microsystems UK Ltd). To allow paraffin bands to smooth and stick to the glass slides, cut bands were placed in 42°C water bath, then transferred to glass slides. Slides were incubated at 37°C overnight for paraffin bands to dry.

To remove paraffin, serial washing of slides was done in 100% xylene, 100% ethanol, 96% ethanol, 70% ethanol, and water (5 min per wash). The tissue fixative agent formaldehyde forms methylene bridges between proteins, which can hinder epitope recognition by primary antibodies. Therefore, citrate was used to retrieve tissue antigens. 1.4 g of citrate was dissolved in 500 mL of distilled water, and solution PH was measured and maintained at 6. The solution was heated to boiling point in the microwave. Slides were placed in citrate and heated in solution for 15 min using the microwave. Slides were allowed to slowly cool down to reach room temperature, then washed three times using PBS with 1 in 1000 tween®20 (Tween®20, Promega, UK). Following that, tissue permeabilization was done using PBS with 0.3% tween®20. Slides were incubated in permeabilizing solution for 10 min at room temperature. The permeabilizing solution was removed, and slides were washed three times using PBS with 1 in 1000 tween20 (5 min per wash). Next, 100 µl per tissue section of blocking solution (50 mL PBS with 1 in 1000 tween®20 with 1.5-gram Bovine Serum Albumin (BSA) was added and incubated for 10 minutes at room temperature. The blocking

solution was removed, and slides were washed three times using PBS with 1 in 1000 tween®20 (5 min per wash).

Primary antibodies (Mouse anti-Human, Santa Cruz Biotechnology, Inc. USA) was prepared at 1 in 100 concentration in blocking solution. To keep reagents localized on tissue specimens and prevents mixing of antibodies, a hydrophobic barrier was created around tissue sections using Liquid Blocker Super PAP Pen (Pyramid Innovation Ltd). 60-100 µl of the primary antibody solution was added per tissue section (slides were labelled with antibodies names to avoid confusion). To maximise antibody affinity to its target antigen and maintain the overall solution, reaction, slides were incubated overnight at 4°C and were kept under damp conditions using PBS wet towels. Next, slides were washed three times using PBS with 1 in 1000 tween®20 (5 min per wash). Secondary antibody (IgG (H+L) Highly Cross-Adsorbed Goat anti-Mouse, Alexa Fluor® 488, Invitrogen) was prepared at 1 in 100 concentrations with DAPI 1 in 500 concentration in the previously described blocking solution. 60-100 µl of the primary antibody solution was added per tissue section. Slides were incubated in the dark for 1 hour at room temperature. After that, slides were washed three times using PBS with 1 in 1000 tween®20 (5 min per wash). Cover slips were mounted onto the slides using 5 µl of PBS with Glycerol (50:50). Slides were stored in a dark box at 4°C. Tumour tissue sections IF images were taken using Zeiss 780 confocal microscope.

4.10 Bioinformatics and statistics

In collaboration with Prof. Rakha and Mongan group, we analysed The Cancer Genome Atlas BC cohort (TCGA, n=1211) data and the Nottingham TNBC cohort (n=112) RNA-seq patients' samples data. Nottingham TNBC cohort RNA sequencing was performed on an in house TNBC cohort FFPE tissue. Samples have been previously assessed histopathologically for tumour burden[87].

Data sets used to validate HOXA5 expression in TNBC subtypes were obtained from the publicly available cancer gene expression database R2 Genomics Analysis and Visualization Platform data analysis tool (https://hgserver1.amc.nl/cgi-bin/r2/main.cgi). The selected data set (Brown - 198 – MAS5.0 – u133p2) includes tissue microarray RNA profiling of 198 TNBC tumours[88]. R2 platform was also used to stratify HOXA5 expression levels in the different BC cell lines among the Hoeflitch - 51 - MAS5.0 - u133pu dataset (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=gse12777)[89]. The molecular subtype for this data set was determined using gene expression and HER2 status assessed by HER2 FISH analysis. Analysis of HOXA5 expression in cells from different TNBC subtypes was conducted on data set from the Library of Integrated Network-based Cellular Signatures (LINCS)[90] (https://lincs.hms.harvard.edu/db/datasets/20348/). Box-whisker plots depict mean, 5 and 95% quartile, min/max and standard error of the mean. Statistical significance of difference was assessed within the platform using the One-Way Analysis of variance (ANOVA). The data were considered not significant (ns) for $p > 0.05$.

Overall Survival (OS) and Relapse Free Survival analysis (RFS) was performed using the Kaplan-Meier Plotter online tool (https://kmplot.com/analysis/). Differences between survival distributions were analysed using the Log-rank test. Cox-Regression analysis was performed to assess covariates. Distant Metastasis Free Survival and 5 years Distant metastasis Free survival analysis was performed in collaboration with another group and using their patient data using SPSS. Differences between survival

distributions were analysed using the Log-rank test. Cox-Regression analysis was performed to assess covariates.

Statistical computations of qPCR, clonogenic assay, and mammospheres formation assay data were performed using GraphPad Prism.8. Experimental data were presented as mean ± standard deviation, and the means of the tested groups were compared using two-tailed unpaired student's t-test. The data were considered not significant (ns) for $p > 0.05$.

5 Results

5.1 HOXA5 expression in Breast Cancer (BC)

To understand the potential role of HOXA5 in breast cancer, initially we studied the expression of HOXA5 in healthy breast and various breast cancer tissues datasets in collaboration with Prof. Rakha and Mongan (see Methods). The analysis showed that HOXA5 expression was significantly increased in normal breast tissue compared to breast tumours (*p< 0.001*) (Figure. 5A). We also investigated the variation in HOXA5 expression among the different BC subtypes. Therefore, we classified patient's samples according to their molecular subtypes. Our analysis showed that HOXA5 expression was higher in TNBC tissue compared to luminal and HER +ve BC subtypes (*p< 0.001, p< 0.001*, respectively) (Figure. 5B).

To identify whether the variation in HOXA5 expression also exists within the different TNBC subtypes previously described by *Lehmann et.*[15], we examined HOXA5 expression in BL-1, BL-2, M, and LAR TNBC subtypes using a pre-existing cancer gene expression database through the R2 Genomics Analysis and Visualization Platform data analysis tool (see Methods) (Figure. 5C). The analysis showed that the M subtype expresses a significantly higher expression of HOXA5 compared to BL-1, BL-2, and LAR subtypes (*p<0.001, p< 0.001, and p=0.005,* respectively) (Figure. 5C).

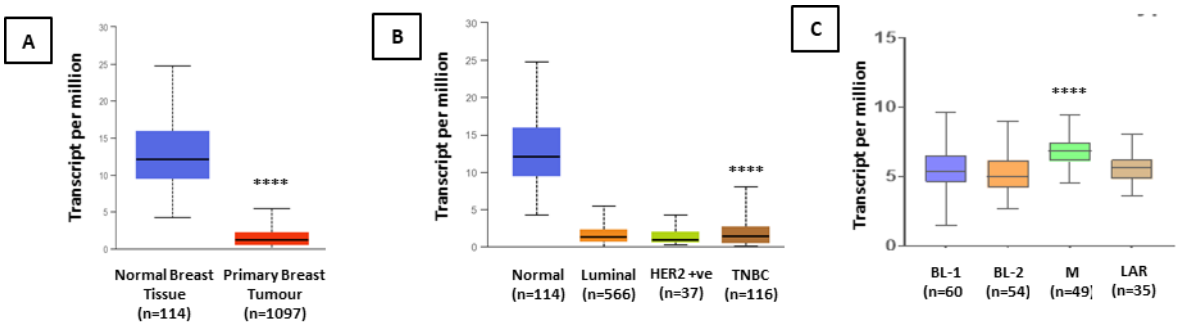

Figure 5: HOXA5 expression in BC: A. *Comparison of HOXA5 mRNA expression levels in normal breast tissue and breast cancer tissue.* ***B.*** *Comparison of HOXA5 expression in different BC molecular subtypes in TCGA data.* ***C.*** *Comparison of HOXA5 expression in different TNBC molecular subtypes (Basal Like-1, Basal Like-2, Mesenchymal, and Luminal androgen receptor (LAR). The Patients RNA samples were analysed using tissue microarrays for A and B. For C RNAseq was used. One-way analysis of variance was used to assess expression variation, * p <0.05, ** p<0.01, *** p<0.001.*

To explore our results further and investigate their relevance in BC cell lines genome, we also compared HOXA5 expression levels in human cell lines from luminal and basal BC subtypes, using a pre-existing cancer gene expression database of 51 cell lines (see Methods) using the R2 platform. Interestingly, the analysis showed that similarly, HOXA5 expression was significantly upregulated in TNBC cell lines relative to luminal BC cell lines (*p<0.001*) (Figure. 6A).

We also tested HOXA5 expression in different TNBC subtypes through analysing the LINCS dataset comprising the information of human cell lines representing four different TNBC subtypes; BL-1, BL-2, M, and MSL (see Methods). The results showed a higher expression of HOXA5 mRNA within the M subtype. However, the difference

was not statistically significant (M Vs BL-1 *p=0.8762*, M Vs BL-2 *p=0.9961*, and M Vs MSL *p=0.7622*) (Figure. 6B).

We also compared HOXA5 mRNA expression within the TNBC cell lines used in this study using RT-qPCR. We selected cell lines representing the BL-1, BL-2, M, and MSL (for cell line classification, see Methods Table. 3). Data showed that the BT549 expressed higher HOXA5 levels compared to the MDA-MB-231, MDA-MB-157, and MDA-MB-468 (*p<0.0001*, *p<0.0001,* and *p<0.0001* respectively) (Figure. 6C). Also, the HCC1806 significantly expressed high HOXA5 levels compared to all the other tested cell lines (*p<0.0001,* for all comparisons) (Figure. 6C). No significant difference was noted in HOXA5 expression between MDA-MB-468, MDA-MB-231 and MDA-MB-157 (*p=0.999, p=0.9978*, respectively) (Figure. 6C).

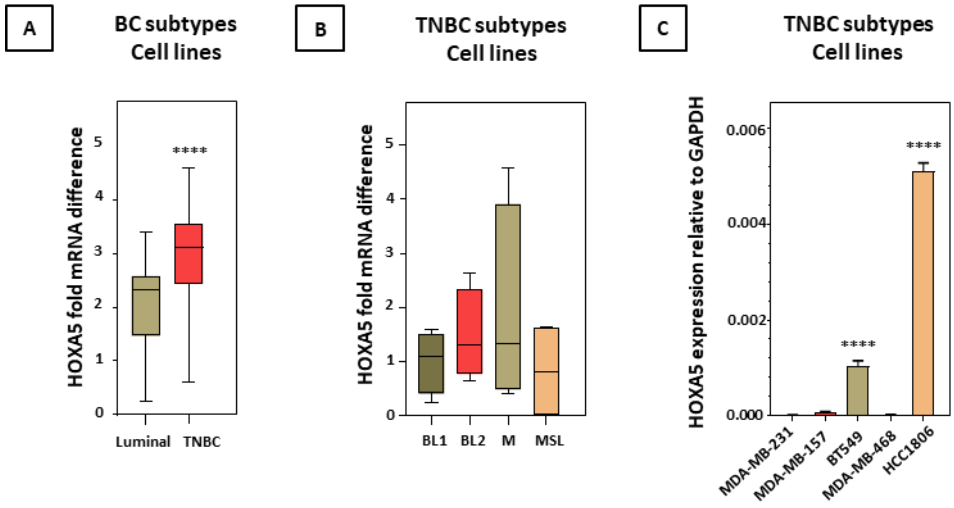

Figure 6: HOXA5 expression in Breast Cancer (BC) cell lines: ***A.*** *Comparison of HOXA5 mRNA expression levels in luminal and TNBC breast cancer cell lines.* ***B.*** *Comparison of HOXA5 expression within cell lines representing different TNBC molecular subtypes, Basal Like-1 (BL-1), Basal Like-2 (BL-2), Mesenchymal (M), and Mesenchymal-stem Like (ML) subtypes.* ***C.*** *Comparison of HOXA5 expression among the tested TNBC cell lines in this study. Cell lines mRNA expression was quantified using qPCR. Bars indicate averages ± standard errors. To assess expression variation two-tailed unpaired student's t-test for A and One-way analysis of variance for B and C was used, * p <0.05, ** p<0.01, *** p<0.001.*

5.2 HOXA5 expression in Survival Analyses

To investigate the prognostic value of HOXA5 for BC (all subtypes), we first investigated the survival analysis from the Kaplan-Meier Plotter database. We assessed the overall survival probability (OS) relative to HOXA5 expression levels in a dataset comprising 1880 patient samples. The analysis showed that HOXA5 mRNA levels are linked to BC patient's clinical outcome, where high HOXA5 mRNA expression was significantly associated with high OS (median, OS: 219 vs 219-month, *p=0.0036*). The calculated hazard ratio (HR) indicated that increased HOXA5 expression is associated with a 34% lower risk of death within the tested population (HR= 0.76, 95% CI 0.63 to 0.91) (Figure. 7A).

Also, 4934 samples from BC (all subtypes) were analysed to assess HOXA5 association with Relapse Free Survival (RFS). Analysis showed a significant increase

in RFS with high HOXA5 expression (median, OS: 231 Vs 231, *p=0.0023*). The HR indicated that increased HOXA5 expression increases patient relapse-free duration by 15% within the tested groups (HR= 0.85, 95% CI 0.77 to 0.95) (Figure. 7B).

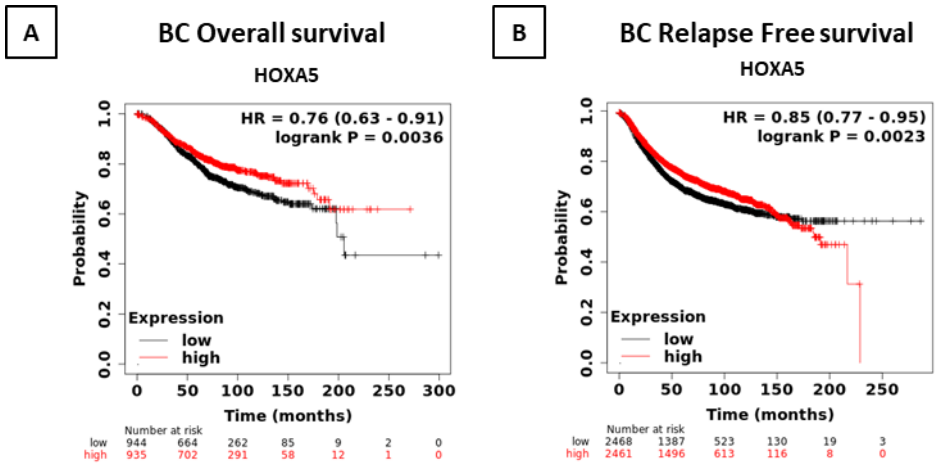

***Figure 7**: **Kaplan-Meier Survival curves based on HOXA5 expression status in BC** (black line indicates patients with low HOXA5 expression; red line indicates patients with high HOXA5 expression). **A.** Cumulative survival curves of BC patients with higher HOXA5 expression displayed higher OS (n=1880, median, OS: 219 vs 219-month, p=0.0036). **B.** Cumulative survival curves of BC patients with higher HOXA5 expression displayed a better RFS compared to patients expressing lower HOXA5 expression (n=4934, median, OS: 231 vs 231-month, p=0.0023). Kaplan-Meier method was used to estimate survival curve, and Long-rank test was used to assess difference in death probability between the two groups. Cox-Regression analysis was performed to assess covariates.*

We also explored HOXA5 prognostic value in only TNBC patients. HOXA5 mRNA expression did not affect patients OS significantly in 154 TNBC patients' samples (median, OS: 205 vs 205-month, *p=0.51*). However, the calculated HR indicated that increased HOXA5 expression is associated with a 17% lower risk of death within the tested population (HR= 0.83, 95% CI 0.49 to 1.43) (Figure. 8A). On the other hand, assessment of RFS in 405 TNBC patients showed that high HOXA5 expression significantly reduces RFS (median, RFS: 250 vs 250-month, *p=0.031*). The HR indicated that increased HOXA5 expression reduces patient relapse-free duration by 34% within the tested groups (HR= 1.34, 95% CI 1.03 to 1.74) (Figure. 8B).

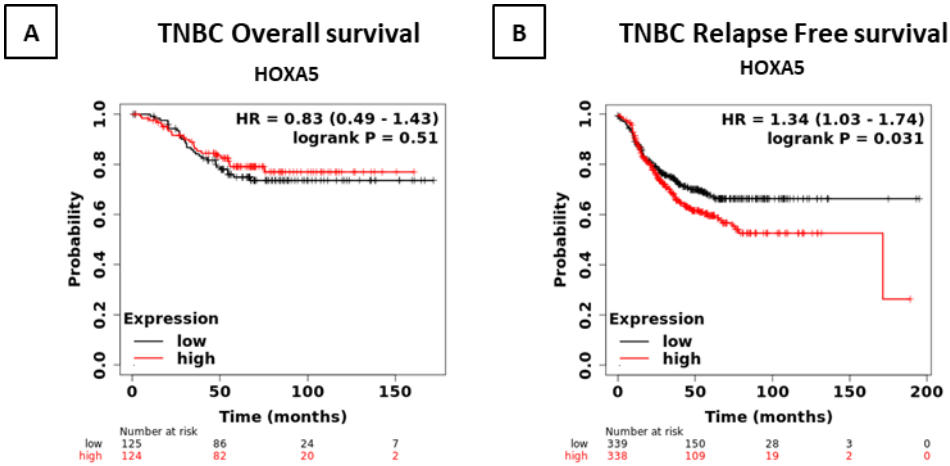

Figure 8: Kaplan-Meier Survival curves based on HOXA5 expression status in TNBC *(black line indicates patients with low HOXA5 expression; red line indicates patients with high HOXA5 expression).* ***A.*** *Cumulative survival curves of TNBC patients show no effect of HOXA5 expression levels on OS (n=154, median, OS: 205 vs 205-month, p=0.51).* ***B.*** *Cumulative survival curves of BC patients with higher HOXA5 expression displayed a worse RFS compared to patients expressing lower HOXA5 expression (n=405, median, RFS: 250 vs 250-month, p=0.031). Kaplan-Meier method was used to estimate survival curve, and Long-rank test was used to assess difference in death probability between the two groups. Cox-Regression analysis was performed to assess covariates.*

To expand our analysis, we investigated whether there is a difference in HOXA5 prognostic value between different TNBC subtypes by assessing its effects on patients OS and RFS. Kaplan-Meier analysis showed that BL-1 patients expressing high HOXA5 mRNA levels have a higher OS (n=103, median, OS: 103 vs 103-month, *p=0.058*). The estimated HR indicated that increased HOXA5 expression is accompanied by a 24% lower risk of death within the tested population (HR= 0.76, 95% CI 0.21 to 1.05) (Figure. 9A). For the BL-2 TNBC, HOXA5 expression status resulted in no significant difference in OS among the tested groups (n=58, median, OS: 202 vs 202-month, *p=0.32*) (HR= 0.6, 95% CI 0.22 to 1.66) (Figure. 9A).

Similarly, HOXA5 expression levels resulted in no significant difference in OS among the tested M and MSL. For the M subtype (n=114, median, OS: 189 vs 189-month, *p=0.86*) (HR= 0.94, 95% CI 0.49 to 1.82). For the MSL subtype (n=39, median, OS: 240 vs 240-month, *p=0.91*) (HR= 0.94, 95% CI 0.34 to 2.6) (Figure. 9B). Moreover, no significant difference in RFS was noted for different HOXA5 expression levels in the different TNBC subtypes (data not shown).

