Institute for Animal Breeding and Genetics
Department of Molecular Genetics,
University of Veterinary Medicine, Vienna
and
Department of Pathology,
Beth Israel Deaconess Medical Center, Harvard Medical School, Boston

Stromal Regulation of long non-coding RNAs in Breast Cancer Metastasis

MASTER'S THESIS
for the degree of
Master of Science (MSc)
University of Veterinary Medicine, Vienna

Submitted by
Johannes Schmöllerl, BSc
1145128 / Biomedicine and Biotechnology

Vienna, December 2013

Internal Supervisors:

Ao. Univ. Prof. Dr. rer. nat. Marina Karaghiosoff

Dr. rer. nat. Sabine Macho-Maschler

Institute for Animal Breeding and Genetics

Department of Molecular Genetics

University of Veterinary Medicine, Vienna

External Supervisor:

Antoine E. Karnoub, Ph.D., Assistant Professor of Pathology

Beth Israel Deaconess Medical Center and Harvard Medical School

Department of Pathology

Boston, USA

Table of Contents

Abbreviations

%	Percent
°C	Degree Celsius
Ad-MSC	Human Adipose-derived MSC
ALDH	Aldehyde dehydrogenase
BAA	BODIPY-amino-acetate
BAAA	BODIPY-amino-acetaldehyde
BCC	Breast cancer cells
BLC	Basal-like Carcinoma
bp	Base pair
CAF	Cancer-associated fibroblasts
CD	Cluster of differentiation
DMEM	Dulbecco's