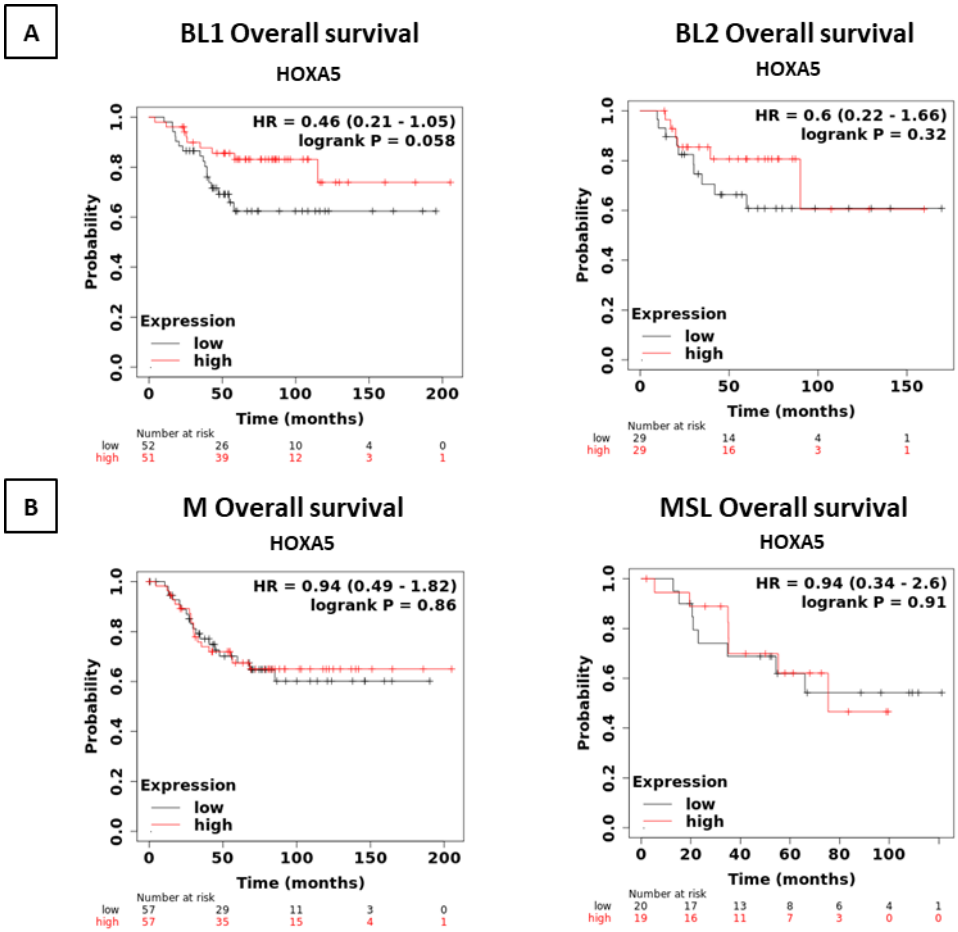

Figure 9: Kaplan- Meier Survival curves based on HOXA5 expression status in different TNBC subtypes*. Black line indicates patients with low HOXA5 expression; red line indicates patients with high HOXA5 expression.* ***A.*** *Cumulative survival curves of Basal Like-1 (BL-1) and Basal Like-2 (BL-2) TNBC patients.* ***B.*** *Cumulative survival curves of Mesenchymal (M) and Mesenchymal-Stem Like (MSL) Kaplan-Meier method was used to estimate survival curve, and Long-rank test was used to assess difference in death probability between the two groups. Cox-Regression analysis was performed to assess covariates.*

As part of assessing HOXA5 prognostic value, we also used Kaplan-Meier analysis from the Nottingham Cohort composed of 112 TNBC patient's samples to compare the HOXA5 expression state effects on patients' distant metastatic free survival time (DMFS) overall and for 5 years' time. The results suggested that high HOXA5 expression is significantly accompanied by low DMFS compared to low HOXA5 expression (*p=0.035*) (Figure. 10A). This was also noted at 5 years survival analysis (*p=0.017*) (Figure. 10B). The OS survival analysis for the Nottingham BC cohort showed no difference in survival probability in relevance to HOXA5 expression state (data not shown)

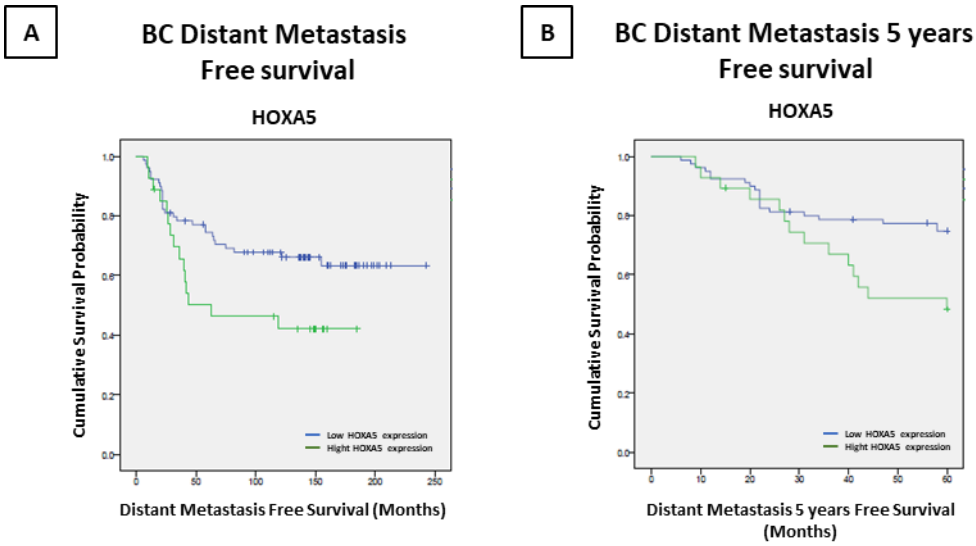

Figure 10: Kaplan-Meier Distant Metastasis Free Survival (DMFS) curves based on HOXA5 expression status in Breast Cancer (BC) *(blue line indicate patients with low HOXA5 expression; green line indicates patients with high HOXA5 expression).* ***A.*** *Cumulative survival curves of BC patients with higher HOXA5 expression displayed low DMFS compared to the low expressing HOXA5 group (p=0.035).* ***B.*** *Cumulative survival curves of BC patients with higher HOXA5 expression displayed a lower distal metastasis 5-years survival probability compared to patients expressing lower HOXA5 expression (p=0.017). Kaplan-Mier method was used to estimate survival curve, and Long-rank test was used to assess difference in death probability between the two groups.*

5.3 Generation and Characterisation of HOXA5 expressing TNBC cell lines

We found interesting the upregulation of HOXA5 in TNBC compared to other BC subtypes. Especially with its expression effects on TNBC survival. This suggested a possible relevant role for HOXA5 in TNBC and possibly between the different subtypes. Therefore, we decided to study HOXA5 further and test the effects of its overexpression and knockdown in different human TNBC subtypes cell lines.

5.3.1 Generation of HOXA5 overexpressing TNBC cells

To investigate the effects of HOXA5 upregulation in the MSL and the BL-1 TNBC subtypes. We generated HOXA5-GFP overexpressing MDA-MB-231 and MDA-MB-468 cells using a lentivirus vector. The selection of the cell lines was based on their low HOXA5 expression levels (Figures. 6C). Successful HOXA5-GFP expression was assessed at this point to ensure infection efficiency through fluorescence imaging of the co-expressed GFP reporter (Figure. 11A). For the selection of a pure HOXA5 expressing population, cells were sorted according to GFP expression using Astrios Cells sorter (Figure. 11B). Microscopic analysis of the infected cells shows that, upon HOXA5 expression, no apparent phenotypic changes were noted on the MDA-MB-231 cells (Figure. 10A). However, the MDA-MB-468 cells appeared more rounded compared to the wild-type (Figure. 10B).

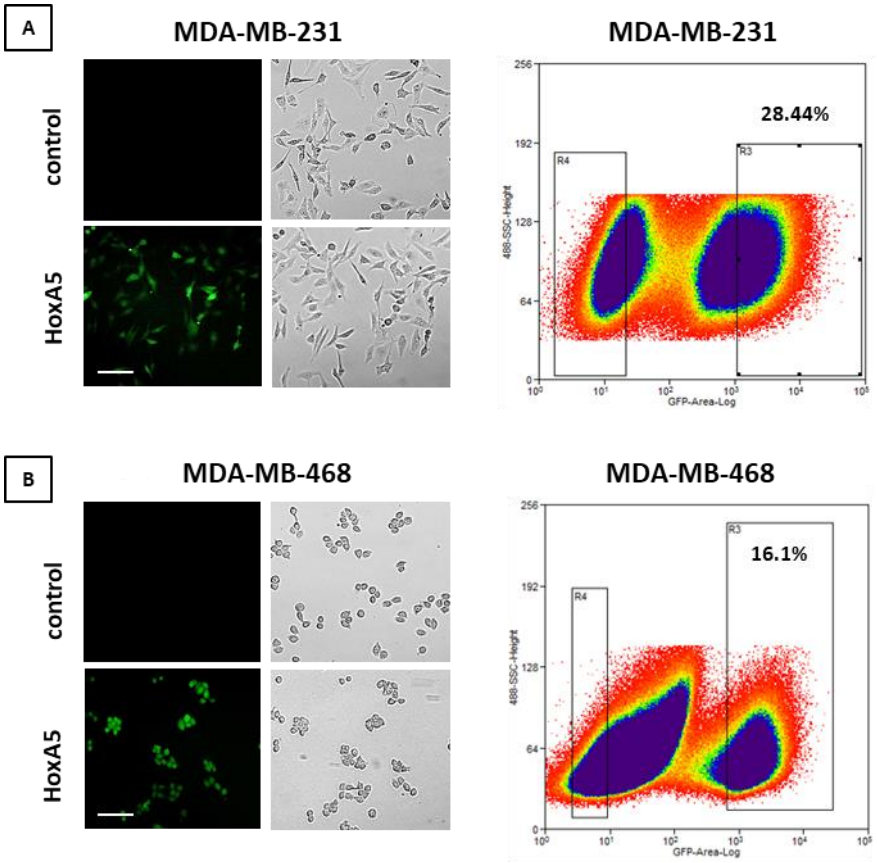

Figure 11: Generation of HOXA5 overexpressing TNBC cells: A. *The mesenchymal stem like TNBC cells MDA-MB-231 HOXA5-GFP expressing cells fluorecent (left) and contrast phase (right) images of the control and infected cells. On the right, Astrios Cell sorter gates showing the GFP positive population (28.44%) and the negative control.* ***B.*** *The Basal like-1TNBC cells MDA-MB-468 HOXA5-GFP expressing cells fluorecent (left) and contrast phase images (right) of the control and infected cells. On the right, Astros Cell sorter gates showing the GFP positive population (16.1%) and the negative control.*

5.3.2 Characterization of HOXA5 expressing cells: gene expression and protein markers

To investigate the differential effects of HOXA5 expression between the MSL and BL-1 TNBC subtypes, we explored the expression of different markers that are known to be involved in important biological processes regulating self-renewal, differentiation, proliferation, and others through quantifying their mRNA expression using RT-qPCR.

The MDA-MB-231-HOXA5 cells significantly expressed lower levels of the adhesion and phenotypic markers CDH1 and KRT14 when compared to the wild-type (*p=0.0033*, and *p=0.0033*, respectively) (Figure. 12A).

Moreover, no significant difference was seen in the levels of the marker of proliferation MKI67 (*p=0.7473*) (Figure. 12B). Similarly, no difference was seen in the expression of the Wnt pathway target gene axis inhibition protein 2 (AXIN2) (p=0*.0904*) (Figure. 12D). HOXA5 is known to counteract stemness through modulating genes that are involved in the Wnt pathway in colorectal cancer[82].

Additionally, HOXA5 expression resulted in different effects on the ALDH1 family markers; Aldehyde Dehydrogenase 1 Family Member A1 (ALDH1A1) and Aldehyde Dehydrogenase 1 Family Member A3 (ALDH1A3), were a significant upregulation in

ALDH1A1 and downregulation in ALDH1A3 levels were noted (*p=0.0002*, and *p=0.0363*, respectively) (Figure. 12C). The ALDH1 family is considered an independent prognostic indicator and important maker and regulator of stemness and proliferation in TNBC[91,92]. The data propose a regulatory role of HOXA5 for the ALDH1 family members ALDH1A1 and ALDH1A3.

A link has been previously identified between retinoids treatment and the yes-associated protein 1 (YAP) gene activation[93]. Retinoids are known to induce HOXA5 and its upstream regulator RAR β in breast cancer[82,85,94]. Therefore, we explored HOXA5 expression effects on important YAP target genes. The MDA-MB-231-HOXA5 cells showed a significant reduction in the mRNA levels of the tested YAP target genes; Tumour Associated Calcium Signal Transducer 2 (TACSTD2), Connective tissue growth factor (CTGF), and Cysteine-rich angiogenic inducer 61 (CYR61) (*p=0.0241*, *p=0.0070* and *p=0.0033*, respectively) (Figure. 12E). Data suggests a strong link between HOXA5 and YAP target genes regulation, implying a negative correlation between the two pathways.

Moreover, similarly, we tested HOXA5 effects on these sets of genes in the MDA-MB-468. The MDA-MB-468-HOXA5 cells also significantly expressed lower levels of the CDH1 compared to the wild-type. Contrarily, higher levels of KRT14 levels accompanied HOXA5 expression (*p<0.0001*, and *p=0.0030*, respectively) (Figure. 13A). Also, a significant reduction of the marker of proliferation MKI67 was noted (*p<0.0001*) (Figure. 13B). Furthermore, opposite to the MDA-MB-231, HOXA5 expression in the MDA-MB-468 reduced ALDH1A1 (*p=0.0004*) (Figure. 13C). While no difference in ALDH1A3 expression was seen (*p=0.3011*) (Figure. 13C). Moreover, a significant upregulation of the Wnt target gene AXIN2 was noted in the MDA-MB-468-HOXA5 cells (*p=0.0232*) (Figure. 13D). Significant changes were also noted in the YAP target genes, where TACSTD2 was upregulated (*p=0.0001*), and CTGF was downregulated (*p=0.0029*). However, there was no change in CYR61 levels between the tested cell lines (p=0.5859) (Figure. 13E).

To further investigate the differential effects of HOXA5 expression between the TNBC subtypes and to assess the relevance of our results on protein expression, we compared the effects of HOXA5 on the expression/localization of a set of proteins in MDA-MB-231 and MDA-MB-468 using fluorescent imaging. In the MDA-MB-231-HOXA5 cells, a higher expression of the CDH1 was noted compared to the wild-type. However, CDH1 did not have membrane localization, so we think this antibody might be not specific (Figure. 14A). While no apparent difference was seen of cytokeratin's were there was no change in pan-cytokeratin expression/localization between the HOXA5 cells and the control (Figure. 14B). Also, there was no difference in staining intensity or change in localization of β-catenin or NFκB/p65 protein between the tested cell lines; however, p65 was noted to be cytoplasmic in both (Figure. 14D/14F). Also, no apparent difference in staining intensity of the proliferation marker MKI67 was seen between the tested cell lines (Figure. 13C). Moreover, a slightly lower intensity and nuclear localisation of YAP was seen within the HOXA5 cells compared to the wild-type (Figure. 13E).

For the MDA-MB-468-HOXA5 cells, no apparent difference was seen in the expression of CDH1 and pan-cytokeratin compared to the wild-type (Figure. 15A/15B). There was also no difference in the staining intensity and localization of β-catenin, YAP, and NFκB/p65 between the tested cell lines. p65 appeared to be cytoplasmic in both (Figure. 15D/15E/15F). Moreover, an apparent reduction in staining intensity of

the proliferation marker Ki67 was seen in the MDA-MB-468-HOXA5 cells (Figure. 15C).

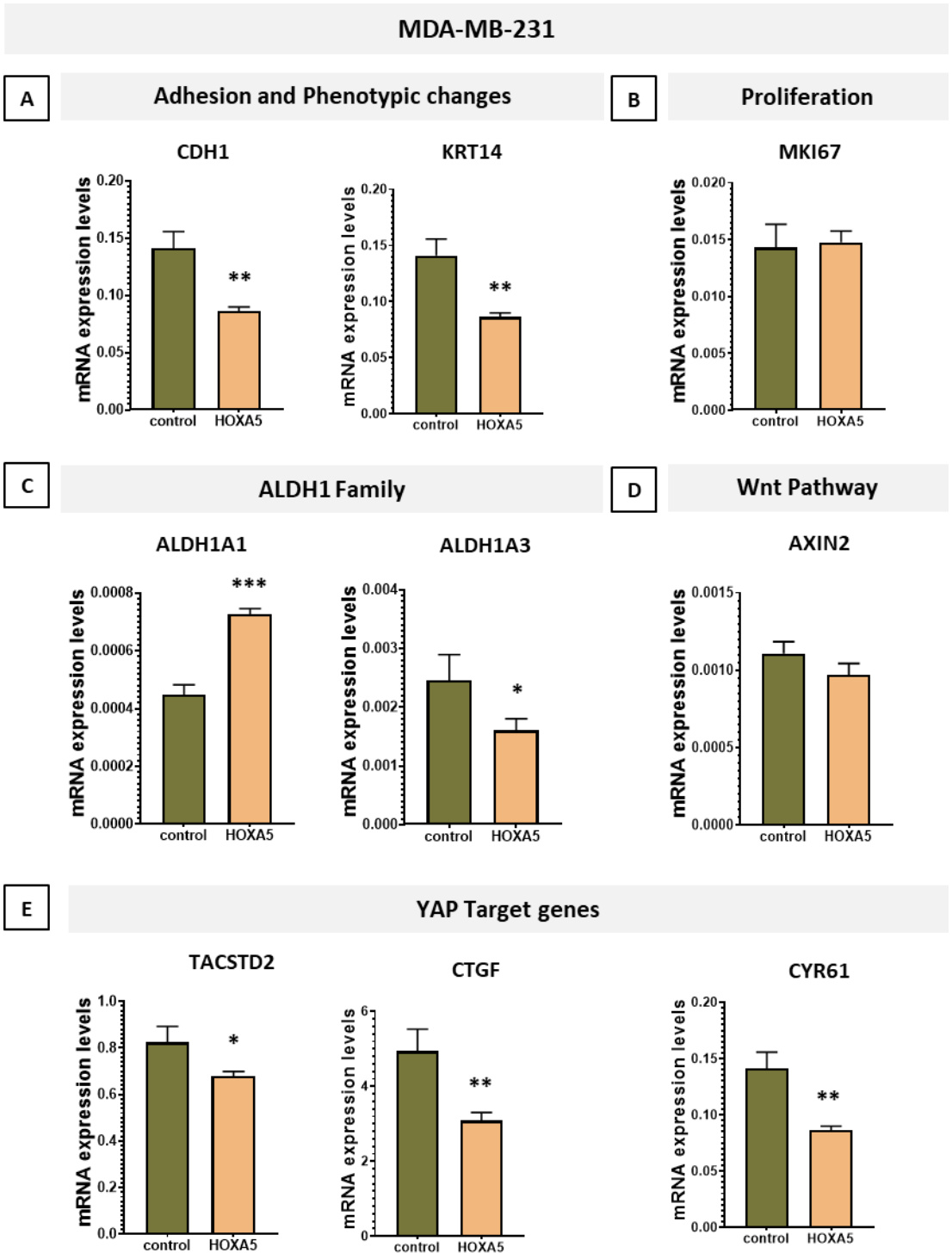

Figure 12: The effects of HOXA5 expression on gene expression in the MSL TNBC cell line MDA-MB-231. *HOXA5 expression effect on the mRNA expression of A. the phenotypic and adhesion markers CDH1 and KRT14.* ***B.*** *the marker of proliferation MKI67.* ***C.*** *the ALDH1 family members ALDH1A1 and ALDH1A3.* ***D.*** *the Wnt pathway target gene AXIN2.* ***E.*** *the YAP target genes; TACSTD2, CTGF, and CYR61. RNA samples were analysed using qPCR. n=3, Bars indicate averages ± standard errors. To assess expression variation two-tailed unpaired student's t-test was used, * p <0.05, ** p<0.01, *** p<0.001.*

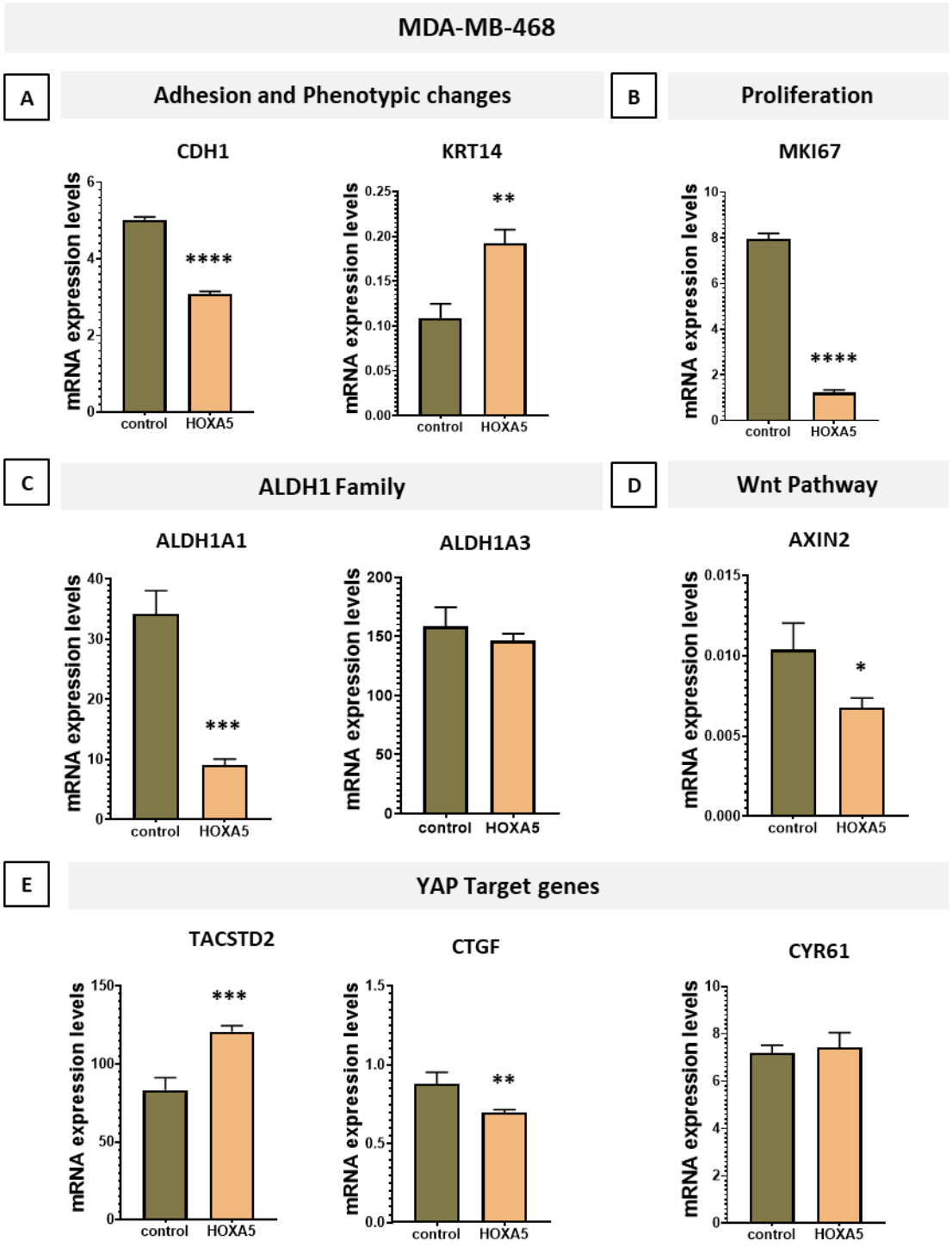

Figure 13: The effects of HOXA5 expression on gene expression in the BL-1 TNBC cell line MDA-MB-468. *HOXA5 expression effect on the mRNA expression of A. the phenotypic and adhesion markers CDH1 and KRT14.* ***B.*** *the marker of proliferation MKI67.* ***C.*** *the ALDH1 family members ALDH1A1 and ALDH1A3.* ***D.*** *the Wnt pathway target gene AXIN2.* ***E.*** *the YAP target genes; TACSTD2, CTGF, and CYR61. RNA samples were analysed using qPCR. n=3, Bars indicate averages ± standard errors. To assess expression variation two-tailed unpaired student's t-test was used, * p <0.05, ** p<0.01, *** p<0.001.*

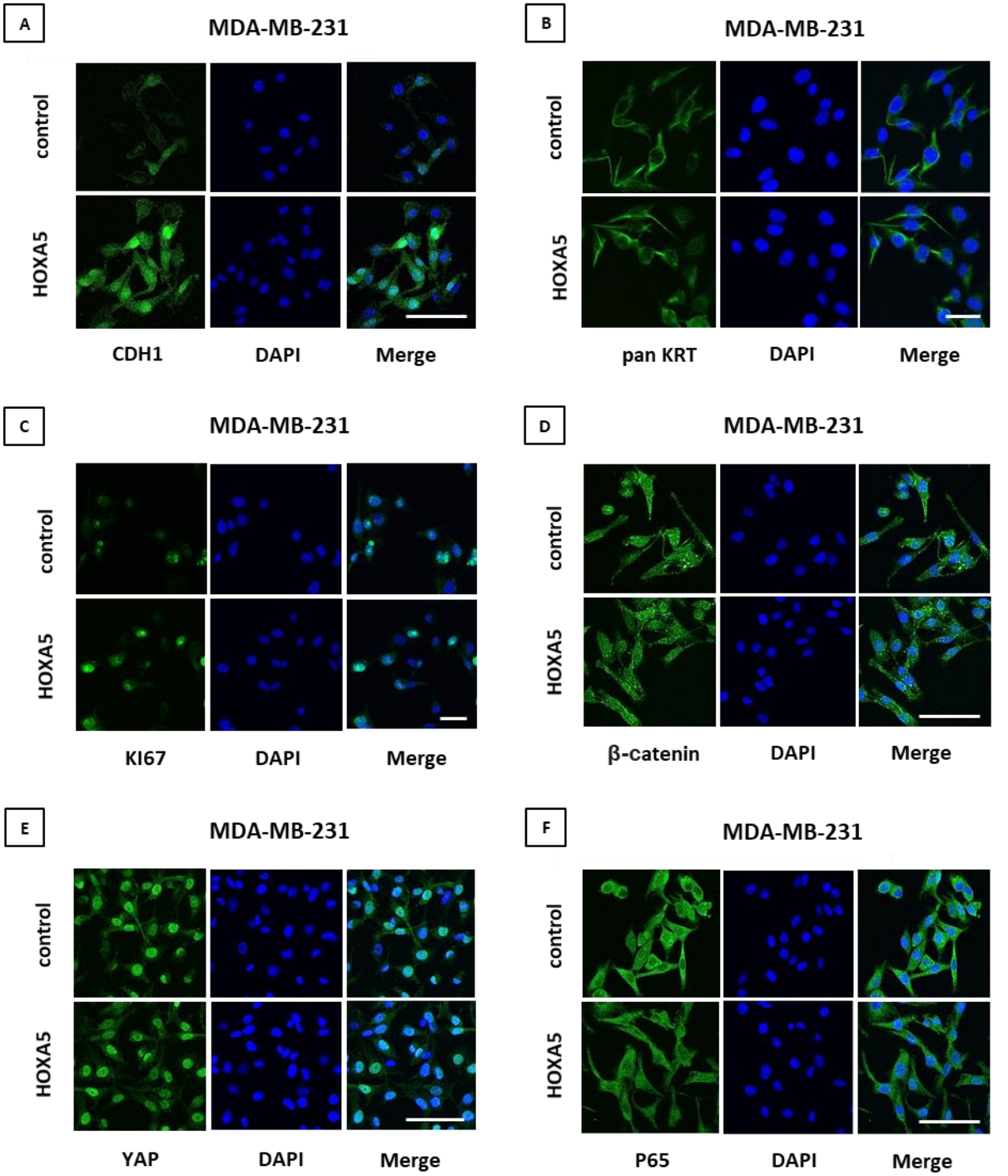

Figure 14: HOXA5 expression effects on protein expression/localization in the MSL TNBC cells MDA-MB-231. *HOXA5 effect on expression of* ***A.*** *CDH1, Scale bar = 74.1 µm.* ***B.*** *Cyto-keratins, scale bar = 36.6 µm.* ***C.*** *MKI67, scale bar = 36.6 µm.* ***D.*** *β-catenin and its localization, scale bar = 74.1 µm.* ***E.*** *YAP and its localization, scale bar = 74.1 µm.* ***F.*** *p65, scale bar = 74.1 µm.*

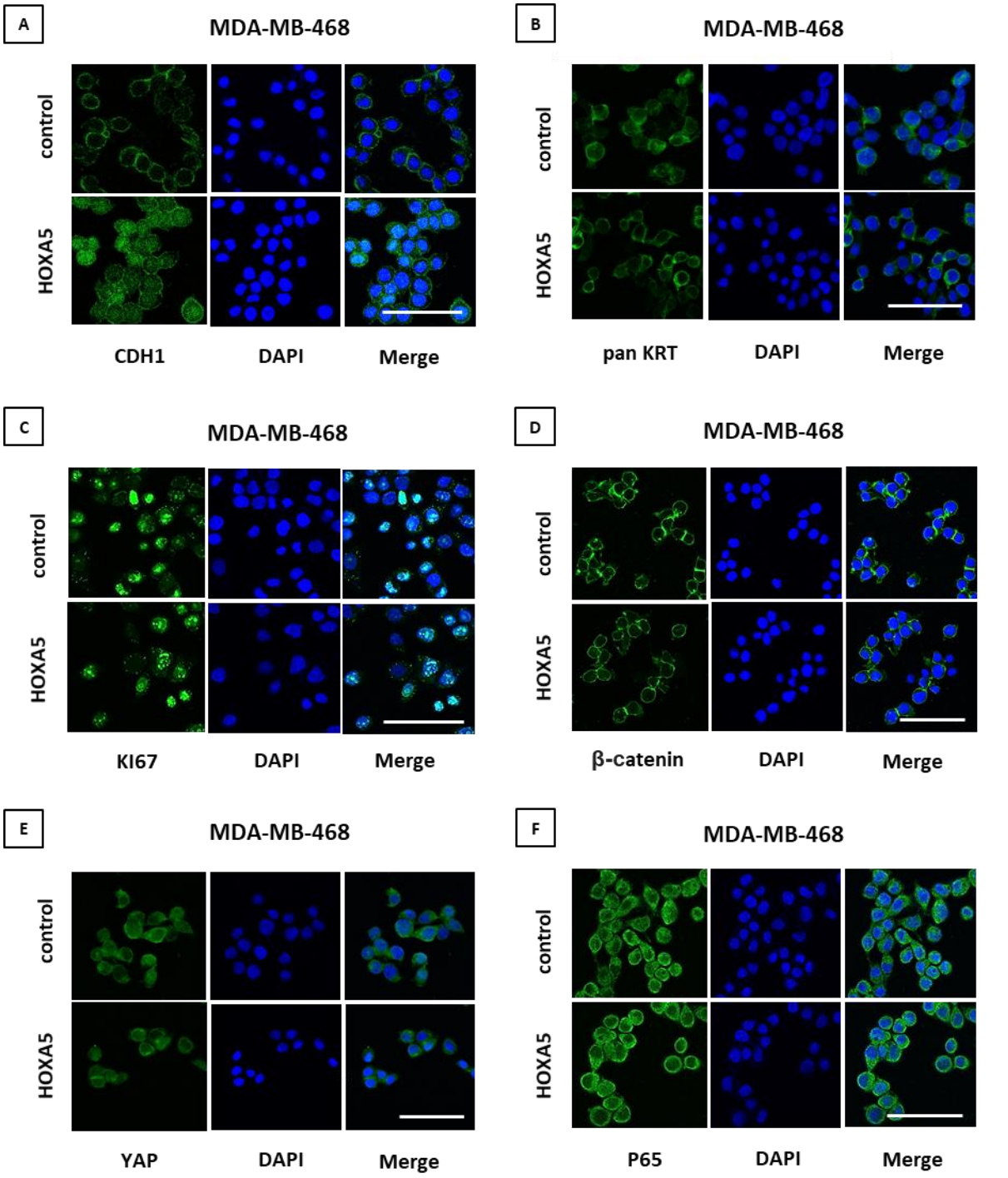

Figure 15: HOXA5 expression effects on protein expression and localization in the BL-1 TNBC cells MDA-MB-468. *HOXA5 effect on expression of **A**. CDH1. **B**. Cyto-keratins. **C**. KI67. **D**. β-catenin and its localization. **E**. YAP and its localization, scale bar. **F**. p65. Scale bar = 74.1 µm.*

To further assess the effects of HOXA5 expression in the MSL TNBC subtypes, we compared protein expression in MDA-MB-231-HOXA5 tumour xenografts using fluorescent imaging. Successful HOXA5-GFP expression is noted in the MDA-MB-231-HOXA5 xenografts (Figure. 16A). The MDA-MB-231-HOXA5 tumours were noted to express higher KI67 levels compared to the wild-type (Figure. 16B). No apparent difference was seen in the expression of NFκB/p65 between the HOXA5 expressing tumour tissue and the wild-type (Figure. 16C). Similarly, no evident difference in YAP staining intensity was seen in response to HOXA5 expression. However, YAP was noted to be localized in the cytoplasm in both tested tissues (Figure. 16D).

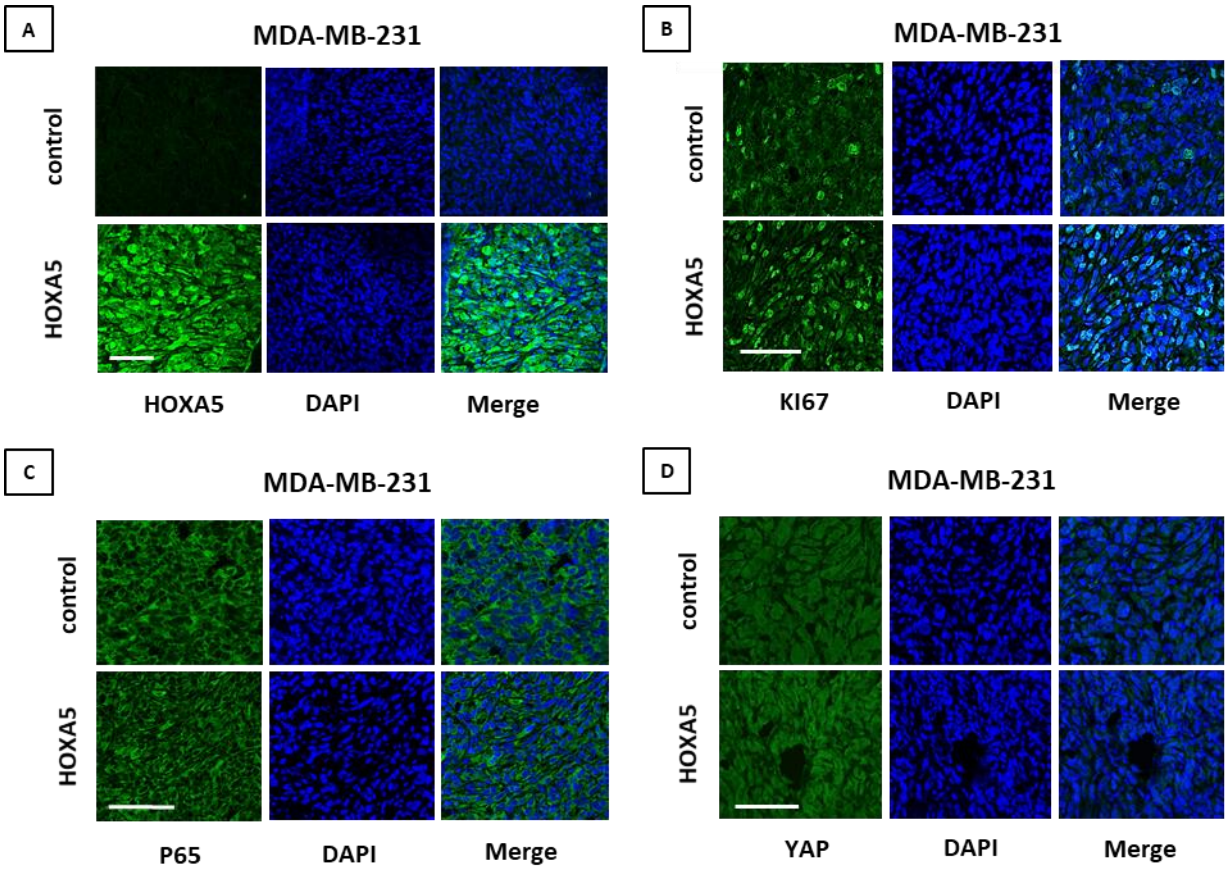

Figure 16: The effects of HOXA5 expression on the MSL TNBC MDA-MB-231 xenografts protein expression. ***A.*** *HOXA5-GFP expression in MDA-MB-231 xenografts.* ***B.*** *HOXA5 expression reduced the intensity of KI67 staining.* ***C.*** *No difference in NFκB/p65 staining and localization was noted following HOXA5 expression.* ***D.*** *No difference in YAP staining and localization was noted between the control and MDA-MB-231-HOXA5. YAP was noted to be localized in the cytoplasm. Scale bar = 73.1 µm.*

To further assess the MDA-MB-231-HOXA5 xenografts, we performed a microscopic examination of the H&E stained tumour section with the assistance of the pathologist. The MDA-MB-231-HOXA5 tumours appeared more proliferative, were it showed a higher number of mitotic events/figures relative to the controls (Figure.17). Moreover, the control tumours margins were noted to be well-circumscribed compared to the HOXA5 tumour margin, which appeared to be speculated. The MDA-MB-231-HOXA5 cells invasion of the surrounding tissue was apparent (Images not included). Dr Ordonez-Moran provided H&E-stained tumour sections slides.

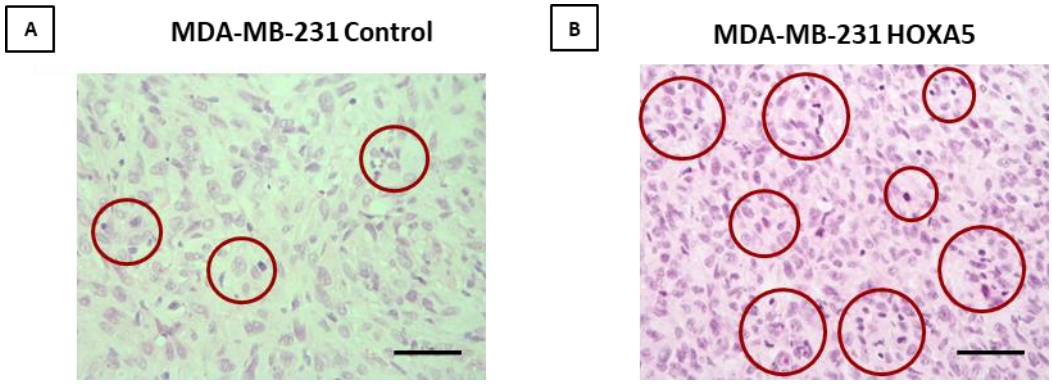

Figure 17: MDA-MB-231-HOXA5 xenografts sections. *Mitotic figures/events marked with red.*

5.3.3 The effects of HOXA5 over-expression on TNBC cells clonogenicity

To assess HOXA5 contribution to cellular survival and proliferation, we performed a colony formation assay using the MDA-MB-231 and MBA-MB-468 HOXA5 expressing cells. HOXA5 expressing cells demonstrated a significant reduction in colony numbers and colony size compared to the wildtype in the MDA-MB-231 (*p=0.0009*) (Figure. 18A). Contrarily, MDA-MB-468 cells showed a significant increase in colony numbers compared to the wild-type (*p=0.0046*) (Figure. 18B).

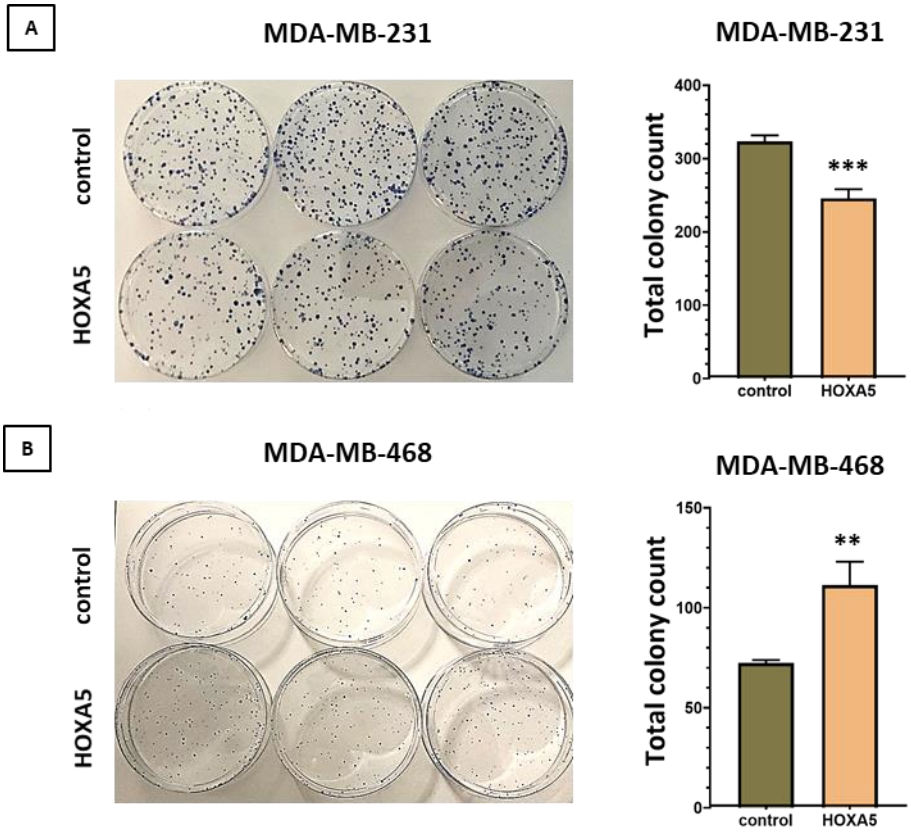

Figure 18: HOXA5 expression modulates TNBC MSL and BL-1 clonogenicity. *On the left a representative image of the colonies. On the right the bar-charts quantifying colony counts.* ***A.*** *HOXA5 expression reduce colony numbers in the MSL TNBC cells MDA-MB-231.* ***B.*** *HOXA5 expression induces colony counts in the BL-1 TNBC cells MDA-MB-468. To assess the significance of colony numbers variation two-tailed unpaired student's t-test was used, * p <0.05, ** p<0.01, *** p<0.001.*

5.3.4 The effects of HOXA5 expression on TNBC cells mammosphere formation ability

To evaluate the effect of HOXA5 on CSCs self-renewal in the different TNBC subtypes, using MDA-MB-231-HOXA5 and MDA-MB-468-HOXA5 cells, we conducted a mammosphere formation assay and performed serial passaging of the generated mammospheres at clonal density. The enhanced CSC population induced by HOXA5 expression resulted in a significant increase in MDA-MB-231-HOXA5 and MDA-MB-468-HOXA5 colony count (*p=0.0422,* and *p=0.0111,* respectively) and mammosphere cell count in the third generation (*p=0.0286,* and *p=0.0024,* respectively) compared to the wild-type (Figure. 19). Results suggest that HOXA5 expression results in enrichment in the stem cell population of TNBC.

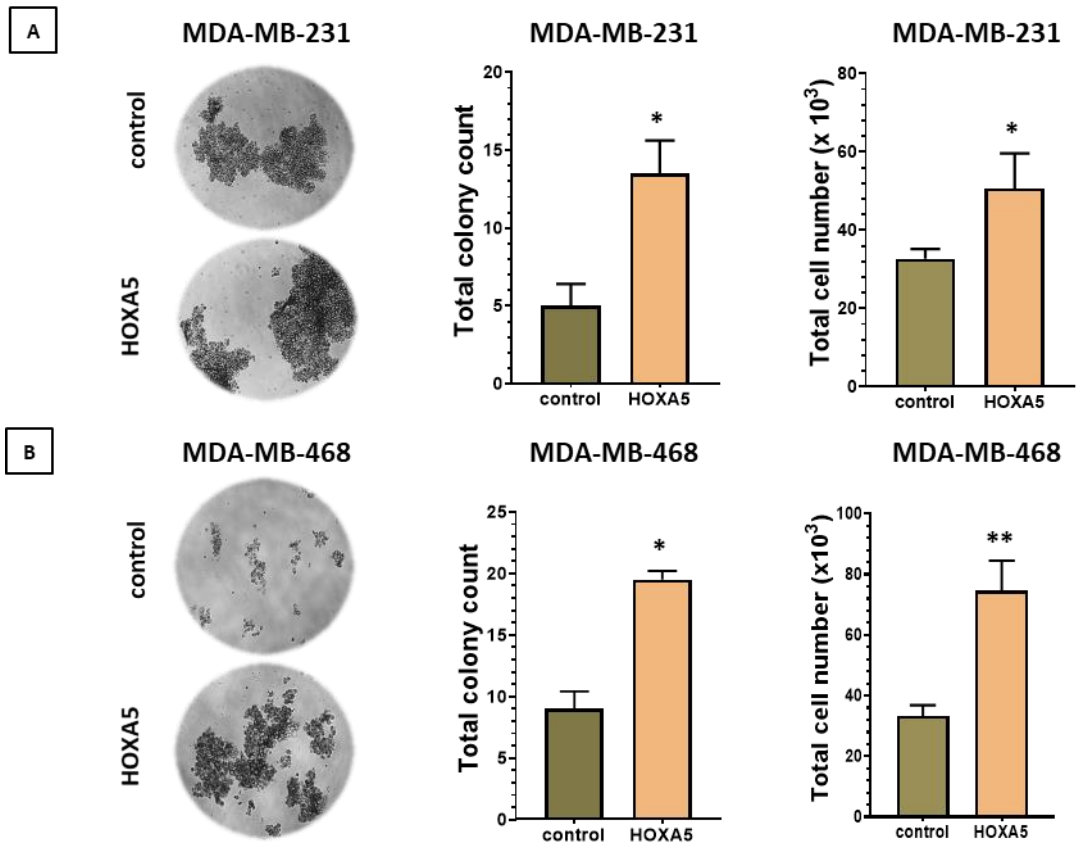

Figure 19: HOXA5 expression induces CSC self-renewal and anchorage independent cell growth in the MSL and BL-1 TNBC Mammospheres*. HOXA5 expression increased number and size of colonies and cell count of both the* ***A.*** *MSL MDA-MB-231 and* ***B.*** *BL-1 MDA-MB-468 cells. On the left a representative image of the colonies. On the right the bar-chart shows increased number of colonies and increased number of cells in the HOXA5 expressing cell lines compared to the wild-type. The number of colonies were calculated by counting 4X fields of view. The number of colonies were calculated by counting 4X fields of view. n = 3, Bars indicate averages ± standard errors. To assess numbers variation two-tailed unpaired student's t-test was used, * p <0.05, ** p<0.01, *** p<0.001.*

5.3.5 The effects of HOXA5 expression on TNBC cells migration

To investigate whether HOXA5 has an effect on cellular migration, we performed the wound healing assay (scratch test) using the MDA-MB-231-HOXA5 and MDA-MB-468-HOXA5 cells. Relative migration distance for the MDA-MB-231 cells was measured at 0 hours, 16 hours, 24 hours (Figure. 20A). Relative migration distance for the MDA-MB-468 cells was measured at 0 hours, 24 hours, and 48 hours (Figure. 20B). The migration assay showed HOXA5 expression significantly increase cell migration speed in both MDA-MB-231 and MDA-MB-468 cell lines. The relative migration distance of cells of the MDA-MB-231-HOXA5 was significantly narrow relative to the wild-type (*p=0.0001*) (Figure. 20A). The relative migration distance of cells of the MDA-MB-468-HOXA5 was also significantly narrow relative to the wild-type (*p=0.0004*) (Figure. 20B). Our results suggest that HOXA5 expression significantly induce cellular migration ability in TNBC cells.

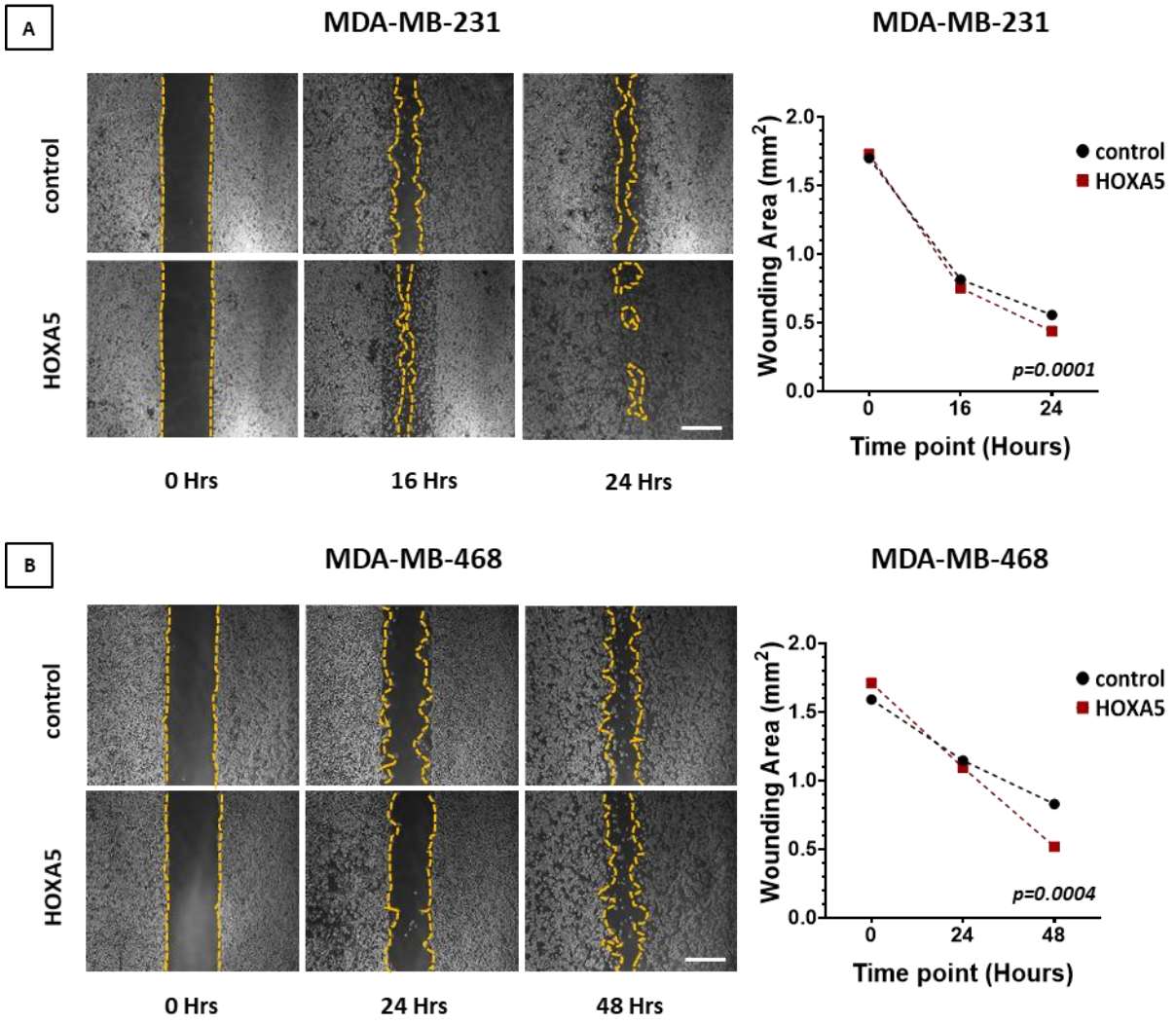

Figure 20: HOXA5 expression induced cellular migration in MSL and BL-1 TNBC cells. *On the left, representative bright-field images comparing wound closure in the HOXA5 expressing cells and the wildtype at different time points. On the right wound closure expressed as the remaining area uncovered by the cells. The black lines represent the control remaining area uncovered by the cell; red line represents HOXA5 expressing cells remaining area uncovered by the cell.* ***A.*** *Migration assay in the MSL TNBC cell line MDA-MB-231.* ***B.*** *Migration assay in the BL-1 TNBC cell line MDA-MB-468. Image taken using Nikon Eclipse Ti2 Widefield Fluorescent Microscope. Scale bar = 500 µm.*

5.4 Generation and Characterisation of HOXA5 Knockdown TNBC cell lines

5.4.1 Generation of HOXA5 Knockdown TNBC cells

To further investigate the effects of HOXA5 on TNBC cell lines, we knocked down HOXA5 expression in the M TNBC cell line BT549. The BT549 cells were previously noted to express high levels of HOXA5 (Figure. 6C). A lentivirus vector was used to knockdown HOXA5. Successful HOXA5 knockdown was assessed through fluorescence imaging of the co-expressed RFP (Figure. 21A). For the selection of a pure HOXA5 KD population, cells were sorted according to RFP expression using Astrios Cells sorter (Figure. 21B). HOXA5 mRNA levels were quantified using RT-qPCR analysis (Figure. 21A). HOXA5 levels were significantly reduced in the infected cell line compared to the wild-type. Microscopic analysis shows that, following HOXA5 knockdown, no apparent phenotypic changes were noted on the BT549 cells (Figure. 21A).

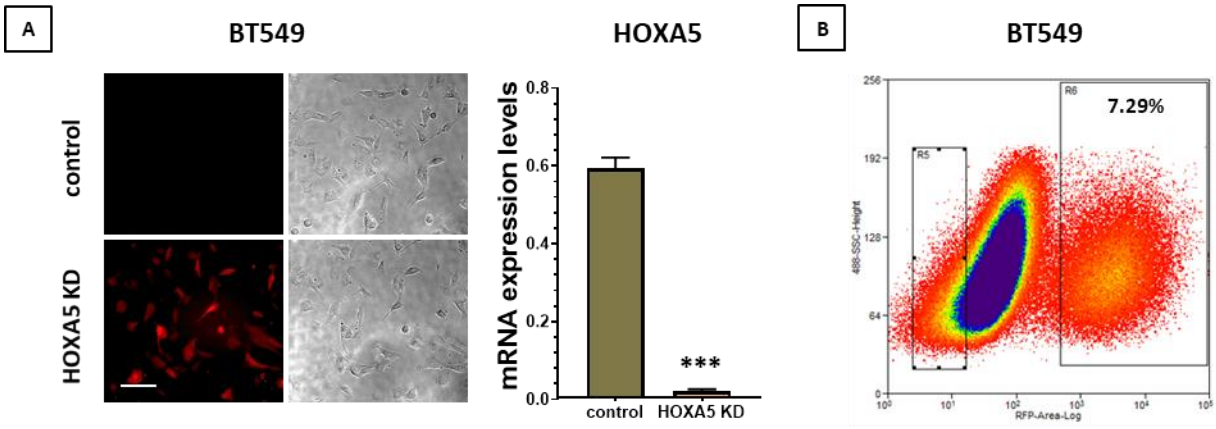

Figure 21: Generation of HOXA5 Knockdown TNBC cells: ***A.*** *The mesenchymal TNBC cells BT549 HOXA5-RFP expressing cells fluorecent (left) and contrast phase (right) images of the control and infected cells. Bars show HOXA5 mRNA expression quantification using RT-qPCR. n=3, Bars indicate averages ± standard errors. To assess expression variation two-tailed unpaired student's t-test was used, * p <0.05, ** p<0.01, *** p<0.001.* ***B.*** *Astrios Cell sorter gates showing the RFP positive population (7.29%) and the negative control.*

5.4.2 Characterization of HOXA5 knockdown cells: gene expression and protein markers

Similar to what we have done in the HOXA5 expressing cells, we investigate the effects of HOXA5 knockdown in the BT549 cells through quantifying the mRNA expression of a set of genes using RT-qPCR.

The results from the MDA-MB-231 cells and the BT549 can be related as the cell lines are classified under MSL and M TNBC subtypes, respectively. According to *Lehmann et.* both subtypes share many mesenchymal cells and metaplastic tumours GE characteristics, and therefore, effects in both lines could be linked[15]. In line with HOXA5 expression data in the MDA-MB-231, the BT549-HOXA5 KD cells significantly expressed high levels of CDH1 compared to the wild-type. This confirmed the suggested HOXA5 negative regulation of CDH1 in TNBC. However, no difference in the expression of KRT14 was noted following HOXA5 KD (*p=0.0009*, and *p=0.2399*, respectively) (Figure. 22A). Also, no difference was seen in the levels of the marker of

proliferation MKI67 (*p=0.0795*) or the expression of the Wnt pathway target gene AXIN2 (p=0*.4671*) (Figure. 22B/22D).

Moreover, HOXA5 KD resulted in different effects on ALDH1 family members, where a significant downregulation of ALDH1A1 was noted, and no difference in the expression of ALDH1A3 was seen compared to the wild-type (*p=0.0023*, and *p=0.1061*, respectively) (Figure. 22C). This can be linked to the suggested regulatory role of HOXA5 for ALDH1A1 in MDA-MB-231 cells (Figure. 12C), supporting that HOXA5 expression upregulates ALDH1A1 levels in TNBC.

Furthermore, the HOXA5 KD effect on YAP target genes was further tested. A significant upregulation in the mRNA levels of the tested YAP target genes; TACSTD2 and CTGF, were noted compared to the control (*p<0.0001* and *p=0.0178*, respectively). On the other hand, no difference was recorded in the expression of the YAP target gene CYR61 (*p=0.2731*) (Figure. 22E). Relating this to the MDA-MB-231 data (Figure. 12E) could support the proposed negative regulation between HOXA5 and YAP target genes regulation.

To further investigate the effects of HOXA5 KD in the BT549 cells, we also explored the effects of HOXA5 KD on the expression of a set of proteins using fluorescent imaging. HOXA5 KD did not affect the staining intensity of the proliferation marker KI67 (Figure. 23A). No apparent difference in YAP localization was noted between the tested cell lines, where nuclear localization of YAP was seen in both (Figure. 23B). Moreover, high staining intensity of NFκB/p65 was seen in the wild-type. NFκB/p65 appeared to be localized in the nucleus (Figure. 23C).

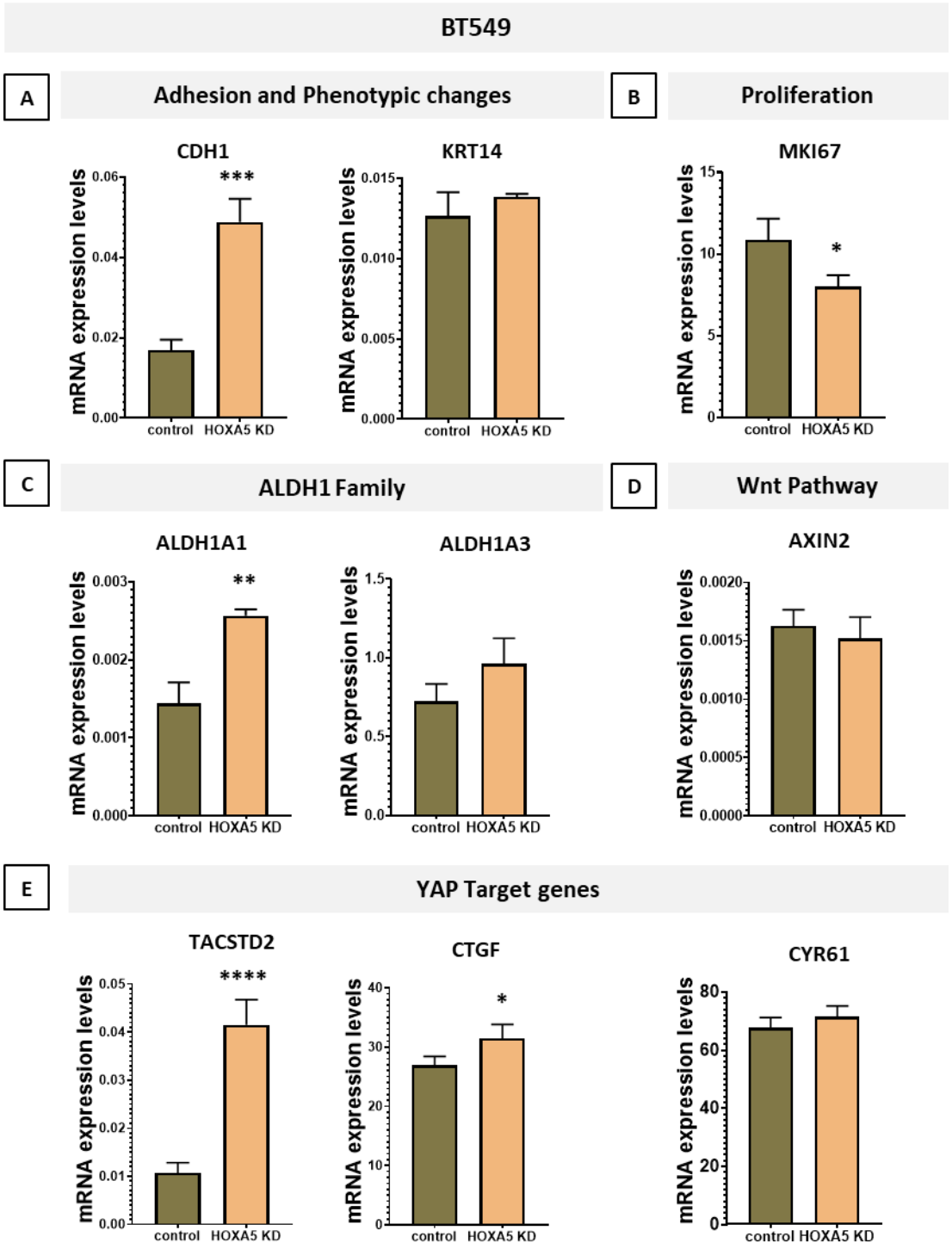

Figure 22: The effects of HOXA5 knockdown on gene expression in the M TNBC cell line BT549. *HOXA5 expression effect on the mRNA expression of A. the phenotypic and adhesion markers CDH1 and KRT14.* ***B.*** *the marker of proliferation MKI67.* ***C.*** *the ALDH1 family members ALDH1A1 and ALDH1A3.* ***D.*** *the Wnt pathway target gene AXIN2.* ***E.*** *the YAP target genes; TACSTD2, CTGF, and CYR61. RNA samples were analysed using qPCR. n=3, Bars indicate averages ± standard errors. To assess expression variation two-tailed unpaired student's t-test was used, * p <0.05, ** p<0.01, *** p<0.001.2R*

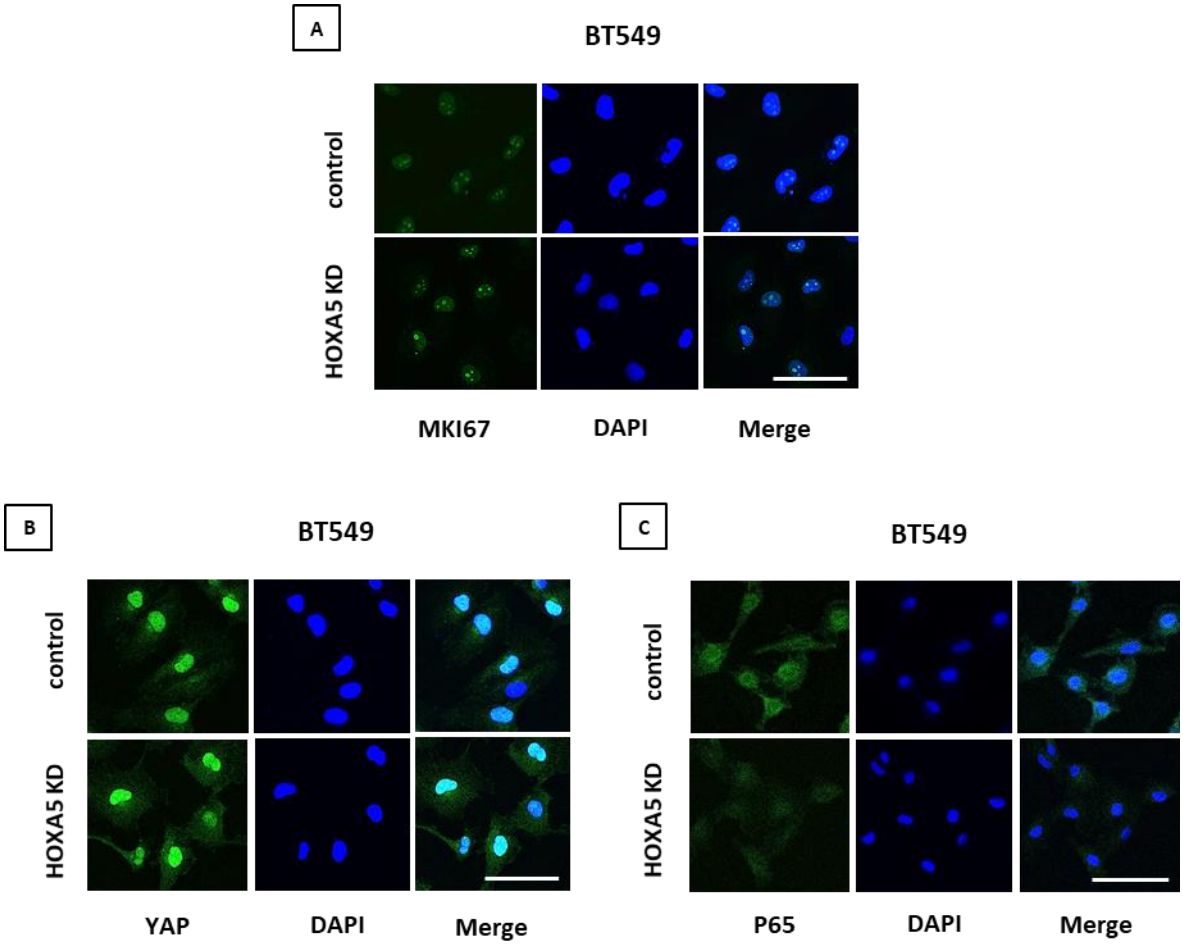

Figure 23: HOXA5 knockdown effects on protein expression and localization in M TNBC cell line BT549. *HOXA5 KD effect on the expression of* ***A.*** *KI67.* ***B.*** *YAP and its localization.* ***C.*** *p65 and its localization, scale bar = 74.1 µm.*

5.4.3 The effects of HOXA5 Knockdown on TNBC cells clonogenicity

We further assessed HOXA5 effects on clonogenicity and cellular survival through performing the clonogenic assay using the HOXA5 KD cells. HOXA5 KD produced a high number of large colonies in BT549 cells compared to the wild-tyype (*p<0.0001*) (Figure. 24). This finding supports HOXA5 overexpression data in the MDA-MB-231 cells, were its expression reduced colony numbers and size in the MSL cells (Figure. 18A).

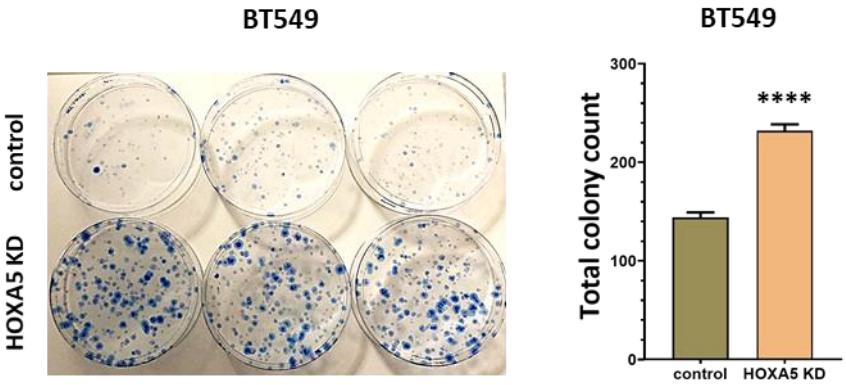

Figure 24: HOXA5 Knockdown induce the M TNBC cells BT549 clonogenicity and cell survival. *On the left a representative image of the colonies. On the right the bar-charts quantifying colony counts showing an increase in colony numbers of the BT549 cells. To assess the significance of colony numbers variation two-tailed unpaired student's t-test was used, * p <0.05, ** p<0.01, *** p<0.001.*

5.4.4 The effects of HOXA5 Knockdown on TNBC cells mammosphere formation ability

Our results previously suggested that HOXA5 expression could enhance the stem cell population of TNBC (Figure. 19). Unlike with HOXA5 overexpression, HOXA5 KD in the BT549 cells did not affect mammospheres colony count (*p=0.1679*) higher mammosphere cell count (*p=0.8906*) (Figure. 25)

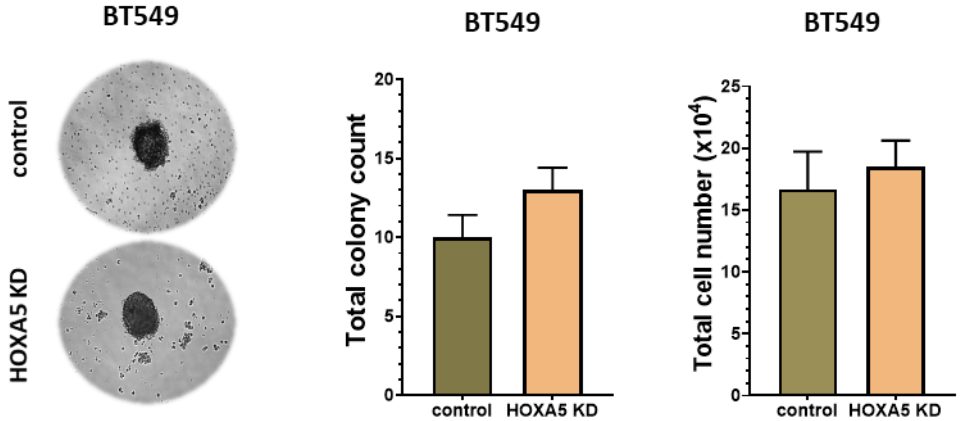

Figure 25: HOXA5 knockdown dose not effect CSC self-renewal and anchorage independent cell growth in the M TNBC Mammospheres. *On the left a representative image of the colonies. On the right the bar-chart HOXA5 knockdown had no effect on colony size, numbers, or cell number compared to the wild-type. The number of colonies were calculated by counting 4X fields of view. n = 3, Bars indicate averages ± standard errors. To assess numbers variation two-tailed unpaired student's t-test was used, * p <0.05, ** p<0.01, *** p<0.001.*

5.5 Retinoid's regulation of HOXA5 expression in TNBC cell lines

5.5.1 The effects of Retinoids Short- and Long-term Treatment on TNBC cells phenotype

Retinoic acid signalling and HOX genes have been linked to cellular differentiation and development in mammary tissue[45]. Retinoids have been previously used to induce HOXA5 expression in colorectal cancer[82]. The expression of HOXA5 direct upstream regulator RAR β was also previously induced in TNBC cells using retinoids[95]. Hence, we tested the possibility of using retinoids to induce HOXA5 expression in TNBC using five cell lines representing the different TNBC subtypes. Treatment regimens were tested on the MSL TNBC cell line MDA-MB-231 and MDA-MB-157, the M TNBC cell line BT549, the BL-1 TNBC cell line MDA-MB-468, and the BL-2 TNBC cell line HCC1806. Moreover, we explored whether retinoids treatment duration has an effect on HOXA5 expression and its subsequent cellular response. Therefore, we designed two treatment regimens of 1 μM all-trans-Retinoic acid in a short retinoids treatment course of 6 days and a long retinoids treatment course of 18 days (Figure. 26).

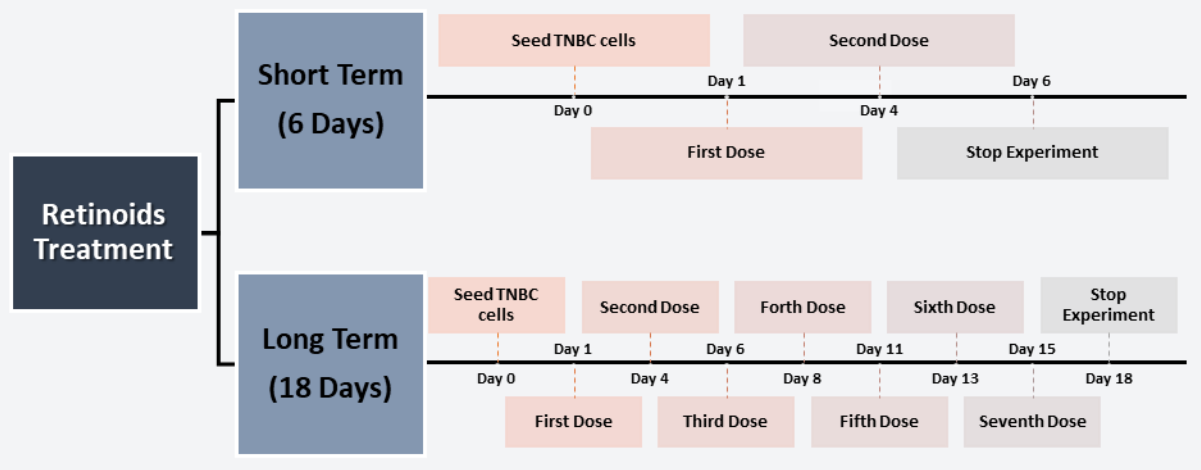

Figure 26: Schematic summary of Retinoids short and long treatment course timeline*: Retinoid dose used in both regimens was 1 μM of all-trans-Retinoic acid (atRA, Sigma-Aldrich).*

Microscopic examination of the effects of treatment on the cell line phenotype was done. In the short term retinoids treatment regimen, no apparent changes in cellular phenotype were noted in all the tested cell lines (Figure. 27A). Contrarily, the long term retinoid treatment regimen induced phenotypic changes in the tested cell line. The MDA-MB-231 and MDA-MB-157 cells appeared to be fewer in number, smaller in size and more rounded in shape compared to the wild-type (Figure. 27B). The BT549 cells were noted to be fewer in number, larger in size and lost their spindle-like shape, acquiring a more asteroid shaped compared to the control (Figure. 27B). Furthermore, the MDA-MB-468 were noted to be less in number and more rounded compared to the control (Figure. 27B). The HCC1806 cells lost their spindle-like shape. Their cells aggregates had a cobblestone appearance, different to the wild-type (Figure. 27B).

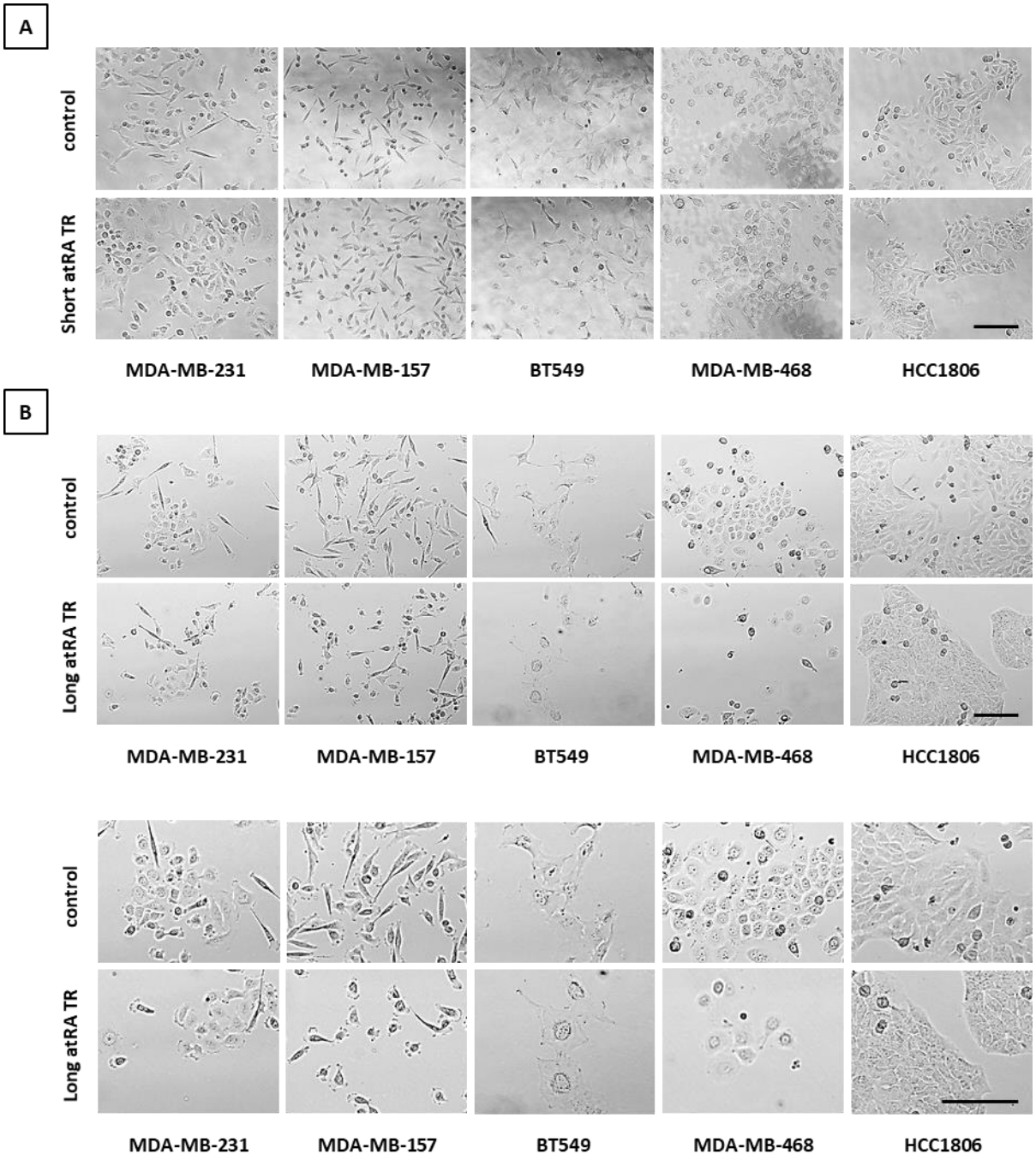

Figure 27: Retinoids short and long treatment course effects on TNBC cells phenotype. *Five cell lines were used to represent different TNBC subtypes: for the MSL subtype: MDA-MB-213 and MDA-MB-157, for the M subtype: BT549, for the BL-1 subtype: MDA-MB-468, and for the BL-2 subtype: HCC1806.* ***A.*** *Short term retinoids treatment (6 days) resulted in no apparent changes in cellular phenotype in all the tested cell lines.* ***B.*** *Long term retinoids treatment (18 days) induced phenotypic changes in the tested cell line. Images were taken using Nikon Eclipse Ti2 Widefield Fluorescent Microscope at 10X magnification. Scale bar = 100 µm.*

5.5.2 The effects of Retinoids Short- and Long-term Treatment on HOXA5 expression and TNBC cells proliferation

To explore the possibility of inducing HOXA5 using retinoids, we tested the effects of both short- and long-term retinoids treatment on HOXA5 expression levels. Retinoic acid receptors are known to direct transcriptional regulators of HOXA5[45]. Therefore we also tested RAR β expression in different TNBC subtypes through quantifying both markers mRNA levels using RT-qPCR. We also investigated the treatment effects on cellular proliferation through quantifying the mRNA levels of the proliferation marker MKI67.

The MDA-MB-231 showed a significant increase in HOXA5 expression compared to the non-treated control in both short- and long-term retinoids treatment (*p=0.0146* and *p=0.0243*, respectively). This was accompanied by high levels of RAR β also in both treatment regimens (*p=0.0028* and *p<0.0001*, respectively) (Figure. 28). The parallel induction of both genes' expression confirmed RAR β previously described regulatory role for HOXA5 and suggest its application in TNBC.

A significant upregulation of the proliferation marker MKI67 was noted in the short-term treatment regimen (*p=0.0260*) (Figure. 28A). Contrarily, significant downregulation of MKI67 was noted following long term retinoids treatment (*p=0.0183*) (Figure. 28B). This indicates that the effect of retinoids treatment on TNBC cellular proliferation might be dependent on the treatment duration.

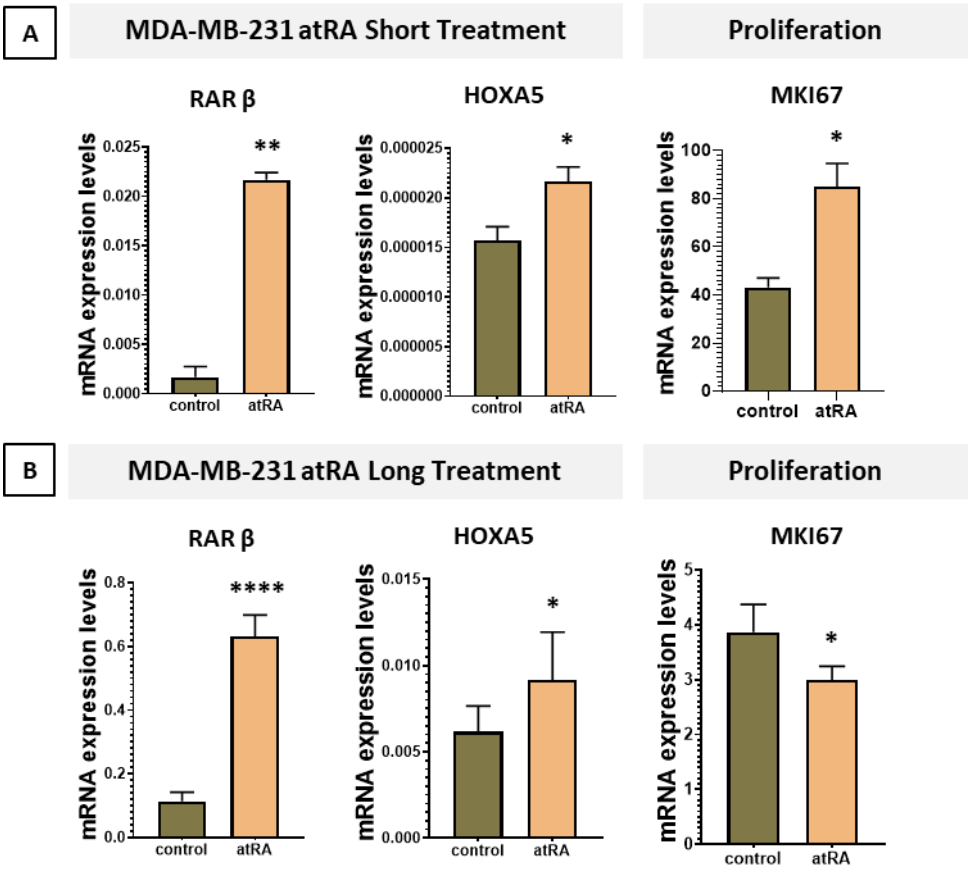

Figure 28: Retinoids short- and long-term treatment effects on HOXA5 pathway regulation and cellular proliferation in MDA-MB-231 cell line. *Both short and long retinoids treatment modulate the expression of RAR-β, HOXA5, and the proliferation marker MKI67.* ***A.*** *Short term retinoids treatment.* ***B.*** *Long term retinoids treatment. RNA samples were analysed using RT-qPCR. n=3, Bars indicate averages ± standard errors. To assess expression variation two-tailed unpaired student's t-test was used, * p <0.05, ** p<0.01, *** p<0.001.*

On the other hand, the MDA-MB-157 cells showed a significant increase in HOXA5 expression in the long-term retinoid's treatment only (*p=0.0017*). This was also accompanied by high levels of RAR β (*p<0.0001*) and an upregulation of the proliferation marker MKI67 (*p=0.0038*) (Figure. 29B). However, unexpectedly the short-term retinoid's treatment showed no effect on HOXA5 expression (*p=0.7385*), despite the significant induction in RAR β levels (*p<0.0001*). No difference in MKI67 was noted in the short-term retinoid's treatment (*p=0.8175*) (Figure. 29A).

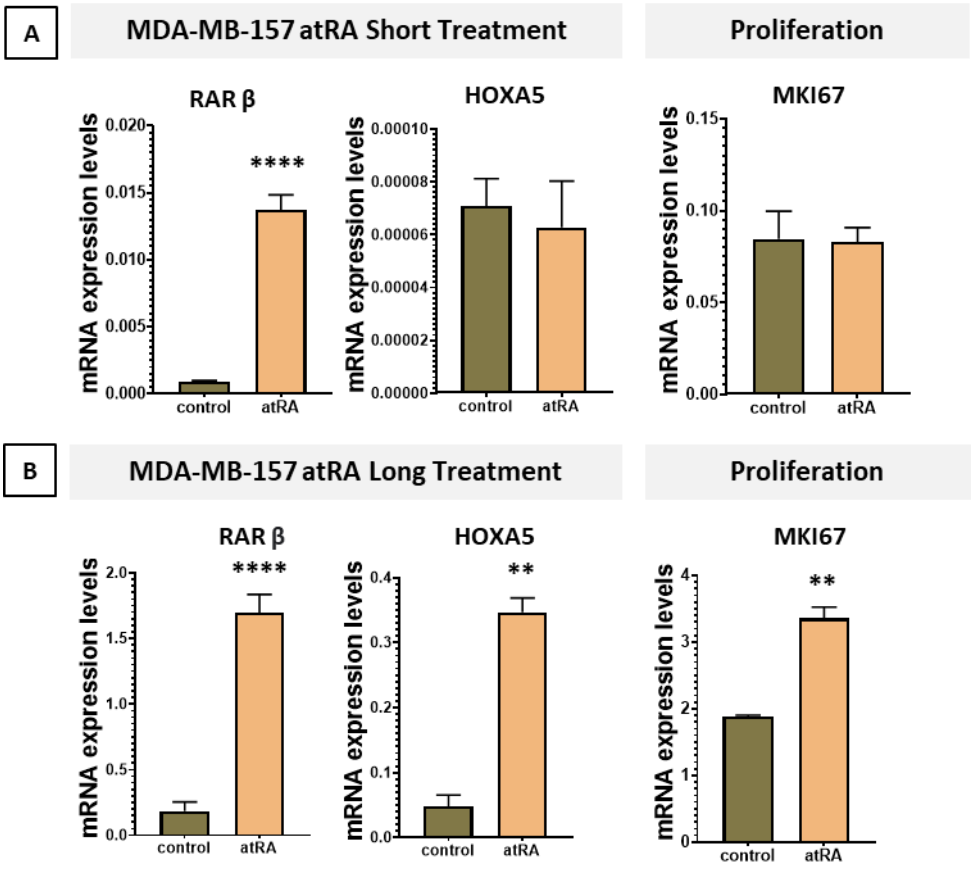

Figure 29: Retinoids short- and long-term treatment effects on HOXA5 pathway regulation and cellular proliferation in MDA-MB-157 cell line. *Both short and long retinoids treatment modulate the expression of RAR-β, HOXA5, and the proliferation marker MKI67.* ***A.*** *Short term retinoids treatment.* ***B.*** *Long term retinoids treatment. RNA samples were analysed using RT-qPCR. n=3, Bars indicate averages ± standard errors. To assess expression variation two-tailed unpaired student's t-test was used, * p <0.05, ** p<0.01, *** p<0.001.*

A similar pattern of retinoids regulation was noted in the BT549 cells, where a significant increase in HOXA5 expression was seen compared to the non-treated control in both short- and long-term retinoids treatment (*p<0.0001* and *p=0.0002*, respectively). Also, this was accompanied by high levels of RAR β also in both treatment regimens (*p<0.0001* and *p=0.0028*, respectively) (Figure. 30). No difference in the expression of the proliferation marker MKI67 was noted in the short-term treatment regimen (*p=0.8890*) (Figure. 30A). Oppositely, significant downregulation of MKI67 was noted following long term retinoids treatment (*p=0.0097*) (Figure. 30B).

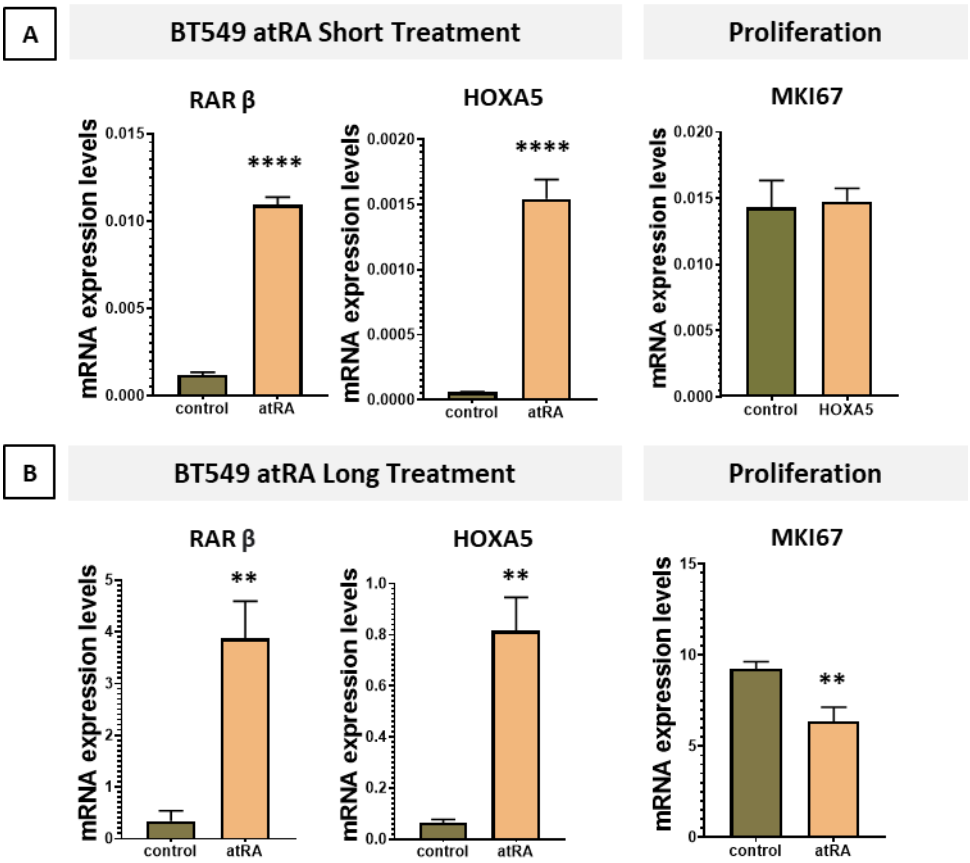

Figure 30: Retinoids short and long treatment effects on HOXA5 pathway regulation and cellular proliferation in BT549 cell line. *Both short and long retinoids treatment modulate the expression of RAR β, HOXA5, and the proliferation marker MKI67.* ***A.*** *Short term retinoids treatment.* ***B.*** *Long term retinoids treatment. RNA samples were analysed using RT-qPCR. n=3, Bars indicate averages ± standard errors. To assess expression variation two-tailed unpaired student's t-test was used, * p <0.05, ** p<0.01, *** p<0.001.*

In the BL-1 TNBC cell line MADA-MB-468 HOXA5 expression was upregulated in response to short term retinoids treatment but not in the long term (*p=0.0290* and *p=0.5711*, respectively) (Figure. 31). In consistence with the previous cell lines and with HOXA5 upregulation, RAR β was also noted to be increased in response to the short treatment (Figure. 31A). However, surprisingly in the long-term, retinoids resulted in significant inhibition of RAR β levels (Figure. 31B). Furthermore, MKI6 regulation was only seen in the short treatment (*p=0.0002*) (Figure. 31A).

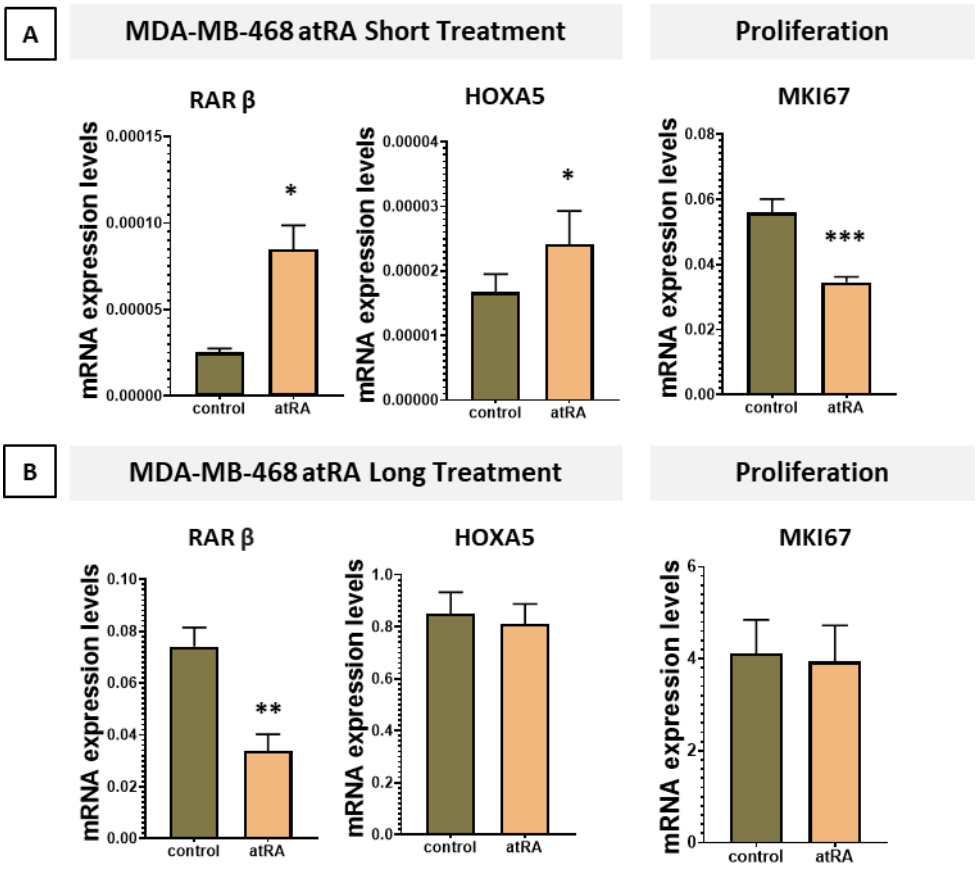

Figure 31: Retinoids short and long treatment effects on HOXA5 pathway regulation and cellular proliferation in MDA-MB-468 cell line. *Both short and long retinoids treatment modulate the expression of RAR-β, HOXA5, and the proliferation marker MKI67.* ***A.*** *Short term retinoids treatment.* ***B.*** *Long term retinoids treatment. RNA samples were analysed using RT-qPCR. n=3, Bars indicate averages ± standard errors. To assess expression variation two-tailed unpaired student's t-test was used, * p <0.05, ** p<0.01, *** p<0.001.*

Similarly, the HCC1806 cells showed a significant increase in HOXA5 expression compared in both short- and long-term retinoids treatment (*p<0.0001* and *p=0.0450*, respectively) (Figure. 32). This was accompanied by high levels of RAR β also in short treatment (*p<0.0001*) (Figure. 32A). However, unexpectedly RAR β was markedly reduced in the long term (*p=0.0117*, respectively) (Figure. 32B). No difference in MKI67 expression was noted in the short-term treatment regimen (*p=0.9482*) (Figure. 32A). Contrarily, MKI67 was significantly downregulated following long term retinoids treatment (*p=0.0002*) (Figure. 32B).

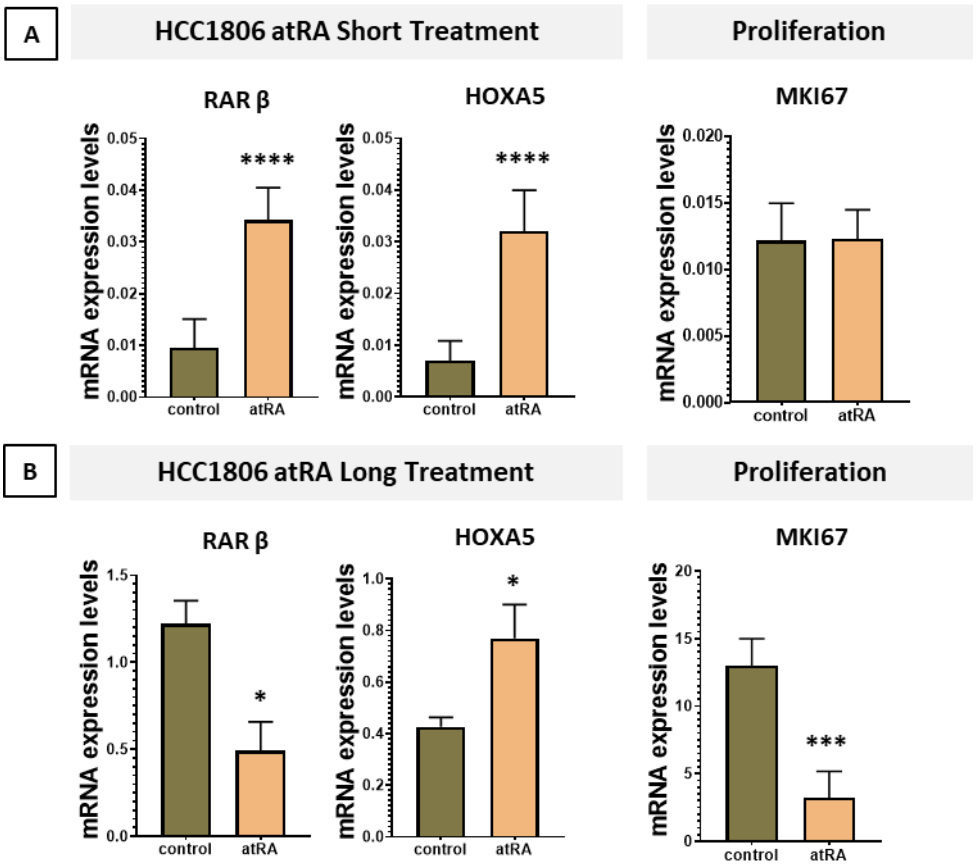

Figure 32: Retinoids short- and long-term treatment effects on HOXA5 pathway regulation and cellular proliferation in HCC1806 cell line. *Both short and long retinoids treatment modulate the expression of RAR β, HOXA5, and the proliferation marker MKI67.* ***A.*** *Short term retinoids treatment.* ***B.*** *Long term retinoids treatment. RNA samples were analysed using RT-qPCR. n=3, Bars indicate averages ± standard errors. To assess expression variation two-tailed unpaired student's t-test was used, * p <0.05, ** p<0.01, *** p<0.001.*

5.5.3 The effects of Retinoids Treatment on TNBC cells mammosphere formation

Our data showed that HOXA5 expression induces the stem cell population within TNBC cells. Our data also showed that retinoids significantly induce HOXA5 expression in TNBC. Therefore, we extended our work to investigate the effects of retinoids on CSC self-renewal in the different TNBC subtypes. Therefore, we conducted a mammosphere formation assay using the five cells. We treated the mammospheres regularly using the long-term retinoids treatment regimen (Figure. 26) and performed serial passaging of the generated mammospheres at clonal density for three generations. Retinoid's treatment enhanced the CSCs population in all the tested cell lines. This was indicated by a significant increase in mammospheres colony counts of the MDA-MB-231 and MDA-MB-157 (*p=0.0140*, *p=0.0033,* respectively) (Figure. 33), the BT549 (*p=0.0076*) (Figure. 34), and the MDA-MB-468 cell line (*p=0.0283*) (Figure. 35). Moreover, the enhanced CSC population induced by retinoids treatment also resulted in a significant increase in mammosphere cell count in the MDA-MB-231 and MDA-MB-157 (*p=0.0213*, and *p<0.0001*, respectively) (Figure. 33), and the MDA-MB-468 (*p=0.0121*) (Figure. 35). But not in the BT549 cell line (*p=0.400*) (Figure. 34).

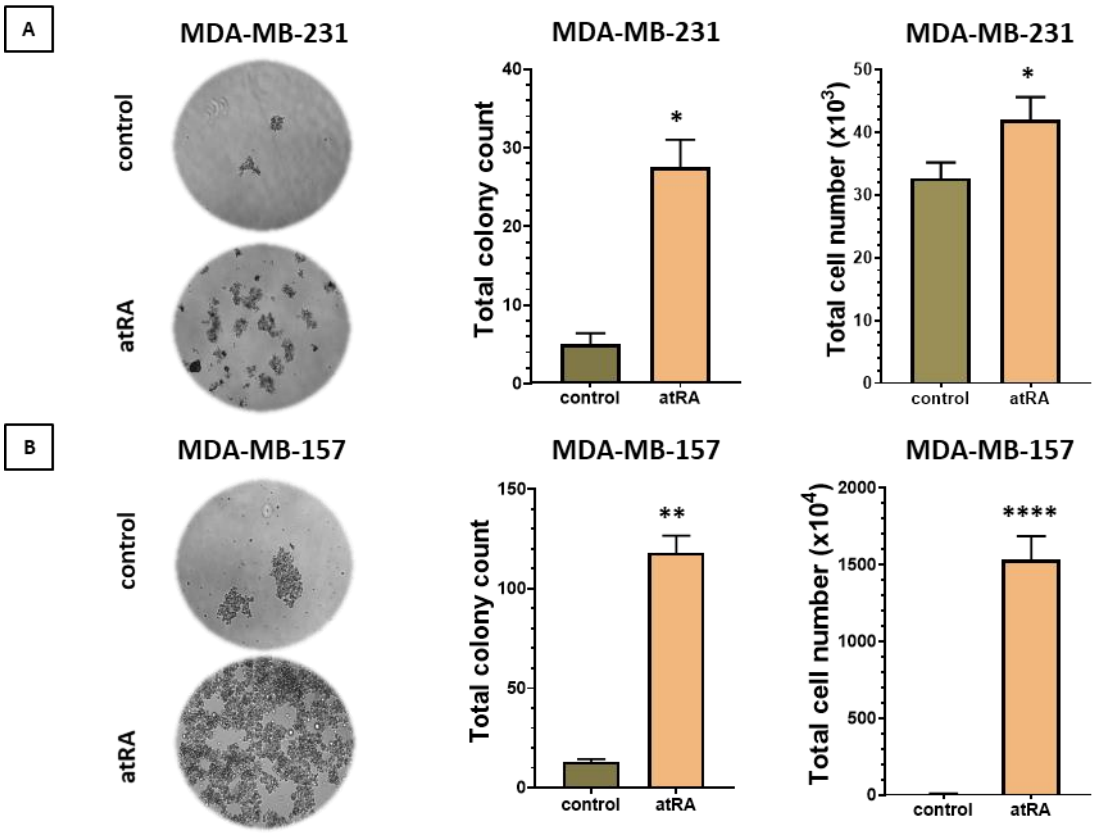

Figure 33: Retinoid's treatment induces CSC self-renewal and anchorage independent cell growth in the MSL TNBC cell lines MDA-MB-231 and MDA-MB-157 Mammospheres*. Retinoid's treatment increase both mammosphere colony count and cell numbers in the MSL subtype.* ***A.*** *MDA-MB-231.* ***B.*** *MDA-MB-157. On the left a representative image of the colonies. On the right the bar-charts quantifies the number of colonies and cell number in the treated cell lines compared to the un-treated controls. The number of colonies were calculated by counting 4X fields of view. n = 3, Bars indicate averages ± standard errors. To assess numbers variation two-tailed unpaired student's t-test was used, * p <0.05, ** p<0.01, *** p<0.001. Images were taken at 10X magnification.*

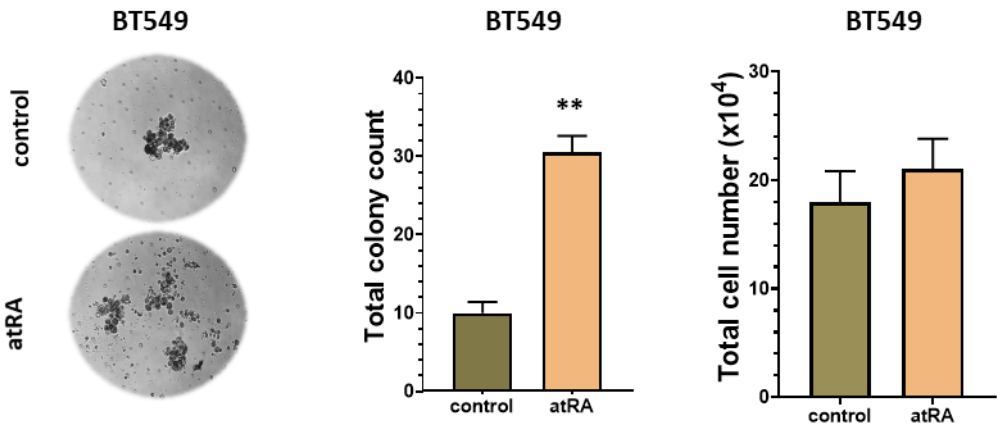

Figure 34: Retinoid's treatment modulates CSC self-renewal and anchorage independent cell growth in the M TNBC cell line BT549 Mammospheres*. Retinoid's treatment increase mammosphere colony count but not their cell numbers in the BT549 cell line. On the left a representative image of the colonies. On the right the bar-charts quantifies the number of colonies and cell number in the treated cell lines compared to the un-treated controls. The number of colonies were calculated by counting 4X fields of view. n = 3, Bars indicate averages ± standard errors. To assess numbers variation two-tailed unpaired student's t-test was used, * p <0.05, ** p<0.01, *** p<0.001. Images were taken at 10X magnification.*

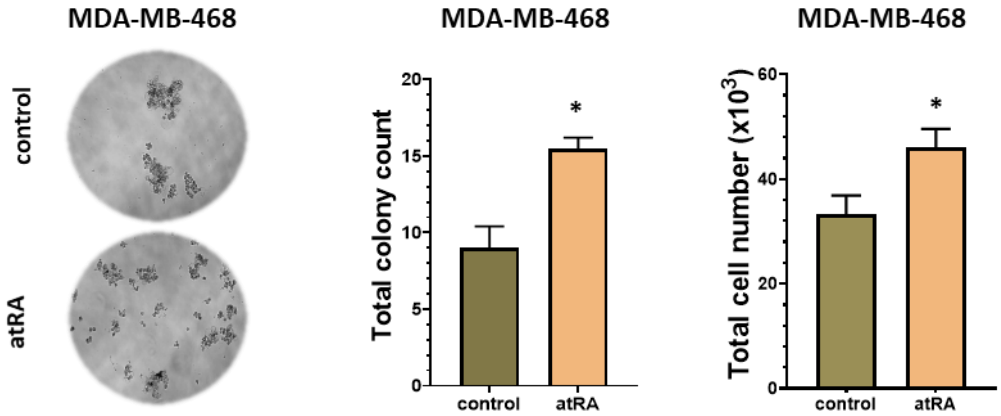

Figure 35: Retinoid's treatment induces CSC self-renewal and anchorage independent cell growth in the BL-1 TNBC cell line MDA-MB-468 Mammospheres*. Retinoid's treatment increase both mammosphere colony count and cell numbers in the MDA-MB-468 cell line. On the left a representative image of the colonies. On the right the bar-charts quantifies the number of colonies and cell number in the treated cell lines compared to the un-treated controls. The number of colonies were calculated by counting 4X fields of view. n = 3, Bars indicate averages ± standard errors. To assess numbers variation two-tailed unpaired student's t-test was used, * p <0.05, ** p<0.01, *** p<0.001. Images were taken at 10X magnification.*

6 Discussion

TNBC comprises a collection of disease entities, each characterised by distinct genomic, transcriptomic and phenotypic features[15]. Despite the improvement in clinical outcomes with the evolution of novel therapeutics, TNBC still bears the worst prognosis among all clinical breast cancer subtypes. This has been referred to the lack of effective therapeutic targets and the heterogeneity of the disease. TNBC is a complex disease; its biological regulation still remains unclear. This is the key behind its challenging therapy and poor clinical outcome. Recently, a new concept has been identified as the main contributor to TNBC rapid, aggressive progression and resistance to therapy. The enrichment of TNBC tumours with CSCs has been determined as their element of tumerogenesis[28]. CSCs have the capacity for self-renewal and multipotency. *Lehmann et.* identification of new TNBC molecular subtypes has further supported this theory. Where a pure mesenchymal stem cell-like subtypes was distinguished[15]. It was previously proposed that TNBC growth originates from a CSCs population[28]. The study supported its hypothesis by identifying a CSC population within the normal breast tissue in TNBC patients[28]. Evidence in the literature around CSCs and their importance in BC is continuously growing, however, their role in TNBC is still unclear. Identifying new contributors to TNBC stemness might assist in resolving this complexity and help identify new TNBC therapeutic strategies.

HOX gene is a family of transcription factors that are well-known for their contribution to genomic regulation of cellular differentiation pathways[31]. HOX gene aberrations can lead to increased cellular plasticity and loss of differentiation ability[37]. In tumours, this was proven to drive cells to acquire stem cells phenotype and self-renewal characteristics, leading to the development of CSCs[38]. HOXA5 is a transcription factor that plays a critical role in embryonic development. In cancer, its expression was identified as an important regulator of cancer growth and progression. The HOXA5 modulation/downregulation is correlated with higher pathological grade and poorer disease outcome in BC[43]. In 2016, *Sukumar et.* reported that HOXA5 expression in mammary epithelial cells stimulates transitioning of the progenitor population within the tumour into a more differentiated state in luminal BC[43]. However, HOXA5 role in TNBC has not been established yet, and the detailed mechanism by which HOXA5 could influence differentiation and development in TNBC is still emerging and less understood. Herein, in this study, we investigated for the first time the role of HOXA5 in TNBC. We explored the possibility of HOXA5 regulation/effect on stemness in TNBC and its consequent effect on cellular proliferation and migration properties. We also investigated whether HOXA5 have a differential regulatory role in different TNBC subtypes based on the existence of six subtypes with distinct molecular characteristics[15].

In this study, we showed that despite HOXA5 low expression in BC compared to the normal breast tissue. Interestingly HOXA5 is more abundant in TNBC tissue and cell lines relative to the other BC subtypes. This suggested that HOXA5 possibly play an important role in TNBC. Thus, we tried to link this finding to HOXA5 prognostic value in BC. Kaplan-Meier survival analysis revealed a direct link between HOXA5 levels and BC patient's clinical outcome. BC patients with high HOXA5 mRNA expression showed higher OS and RFS. Nevertheless, when we restricted the study to only TNBC patients, the patients expressing high HOXA5 levels were noted to have a low RFS probability accompanied by a short metastasis-free survival period compared to other BC patients. The analysis clearly showed that unlike in other BC subtypes, in the long

term, HOXA5 expression confers TNBC patient's a lower prognosis and poor clinical outcome.

We expanded our survival analysis to explore whether HOXA5 prognostic value would vary between the different TNBC subtypes. Survival analysis showed that an increase in HOXA5 expression is associated with better OS in BL-1 TNBC subtype only. However, this did not affect patient RFS. This finding can possibly explain the lower expression of HOXA5 in the BL-1 TNBC tissue subtype. The more HOXA5 is expressed in this subtype, the more chance of survival it confers the patients, reducing cancer cells survival possibility. HOXA5 induction in this subtype could possibly improve patient survival and outcome.

To further explore our results and to understand the role of HOXA5 in different TNBC subtypes, we studied five TNBC cell lines representing the BL-1, BL-2, M, and MSL subtypes. We quantified HOXA5 expression in the tested cell lines, and accordingly, we either overexpressed or knockdown HOXA5 levels. As previously published, the M and MSL subtypes express high levels of HOX genes, including HOXA5[20]. Our data also showed higher expression of HOXA5 in the M subtypes, and therefore using an shRNA loaded lentivirus vector, we knocked down HOXA5 in the BT549 cells. We also expressed HOXA5 in the non-expressing MDA-MB-231 cells.

First, we aimed to explain our survival results further and determine whether stemness play a role in the noted effect. Herein, we tested the effects of HOXA5 on CSCs self-renewal in the different TNBC cells by assessing cell growth in mammospheres. HOXA5 expression enhanced CSCs population in the MDA-MB-231 and MDA-MB-468 cells. This was clearly observed by the increase in mammospheres colonies and cell counts in both cell lines. Moreover, its KD in BT549 cells did not reverse this effect, where it did not change the stem cell population. This would indicate that HOXA5 may not be the only factor regulating stemness in the BT549 cells. HOXA5 was also proved to be upregulated in the TNBC subtype by our GE analysis. The co-existence of a high CSCs population and high HOXA5 expression in TNBC in our data suggest an interesting correlation. Therefore, we further explored that HOXA5 is possibly responsible for the enrichment of TNBC tumours with CSCs.

To explain our results further and explore HOXA5 role in different TNBC subtypes we assessed cellular and molecular characteristics of the MDA-MB-231-HOXA5 cells and the MDA-MB-468-HOXA5 cells. We noted that HOXA5 expression did not affect the cellular morphology and phenotype of the MDA-MB-231 cell line. Whilst the MDA-MB-468 cells appeared more rounded and formed more clusters compared to the wild-type. This was reflected in the colony formation assay, where the HOXA5 cells formed more colonies, and therefore, we suggest that HOXA5 increases clonogenicity in the BL-1 TNBC subtype.

Despite the absence of phenotypic changes in the MDA-MB-231-HOXA5 cells, HOXA5 modulated the expression of different genes and proteins within these cells. MDA-MB-231-HOXA5 cells showed a significant reduction in the expression of the adhesion and phenotypic markers CDH1 and KRT14, producing possibly a less differentiated state of these cells. Similarly, HOXA5 expression suppressed CDH1 in the MDA-MB-468. It was previously reported by *Sukumar et.* that the loss of HOXA5 in non-malignant and luminal breast cancer cells is accompanied by loss of the cell adhesion markers CDH1[43]. The study demonstrated that the loss of CDH1 promotes the process of cellular dedifferentiation and transformation in BC. The study also

reported that HOXA5 expression in mammary epithelial cells stimulates the transition of the progenitor population within the tumour into a more differentiated state, where it identified HOXA5 as a tumour suppressor in BC through its direct transcriptional regulation for CDH1[43]. Contrarily, our findings suggest that inducing HOXA5 expression in the TNBC cells might suppress cellular differentiation through downregulating CDH1 expression. Also, HOXA5 KD in the BT549 cells induced CDH1 expression, confirming HOXA5 negative regulation of CDH1 in TNBC cells. However, this assessment is not sufficient. Western blotting can also be used for the comparison of CDH1 protein expression. Moreover, assessing the expression status of the stemness markers CD44 and CD24 using FACS analysis in the HOXA5 expressing and knockdown cells could be used to evaluate further the cells stemness/differentiation status and support our findings.

The previously described role of CDH1 in BC and the fact that its expression is lost in response to HOXA5 expression in both the MSL and BL-1 subtypes supports the possibility that HOXA5 regulates CDH1 in TNBC. Our data also suggests a possible oncogenic property of HOXA5 in both MSL and BL-1 TNBC subtypes despite its tumour suppresser characteristics in other BC subtypes[43]. HOXA5 subsequent loss of CDH1 generates a less differentiated cell population. This finding corresponds to our mammospheres data and further support that HOXA5 expression could be the key behind CSCs enrichment in TNBC. HOX genes expression has been reported to be linked with cancer under different circumstances, where depending on cancer type , HOX genes can act as both tumour suppressor or a proto-oncogenes[96], supporting the possibility of our projected role of HOXA5 in TNBC.

Additionally, in the MDA-MB-468-HOXA5 cells, the high CDH1 expression accompanied by the low KRT14 expression indicate that HOXA5 increase BL-1 subtype aggressiveness. KRT14 expression characterizes the basal epithelial layer. Its expression is linked to aggressive BC tumours, where is known to be expressed in sporadic invasive ductal breast cancers that are mostly of aggressive grade III[97].

Our data also suggest that the ALDH1 family are critical for HOXA5 induced stemness in TNBC. It was previously reported that the MSL subtype expresses high stemness-related genes, including ALDH1A genes[20]. Our results showed that HOXA5 expression induces ALDH1A1 expression in the MDA-MB-231. This was supported by a significant inhibition of ALDH1A1 levels following HOXA5 KD in BT549 cells, confirming HOXA5 positive regulatory role for ALDH1A1 in TNBC. ALDH1A1 is a member of the ALDH1 superfamily of enzymes. It is involved in the bioactivation and\or detoxification of different aldehydes. The ALDH1A1 and ALDH1A3 mainly are known to be involved in retinoic acid metabolism[92]. Therefore, modulation of their expression in response to HOXA5 is expected, where the retinoic acid receptor RAR β is a known upstream regulator of HOXA5[45]. ALDH1A1 is an independent prognostic marker in TNBC; its expression was linked to poor clinical outcome in BC patients[92].

In 2017 a study done by *John D. Lewis and Alison L. Allan* revealed that ALDH1A1 expression provokes proliferation, metastatic behaviour and resistance to therapy in BC[98]. The previously described metastatic function for ALHDH1A1 clearly explains our migration assay results. We performed a wound-healing assay to test the effects of HOXA5 on cellular migration. The assay showed that HOXA5 expression significantly increases cell migration speed in the MDA-MB-231 cells. The H&E staining of MDA-MB-231-HOXA5 tumours also revealed the invasive properties of these tumours.

Tumour cells were clearly escaping the tumour margin, invading the surrounding tissue. We suggest that ALDH1A1 induction could be the mechanism by which HOXA5 induces this metastatic behaviour in the MSL TNBC subtype. The described proliferation role of ALDH1A1 can also be related to the increased levels of the proliferation marker MKI67 in the MDA-MB-231-HOXA5 cells. Despite that, no difference was seen in MKI67 mRNA expression following HOXA5 expression. The IF staining showed high intensity of the KI67 in the nucleus in the MDA-MB-231-HOXA5 cells. Also, with H&E staining, the MDA-MB-231-HOXA5 tumour sections appeared to be highly proliferative. Therefore, we believe that HOXA5 induces cellular proliferation in the MSL TNBC. Quantifying the stained tumour proliferation through calculating their mitotic-index will assist in concluding the result. Testing more breast cancer proliferation markers such as the proliferating cell nuclear antigen (PCNA) or mini-chromosome maintenance (MCM) protein relative to HOXA5 expression can also be used to support our findings further.

Moreover, unexpectedly, unlike ALDH1A1, ALDH1A3 levels were noted to be reduced in MDA-MB-231-HOXA5 cells. This is thought to be a result of the high HOXA5 expression inducting a feedback inhibition on its upstream regulators[92]. A regulatory feedback loop was previously described between ALDH1 genes and the retinoic acid pathway in different cancer tissues[92]. The high HOXA5 expression might be inducting a feedback inhibition on its upstream regulators, namely retinoic acid and its metabolizing enzymes, the ALDH1 family[92]. In *John D. Lewis and Alison L. Allan* work ALDH1A3 was liked to cellular clonogenicity, where ALDH1A3 knockdown reduced colony formation in TNBC cells[98]. A similar effect was seen in our work, where a significant reduction in colony numbers was seen in MDA-MB-231-HOXA5 cells. The induced ALDH1A3 inhibition could explain the inhibition of clonogenicity of HOXA5 expressing cells.

Previously, *Ordonez-Moran et al. in-vivo* work demonstrated that HOXA5 expression reduces CSCs in colorectal cancer[82]. The observed modulation of the stemness, self-renewal and proliferation markers ALDH1A1 and ALHD1A3 accompanying HOXA5 alteration also supports HOXA5 regulatory in TNBC CSCs. However, unlike in colorectal cancer, we believe that HOXA5 enhanced the CSCs population in the TNBC. The described regulatory role of HOXA5 for colorectal CSCc was linked to a reciprocal feedback between HOXA5 and Wnt signalling. The Wnt pathway suppresses HOXA5 to maintain stemness inside the intestinal crypts. HOXA5 becomes active only outside the intestinal crypt, where it inhibits the Wnt pathway to enforce differentiation. In colon cancer, HOXA5 is downregulated, and its re-expression induces loss of cancer stem cells phenotype[82]. However, we believe that this is not the case in the TNBC. We investigated the HOXA5-Wnt pathway connection in TNBC through testing the expression of the Wnt pathway gene AXIN2 and β-catenin protein localization. HOXA5 did not affect AXIN2 mRNA expression, neither did it change β-catenin in the MDA-MB-231. HOXA5 KD also showed no effects on the expression of both AXIN2 gene and β-catenin protein localization. Therefore, we suggest that HOXA5 stemness regulatory role in the MSL TNBC subtype is independent of the Wnt pathway.

On The other hand, our data suggest that HOXA5 possibly inhibit of Wnt pathway in the BL-1 TNBC subtype. Where MDA-MB-468-HOXA5 showed lower AXIN2 expression compared to control cells, this finding might suggest an interaction between HOXA5 and the Wnt pathway genes in the BL-1 subtype but not in the other

subtypes. However, no difference in β-catenin localization was noted in the MDA-MB-468 cell line; therefore it is difficult to conclude a connection between the two genes at least in this subtype.

Lehmann et. previously reported that the basal-like subtypes were found to express high levels of proliferation and DNA damage response genes. Their study suggested that BL TNBC tumours could profit from therapies targeting highly proliferative tumours such as DNA-damaging agents and anti-mitotic[17]. Our data showed that HOXA5 expression in the BL-1 subtype inhibits the marker of proliferation MKI67 mRNA expression. This was supported by IF staining, where MDA-MB-468-HOXA5 cells showed an apparent reduction in staining intensity of KI67. However, the variation in MKI67 levels regulation between the subtypes indicates that HOXA5/KI67 interaction in TNBC generally cannot be determined. HOXA5 being involved in different proliferation pathways in the different TNBC subtypes might explain this variation. Therefore, quantifying IF staining of KI67 in single cells, IHC staining of tumour section with the calculation of the mitotic index, and testing the previously mentioned proliferation markers (PCNA and MCM) in all the tested cell lines could help clarify HOXA5 effects on TNBC proliferation.

A recent study conducted by *Jeremy Jenkins & Prisca Liberali* revealed for the first time a connection between retinoids treatment and the yes-associated protein 1 (YAP) gene activation[93]. YAP and its co-activator PDZ-binding motif (TAZ) are transcriptional regulators commonly activated in human malignancies. Studied revealed their importance for cancer initiation and growth in most solid tumours. YAP/TAZ activation induces cancer stem cells characteristics, proliferation, metastasis, and resistance to therapy. Cancer cell addiction to YAP/TAZ was suggested to potentially represents a promising therapeutic target in different malignancies[100]. In cancer, YAP/TAZ intracellular localization is the key determinant of their activity. YAP/TAZ nucleo-cytoplasmic shuttling and aberrant nuclear localization of YAP/TAZ is noted in numerous human malignancies and is identified as an indication of their activity[101]. *Jeremy Jenkins & Prisca Liberali* study aimed to test the effects of a set of compounds on the intestinal organoids and to identify their effects on organoids formation and the regeneration mechanistic. Retinoids were one of the tested compounds. The study showed that atRA treatment of the intestinal organoids resulted in invariable localization of YAP to the cytoplasm. An accompanied enhanced maturation of the organoids was noted, indicating a decrease in the regenerative signature of enterocytes within the organoids. Their results proposed that atRA promotes enterocyte differentiation. The change in YAP localization was suggested to play a part in this regulation, where it was linked to a more differentiated state of the cells[93]. We found interesting the suggested contribution of YAP to cellular differentiation in response to atRA treatment. Thus, based on the fact that HOXA5 is a known cellular self-renewal regulator that is regulated by retinoids, we investigated whether there is a possible interaction between YAP/TAZ and HOXA5.

Our analysis showed a negative correlation between HOXA5 expression and YAP1 target genes in the MSL and M TNBC subtypes. The Tumour Associated Calcium Signal Transducer 2 (TACSTD2), the connective tissue growth factor (CTGF), and Cysteine-rich angiogenic inducer 61 (CYR61) are important YAP target genes that are known to regulate many biological processes and play an important role in response to therapy in cancer[102]. Our RT-qPCR data showed that HOXA5 suppress TACSTD2 and CTGF mRNA levels in the MDA-MB-231 cells. HOXA5 KD reversed this effect in

the BT549. This was further supported by the lower intensity YAP IF staining within the MDA-MB-231-HOXA5 cells compared to control cells. HOXA5/CYP61 interaction data was inconclusive. Therefore, our data suggests negative regulation of HOXA5 for the YAP target genes in the MSL and the M TNBC subtypes; however, this needs to be further evaluated. Moreover, HOXA5 expression in the BL-1 cell line MDA-MB-468 also modulated the expression of TACSTD2 and CTGF. However, there was no difference in the staining intensity of YAP. Our results reveal an effect for HOXA5 on some YAP target genes; nonetheless, it is not clear whether this is a direct effect or if YAP is a possible intermediate. The revealed YAP/TAZ-HOXA5 regulation is an interesting aspect for future studies. Investigating this pathway might reveal a new pathway by which HOXA5 affects TNBC.

Retinoids analogues are promising agents that are being used for managing certain haematological malignancies and many solid tumours, including breast cancer[47,53,65]. Retinoids hold cellular differentiating, anti-proliferative, and apoptotic effects that are largely mediated by activation of the nuclear hormone retinoic acid receptors[46,47]. In BC, these effects are mediated mainly through RAR β[103,104]. Retinoic acid signalling and HOX genes have been linked to cellular differentiation and development in mammary tissue[45]. It is not clear whether this applies to TNBC or not. RAR β is a direct upstream regulator for HOXA5 in BC[95]. Retinoids have been previously used to induce HOXA5 expression in colorectal cancer[82].

Similarly, we were able to successfully induce HOXA5 expression in all the tested TNBC subtypes using atRA. Our data clearly showed that the induction in HOXA5 was in response to retinoids induced upregulation of RAR β in the M and MSL TNBC subtypes. The variation in RAR β expression despite the distinct upregulation in HOXA5 levels in the BL-1 and BL-2 cell lines implies that other RAR could be involved in HOXA5 regulation in the BL TNBC subtypes.

Moreover, through testing short and long term atRA treatment effects, we also showed that the duration of retinoids treatment does not remarkably affect their induction of HOXA5. However, it did affect the treated cell phenotypic characteristics and proliferation rate. There was an obvious variation in MKI67 levels detected by RT-qPCR. Nevertheless, from the lab work, it was noted that with the frequency of the doses, an initial increase in cellular proliferation was observed in the first 10 days, and cells required frequent passaging. Following that, cellular proliferation was apparently reducing with each dose. The GM cells data suggest that HOXA5 induce cellular proliferation, and therefore we suggest that the noted proliferation inhibition in some of the treated cell lines could be possibly due to the effects of retinoids on other cellular pathways.

In addition, our study also investigated the effect of retinoids on the TNBC CSCs population. We believe that retinoids induce stem cell population within the different TNBC subtypes through possibly upregulating HOXA5. Retinoids were reported to have two opposing effects in regulating CSCs depending on the cell type[105]. Retinoic acid was previously used to induce differentiation in the mammary MCF12A cells and T47D luminal breast cancer cells[106]. Nonetheless, *Marcato P, Dean CA et al.* has recently reported that RA induces stemness and tumour xenografts proliferation and progression in the MDA-MB-231 cell line, but this effect was reversed in the MDA-MB-468[107]. RA mediate these effects through activating more than one thousand different genes in TNBC cells[107]. Our data showed that retinoids treatment enhanced the CSCs

population in all the tested cell lines. This was indicated by an apparent increase in mammospheres colony counts in the MDA-MB-231, the MDA-MB-157, the BT549, and the MDA-MB-468 cell line.

7 Conclusion:

BC is a heterogeneous group of diseases, and TNBC is the most aggressive subtype of BC. TNBC patient's poor clinical outcome and response to therapy were referred to the heterogeneity of the disease and lack of therapeutic targets. Recently, the enrichment of TNBC tumours with cancer stem cells has been identified as a key for their oncogenic properties. However, the specific mechanisms that regulate TNBC self-renewal, cellular differentiation and consequently their progression and migration ability is still unclear.

In this study, we identified HOXA5 transcription factor as a regulator of cancer stemness in TNBC. Under different circumstances, and depending on cancer type , HOX genes can act as tumour suppressors or proto-oncogenes[96]. HOXA5 is reported to be beneficial in several breast cancer subtypes; however, HOXA5 behaves differently in TNBC, and this is confirmed by our data. Our data demonstrate that HOXA5 induces 3D mammosphere formation in TNBC. This may also correspond to the different subtypes, where this effect is further evoked in the M and MSL TNBC subtypes. TNBC is known to be rich with EMT and motility related pathways[15]. We believe that HOXA5 stimulates cellular migration in TNBC. Our data suggest that HOXA5 produces in a specific context the proposed effects through modulating the expression of a set of genes involved in cancer cells stemness, proliferation, and migration. HOXA5 can modulate the expression of the adhesion/phenotypic markers CDH1 and KRT14, the ALDH1 family members ALDH1A1 and ALDH1A3, and some YAP target genes (TACSTD2, CTGF, and CYR61).

The balance between self-renewal and differentiation of stem cells is crucial for cancer development and progression. We found that retinoids induce mammosphere formation in TNBC. We identify HOXA5 as an important element that can determine the balance to stemness or differentiation. In TNBC, RA induce HOXA5 to maintain stemness traits and enhance metastatic behaviour. This study may help to identify a potential therapeutic approach to treat TNBC patients by considering HOXA5 expression modulation

8 References

1. Torre, L. A. *et al.* Global cancer statistics, 2012. *CA. Cancer J. Clin.* **65**, 87–108 (2015).
2. Breast cancer statistics | Cancer Research UK. https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/breast-cancer#heading-Zero.
3. Rakha, E. A. *et al.* Triple-negative breast cancer: Distinguishing between basal and nonbasal subtypes. *Clin. Cancer Res.* **15**, 2302–2310 (2009).
4. Goldhirsch, A. *et al.* Personalizing the treatment of women with early breast cancer: Highlights of the st gallen international expert consensus on the primary therapy of early breast Cancer 2013. *Ann. Oncol.* **24**, 2206–2223 (2013).
5. CM, P. *et al.* Molecular portraits of human breast tumours. *Nature* **406**, 747–752 (2000).
6. Rakha, E., Ellis, I. & Reis-Filho, J. Are Triple-Negative and Basal-Like Breast Cancer Synonymous? (2008) doi:10.1158/1078-0432.CCR-07-1943.
7. MC, C. *et al.* Basal-like breast cancer defined by five biomarkers has superior prognostic value than triple-negative phenotype. *Clin. Cancer Res.* **14**, 1368–1376 (2008).
8. M, S., Y, B., M, T., T, H. & T, N. Screening for basal marker expression is necessary for decision of therapeutic strategy for triple-negative breast cancer. *J. Surg. Oncol.* **97**, 30–34 (2008).
9. MC, C. *et al.* Basal-like breast cancer defined by five biomarkers has superior prognostic value than triple-negative phenotype. *Clin. Cancer Res.* **14**, 1368–1376 (2008).
10. Kreike, B. *et al.* Gene expression profiling and histopathological characterization of triple-negative/basal-like breast carcinomas. *Breast Cancer Res.* **9**, R65 (2007).
11. Morris, G. J. *et al.* 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 (2007).
12. Haffty, B. G. *et al.* Locoregional relapse and distant metastasis in conservatively managed triple negative early-stage breast cancer. *J. Clin. Oncol.* **24**, 5652–5657 (2006).
13. Dent, R. *et al.* Triple-negative breast cancer: Clinical features and patterns of recurrence. *Clin. Cancer Res.* **13**, 4429–4434 (2007).
14. Chaudhary, L. N., Wilkinson, K. H. & Kong, A. Triple-Negative Breast Cancer: Who Should Receive Neoadjuvant Chemotherapy? *Surgical Oncology Clinics of North America* vol. 27 141–153 (2018).
15. Lehmann, B. D. *et al.* Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. *J. Clin. Invest.* **121**, 2750–2767 (2011).

16. Kwei, K. A., Kung, Y., Salari, K., Holcomb, I. N. & Pollack, J. R. Genomic instability in breast cancer: Pathogenesis and clinical implications. *Molecular Oncology* vol. 4 255–266 (2010).

17. Liu, Y. *et al.* Retinoic acid receptor beta mediates the growth-inhibitory effect of retinoic acid by promoting apoptosis in human breast cancer cells. *Mol. Cell. Biol.* **16**, 1138–1149 (1996).

18. Juul, N. *et al.* Assessment of an RNA interference screen-derived mitotic and ceramide pathway metagene as a predictor of response to neoadjuvant paclitaxel for primary triple-negative breast cancer: A retrospective analysis of five clinical trials. *Lancet Oncol.* **11**, 358–365 (2010).

19. Bauer, J. A. *et al.* Identification of markers of taxane sensitivity using proteomic and genomic analyses of breast tumors from patients receiving neoadjuvant paclitaxel and radiation. *Clin. Cancer Res.* **16**, 681–690 (2010).

20. Yin, L., Duan, J. J., Bian, X. W. & Yu, S. C. Triple-negative breast cancer molecular subtyping and treatment progress. *Breast Cancer Res.* **22**, 1–13 (2020).

21. Guarino, M. Src signaling in cancer invasion. *J. Cell. Physiol.* **223**, n/a-n/a (2009).

22. Tian, J. *et al.* Dasatinib sensitises triple negative breast cancer cells to chemotherapy by targeting breast cancer stem cells. *Br. J. Cancer* **119**, 1495–1507 (2018).

23. Hennessy, B. T. *et al.* Characterization of a naturally occurring breast cancer subset enriched in epithelial-to-mesenchymal transition and stem cell characteristics. *Cancer Res.* **69**, 4116–4124 (2009).

24. Al Sayed, A. D., El Weshi, A. N., Tulbah, A. M., Rahal, M. M. & Ezzat, A. A. Metaplastic carcinoma of the breast Clinical presentation, treatment results and prognostic factors. *Acta Oncol. (Madr).* **45**, 188–195 (2006).

25. Hayes, M. J., Thomas, D., Emmons, A., Giordano, T. J. & Kleer, C. G. Genetic changes of Wnt pathway genes are common events in metaplastic carcinomas of the breast. *Clin. Cancer Res.* **14**, 4038–4044 (2008).

26. Jung, Y. S. & Park, J. Il. Wnt signaling in cancer: therapeutic targeting of Wnt signaling beyond β-catenin and the destruction complex. *Experimental and Molecular Medicine* vol. 52 183–191 (2020).

27. Doane, A. S. *et al.* An estrogen receptor-negative breast cancer subset characterized by a hormonally regulated transcriptional program and response to androgen. *Oncogene* **25**, 3994–4008 (2006).

28. CJ, O., T, C., I, G., D, C. & Y, P. Cancer stem cells in triple-negative breast cancer: a potential target and prognostic marker. *Biomark. Med.* **12**, 813–820 (2018).

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

30. S, L. *et al.* Breast cancer stem cells transition between epithelial and mesenchymal states reflective of their normal counterparts. *Stem cell reports* **2**,

78–91 (2013).

31. Mark, M., Rijli, F. M. & Chambon, P. Homeobox genes in embryogenesis and pathogenesis. *Pediatric Research* vol. 42 421–429 (1997).

32. Shah, N. & Sukumar, S. The Hox genes and their roles in oncogenesis. *Nature Reviews Cancer* vol. 10 361–371 (2010).

33. Li, B., Huang, Q. & Wei, G. H. The role of hox transcription factors in cancer predisposition and progression. *Cancers (Basel).* **11**, 1–25 (2019).

34. Lawrence, H. J., Sauvageau, G., Humphries, R. K. & Largman, C. The Role of *HOX* Homeobox Genes in Normal and Leukemic Hematopoiesis. *Stem Cells* **14**, 281–291 (1996).

35. Gilbert, P. M. *et al.* HOXA9 regulates BRCA1 expression to modulate human breast tumor phenotype. *J. Clin. Invest.* **120**, 1535–1550 (2010).

36. Novak, P. *et al.* Epigenetic inactivation of the HOXA gene cluster in breast cancer. *Cancer Res.* **66**, 10664–10670 (2006).

37. Smith, J., Zyoud, A. & Allegrucci, C. A case of identity: Hox genes in normal and cancer stem cells. *Cancers (Basel).* **11**, 1–13 (2019).

38. Shah, M. & Allegrucci, C. Stem cell plasticity in development and cancer: Epigenetic origin of cancer stem cells. *Subcell. Biochem.* **61**, 545–565 (2013).

39. Visvader, J. E. & Lindeman, G. J. Cancer stem cells in solid tumours: Accumulating evidence and unresolved questions. *Nature Reviews Cancer* vol. 8 755–768 (2008).

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

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

42. Rodrigues, M. F. S. D., Esteves, C. M., Xavier, F. C. A. & Nunes, F. D. Methylation status of homeobox genes in common human cancers. *Genomics* vol. 108 185–193 (2016).

43. Teo, W. W. *et al.* HOXA5 determines cell fate transition and impedes tumor initiation and progression in breast cancer through regulation of E-cadherin and CD24. *Oncogene* **35**, 5539–5551 (2016).

44. Raman, V. *et al.* Compromised HOXA5 function can limit p53 expression in human breast tumours. *Nat. 2000 4056789* **405**, 974–978 (2000).

45. Chen, H. *et al.* HOXA5 acts directly downstream of retinoic acid receptor β and contributes to retinoic acid-induced apoptosis and growth inhibition. *Cancer Res.* **67**, 8007–8013 (2007).

46. Kalemkerian, G. P. & Ramnath, N. Retinoids and apoptosis in cancer therapy. *Apoptosis* **1**, 11–24 (1996).

47. Nadzan, A. M. Retinoids for the Treatment of Oncological Disease. *Annu. Rep. Med. Chem.* **30**, 119–128 (1995).

48. Epstein, F. H. & Goodman, D. S. Vitamin A and Retinoids in Health and Disease. *N. Engl. J. Med.* **310**, 1023–1031 (1984).

49. O'Byrne, S. M. & Blaner, W. S. Retinol and retinyl esters: Biochemistry and physiology. *Journal of Lipid Research* vol. 54 1731–1743 (2013).

50. Freemantle, S. J., Spinella, M. J. & Dmitrovsky, E. Retinoids in cancer therapy and chemoprevention: Promise meets resistance. *Oncogene* **22**, 7305–7315 (2003).

51. Burt Wolbach, S. & Howe, P. R. Tissue changes following deprivation of fatsoluble a vitamin. *J. Exp. Med.* **42**, 753–778 (1925).

52. Hong, W. K. & Sporn, M. B. Recent advances in chemoprevention of cancer. *Science (80-.).* **278**, 1073–1077 (1997).

53. Smith, M. A., Parkinson, D. R., Cheson, B. D. & Friedman, M. A. Retinoids in cancer therapy. *Journal of Clinical Oncology* vol. 10 839–864 (1992).

54. Evans, TRJ & al. *Retinoids: present role and future potential*. (1999).

55. Gudas, L. J. Retinoids and vertebrate development. *J. Biol. Chem.* **269**, 15399–15402 (1994).

56. Dobrotkova, V., Chlapek, P., Mazanek, P., Sterba, J. & Veselska, R. Traffic lights for retinoids in oncology: Molecular markers of retinoid resistance and sensitivity and their use in the management of cancer differentiation therapy. *BMC Cancer* **18**, 1–13 (2018).

57. Niles, R. M. Signaling pathways in retinoid chemoprevention and treatment of cancer. *Mutat. Res. - Fundam. Mol. Mech. Mutagen.* **555**, 97–105 (2004).

58. Kuppumbatti, Y. S., Rexer, B., Nakajo, S., Nakaya, K. & Mira-Y-Lopez, R. CRBP suppresses breast cancer cell survival and anchorage-independent growth. *Oncogene* **20**, 7413–7419 (2001).

59. Esteller, M. *et al.* Hypermethylation-associated inactivation of the cellular retinol-binding-protein 1 gene in human cancer. *Cancer Res.* **62**, 5902–5905 (2002).

60. Giguere, V., Ong, E. S., Segui, P. & Evans, R. M. Identification of a receptor for the morphogen retinoic acid. *Nature* **330**, 624–629 (1987).

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

62. Heyman, R. A. *et al.* 9-cis retinoic acid is a high affinity ligand for the retinoid X receptor. *Cell* **68**, 397–406 (1992).

63. Levin, A. A. *et al.* 9-Cis retinoic acid stereoisomer binds and activates the nuclear receptor RXRα. *Nature* **355**, 359–361 (1992).

64. Vivat, V. *et al.* A mutation mimicking ligand-induced conformational change yields a constitutive RXR that senses allosteric effects in heterodimers. *EMBO J.* **16**, 5697–5709 (1997).

65. Shiohara, M. *et al.* Effects of novel RAR- and RXR-selective retinoids on myeloid leukemic proliferation and differentiation in vitro. *Blood* **93**, 2057–2066 (1999).

66. Krezel, W. *et al.* RXRγ null mice are apparently normal and compound RXRα+/-/RXRβ-/-/RXRγ-/- mutant mice are viable. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 9010–9014 (1996).

67. Taneja, R. *et al.* Cell-type and promoter-context dependent retinoic acid receptor (RAR) redundancies for RARβ2 and Hoxa-1 activation in F9 and P19 cells can be artefactually generated by gene knockouts. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 6197–6202 (1996).

68. Taneja, R. *et al.* Reexpression of retinoic acid receptor (RAR)γ or overexpression of RARα or RARβ in RARγ-null F9 cells reveals a partial functional redundancy between the three RAR types. *Proc. Natl. Acad. Sci. U. S. A.* **92**, 7854–7858 (1995).

69. Kastner, P. *et al.* Murine isoforms of retinoic acid receptor γ with specific patterns of expression. *Proc. Natl. Acad. Sci. U. S. A.* **87**, 2700–2704 (1990).

70. Brabender, J. *et al. The Role of Retinoid X Receptor Messenger RNA Expression in Curatively Resected Non-Small Cell Lung Cancer 1*. (2002).

71. Li, M. *et al.* Skin abnormalities generated by temporally controlled RXRα mutations in mouse epidermis. *Nature* **407**, 633–636 (2000).

72. Huang, J. *et al. Prostatic Intraepithelial Neoplasia in Mice with Conditional Disruption of the Retinoid X Receptor Allele in the Prostate Epithelium 1*. (2002).

73. Pandolfi, P. P. Oncogenes and tumor suppressors in the molecular pathogenesis of acute promyelocytic leukemia. *Human Molecular Genetics* vol. 10 769–775 (2001).

74. Farias, E. F. *et al.* Retinoic acid receptor α2 is a growth suppressor epigenetically silenced in MCF-7 human breast cancer cells. *Cell Growth Differ.* **13**, 335–341 (2002).

75. Xu, X. C. *et al.* Progressive decrease in nuclear retinoic acid receptor β messenger RNA level during breast carcinogenesis. *Cancer Res.* **57**, 4992–4996 (1997).

76. Hayashi, K. *et al.* Inactivation of retinoic acid receptor β by promoter CpG hypermethylation in gastric cancer. *Differentiation* **68**, 13–21 (2001).

77. Sirchia, S. M. *et al. Endogenous Reactivation of the RAR2 Tumor Suppressor Gene Epigenetically Silenced in Breast Cancer 1*. *CANCER RESEARCH* vol. 62 (2002).

78. De The, H., Del Mar Vivanco-Ruiz, M., Tiollais, P., Stunnenberg, H. & Dejean, A. Identification of a retinoic acid responsive element in the retinoic acid receptor & beta;gene. *Nature* **343**, 177–180 (1990).

79. Sirchia, S. M. *et al.* Evidence of epigenetic changes affecting the chromatin state of the retinoic acid receptor β2 promoter in breast cancer cells. *Oncogene* **19**, 1556–1563 (2000).

80. Yang, Q. *et al.* Biallelic inactivation of retinoic acid receptor β2 gene by epigenetic change in breast cancer. *Am. J. Pathol.* **158**, 299–303 (2001).

81. Widschwendter, M., Berger, J., Müller, H. M., Zeimet, A. G. & Marth, C. Epigenetic downregulation of the retinoic acid receptor-β2 gene in breast

cancer. *J. Mammary Gland Biol. Neoplasia* **6**, 193–201 (2001).

82. Ordóñez-Morán, P., Dafflon, C., Imajo, M., Nishida, E. & Huelsken, J. HOXA5 Counteracts Stem Cell Traits by Inhibiting Wnt Signaling in Colorectal Cancer. *Cancer Cell* **28**, 815–829 (2015).

83. Farmer, P. *et al.* Identification of molecular apocrine breast tumours by microarray analysis. *Oncogene* **24**, 4660–4671 (2005).

84. Kenny, P. A. *et al.* The morphologies of breast cancer cell lines in three-dimensional assays correlate with their profiles of gene expression. *Mol. Oncol.* **1**, 84–96 (2007).

85. Mongan, N. P. & Gudas, L. J. Valproic acid, in combination with all-trans retinoic acid and 5-aza-2′-deoxycytidine, restores expression of silenced RARβ2 in breast cancer cells. *Mol. Cancer Ther.* **4**, 477–486 (2005).

86. Lombardo, Y., Giorgio, A. de, Coombes, C. R., Stebbing, J. & Castellano, L. Mammosphere Formation Assay from Human Breast Cancer Tissues and Cell Lines. *J. Vis. Exp.* **2015**, 52671 (2015).

87. Alsaleem, M. A. *et al.* A novel prognostic two-gene signature for triple negative breast cancer. *Mod. Pathol. 2020 3311* **33**, 2208–2220 (2020).

88. Burstein, M. D. *et al.* Comprehensive genomic analysis identifies novel subtypes and targets of triple-negative breast cancer. *Clin. Cancer Res.* **21**, 1688–1698 (2015).

89. GEO Accession viewer. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=gse12777.

90. Breast Cancer Profiling Project, Gene Expression 1: Baseline mRNA sequencing on 35 breast cell lines - Dataset - HMS LINCS Database - HMS LINCS Project. https://lincs.hms.harvard.edu/db/datasets/20348/.

91. Croker, A. K. *et al.* Differential Functional Roles of ALDH1A1 and ALDH1A3 in Mediating Metastatic Behavior and Therapy Resistance of Human Breast Cancer Cells. *Int. J. Mol. Sci. 2017, Vol. 18, Page 2039* **18**, 2039 (2017).

92. F, M. *et al.* Aldehyde dehydrogenase 1 (ALDH1) expression is an independent prognostic factor in triple negative breast cancer (TNBC). *Medicine (Baltimore).* **96**, (2017).

93. Lukonin, I. *et al.* Phenotypic landscape of intestinal organoid regeneration. *Nat.* | **586**, (2020).

94. Chen, H. *et al.* HOXA5 Acts Directly Downstream of Retinoic Acid Receptor β and Contributes to Retinoic Acid–Induced Apoptosis and Growth Inhibition. *Cancer Res.* **67**, 8007–8013 (2007).

95. Mongan, N. P. & Gudas, L. J. *Valproic acid, in combination with all-trans retinoic acid and 5-aza-2V-deoxycytidine, restores expression of silenced RARb b2 in breast cancer cells*. (2005).

96. Li, B., Huang, Q. & Wei, G. H. The role of hox transcription factors in cancer predisposition and progression. *Cancers (Basel).* **11**, (2019).

97. Laakso, M., Loman, N., Borg, Å. & Isola, J. Cytokeratin 5/14-positive breast

cancer: true basal phenotype confined to BRCA1 tumors. *Mod. Pathol. 2005 1810* **18**, 1321–1328 (2005).

98. Croker, A. K. *et al.* Differential Functional Roles of ALDH1A1 and ALDH1A3 in Mediating Metastatic Behavior and Therapy Resistance of Human Breast Cancer Cells. *Int. J. Mol. Sci. Artic.* doi:10.3390/ijms18102039.

99. Juríková, M., Danihel, Ľ., Polák, Š. & Varga, I. Ki67, PCNA, and MCM proteins: Markers of proliferation in the diagnosis of breast cancer. *Acta Histochem.* **118**, 544–552 (2016).

100. Zanconato, F., Cordenonsi, M. & Piccolo, S. YAP/TAZ at the roots of cancer. *Cancer Cell* **29**, 783 (2016).

101. Shreberk-Shaked, M. & Oren, M. New insights into YAP/TAZ nucleo-cytoplasmic shuttling: new cancer therapeutic opportunities? *Mol. Oncol.* **13**, 1335 (2019).

102. Kim, M.-K., Jang, J.-W. & Bae, S.-C. DNA binding partners of YAP/TAZ. *BMB Rep.* **51**, 126 (2018).

103. LM, Y., C, T.-U., K, W. & P, B. Role of retinoid receptors in the prevention and treatment of breast cancer. *J. Mammary Gland Biol. Neoplasia* **4**, 377–388 (1999).

104. Shilkaitis, A., Green, A. & Christov, K. Retinoids induce cellular senescence in breast cancer cells by RAR-β dependent and independent pathways: Potential clinical implications (Review). *Int. J. Oncol.* **47**, 35–42 (2015).

105. Mezquita, B. & Mezquita, C. Two Opposing Faces of Retinoic Acid: Induction of Stemness or Induction of Differentiation Depending on Cell-Type. *Biomolecules* **9**, (2019).

106. MJ, W., MR, K., YS, C., JY, Y. & CJ, C. Retinoic acid directs breast cancer cell state changes through regulation of TET2-PKCζ pathway. *Oncogene* **36**, 3193–3206 (2017).

107. Khoury, T. *et al.* Aldehyde dehydrogenase 1A1 expression in breast cancer is associated with stage, triple negativity, and outcome to neoadjuvant chemotherapy. *Mod. Pathol. 2012 253* **25**, 388–397 (2011).

9 Appendix

9.1 R2 platform data

9.1.1 Tumor Breast (TNBC) - Brown - 198 - MAS5.0 - u133p2 Dataset

Table 6: Overall, Gene expression across 4 TNBC subtypes

Gene	Description	*P* value
HOXA5	Highly expressed in the **MS** subtype	3.98e-08
CD44	Highly expressed in the **MS** subtype followed by **BL-1**	5.58e-03
ALDH1A1	Highly expressed in the **MS** subtype	2.64e-16
ALDH18A1	Highly expressed in the **BL-2** subtype	3.00e-03
RARA	Highly expressed in **LAR** subtype followed by **MS**	7.76e-07
KRT14	Highly expressed in BL-2 followed by MS subtype, however there is apparent variation in its expression between the samples	2.27e-06
VIM	Highly expressed in the MS subtype	1.39e-12
KRT19	Highly expressed in LAR followed closely by BL-1 and BL-2	1.23e-06
ABCG2	Highly expressed in MS subtype	2.41e-11
NANOG	expression showed high variation within the samples across all subtypes and was not significant	0.191
FOSL1	not significantly expressed in either of the subtypes	0.171
CDH1	Highly expressed in all subtypes	5.50e-12
CDH2	Highly expressed in BL-2 followed by MS, however there is apparent variation in its expression between the tested samples	5.95e-06
CCNA2	Highly expressed in BL-1 and BL-2 subtypes	4.95e-19
MKI67	Highly expressed in BL-1 subtype	6.92e-19

CTNNB1	not significantly expressed in either of the subtypes	0.561
EPCAM	Highly expressed in all subtypes	1.59e-12
CCNB2	Highly expressed in BL subtypes followed by MS but with apparent variation between the tested samples	6.24e-20
FOXM1	Highly expressed in BL subtypes	1.68e-22

Table 7: Gene Correlations to HOXA5 in TNBC

Gene	r- value	*P* value
YAP1	0.161	0.024
ALDH1A1	0.375	9.31e-03

Table 8: Gene Correlations to HOXA5 in MS subtype

Gene	r- value	*P* value
CD44	-0.315	0.031
ALDH1A1	0.375	9.31e-03
ALDH18A1	-0.317	0.030
RARA	-0.058	0.701
KRT14	0.036	0.809
VIM	0.031	0.837
KRT19	-0.111	-0.111
ABCG2	0.501	3.31e-04
NANOG	0.422	3.16e-03
FOSL1	-0.385	7.51e-03
CDH1	-0.139	0.350

CDH2	-0.413	3.96e-03
CCNA2	-0.604	6.88e-06
MKI67	-0.574	2.43e-05
CTNNB1	0.001	1.000
EPCAM	-0.205	0.168
CCNB2	-0.675	1.95e-07
FOXM1	-0.621	3.22e-06

Table 9: Gene Correlations to HOXA5 in BL-1 subtype

Gene	**r- value**	***P* value**
CD44	0.026	0.850
ALDH1A1	-0.122	0.381
ALDH18A1	-0.222	0.107
RARA	0.062	0.654
KRT14	-0.030	0.827
VIM	0.131	0.345
KRT19	-0.091	0.515
ABCG2	0.281	0.040
NANOG	-0.157	0.258
FOSL1	0.038	0.785
CDH1	0.053	0.703
CDH2	-0.046	0.739
CCNA2	-0.124	0.372
MKI67	-0.191	0.168
CTNNB1	-0.008	0.957

EPCAM	-0.091	0.513
CCNB2	0.008	0.955
FOXM1	-0.084	0.545

Table 10: Gene Correlations to HOXA5 in BL-2 subtype

Gene	r- value	*P* value
CD44	0.056	0.673
ALDH1A1	-0.241	0.063
ALDH18A1	0.008	0.953
RARA	0.141	0.283
KRT14	0.177	0.176
VIM	-0.389	2.13e-03
KRT19	0.028	0.833
ABCG2	0.146	0.266
NANOG	0.052	0.692
FOSL1	-0.116	0.378
CDH1	-0.075	0.568
CDH2	-0.156	0.232
CCNA2	-0.061	0.642
MKI67	-0.182	0.163
CTNNB1	-0.066	0.614
EPCAM	-0.052	0.695
CCNB2	-0.072	0.583
FOXM1	-0.136	0.299

9.1.2 Tumor Breast (TNBC) - Brown - 198 - fRMA - u133p2

Table 11: Overall, Gene expression across 4 TNBC subtypes

Gene	Description	*P* value
HOXA5	Highly expressed in the **MS** subtype	4.99e-09
CD44	Highly expressed in the **MS** subtype	0.015
ALDH1A1	Highly expressed in the **MS** subtype but with high variation	6.86e-22
ALDH18A1	Highly expressed in the **BL-2** subtype	2.02e-03
RARA	not significantly expressed in either of the subtypes	0.214
ABCG2	Highly expressed in **MS** subtype	6.24e-12
NANOG	Highly expressed in **LAR** subtype but with high variation	3.35e-04
FOSL1	Highly expressed in **MS** subtype	0.025
CDH1	Highly expressed in all subtypes more in **LAR**	1.44e-12
CDH2	Highly expressed in **BL-2** followed by **MS**, however there is apparent variation in its expression between the tested samples	9.76e-06
CCNA2	Highly expressed in **BL-1** and **BL-2** subtypes	5.11e-19
MKI67	Highly expressed in **BL-1** subtype	3.40e-21
CCNB2	Highly expressed in **BL** subtypes followed by MS but with apparent variation between the tested samples	1.28e-20
FOXM1	Highly expressed in BL subtypes	1.29e-21

Table 12: Gene Correlations to HOXA5 in MS subtype

Gene	r- value	*P* value
CD44	-0.311	0.033
ALDH1A1	0.392	6.45e-03

ALDH18A1	-0.349	0.016
RARA	-0.319	0.029
ABCG2	0.500	3.45e-04
NANOG	0.472	8.05e-04
FOSL1	-0.409	4.33e-03
CDH1	-0.214	0.148
CDH2	-0.322	0.027
CCNA2	-0.614	4.44e-06
MKI67	-0.550	6.31e-05
CCNB2	-0.659	4.86e-07
FOXM1	-0.600	8.39e-06

Table 13: Gene Correlations to HOXA5 in BL-1 subtype

Gene	r- value	*P* value
CD44	0.158	0.229
ALDH1A1	-0.248	0.056
ALDH18A1	0.025	0.849
RARA	0.070	0.596
ABCG2	0.400	2.70e-03
NANOG	-0.115	0.406
FOSL1	-0.047	0.738
CDH1	0.003	0.983
CDH2	0.072	0.603
CCNA2	-0.124	0.370
MKI67	-0.228	0.097
CCNB2	-0.056	0.685
FOXM1	-0.062	0.654
YAP1	0.499	3.52e-04

Table 14: Gene Correlations to HOXA5 in BL-2 subtype

Gene	r- value	*P* value
CD44	0.056	0.673
ALDH1A1	-0.241	0.063
ALDH18A1	0.008	0.953
RARA	0.141	0.283
ABCG2	0.047	0.721
NANOG	-0.086	0.511

FOSL1	-0.191	0.143
CDH1	-0.011	0.933
CDH2	-0.176	0.178
CCNA2	-0.056	0.672
MKI67	-0.107	0.414
CCNB2	0.003	0.983
FOXM1	-0.047	0.723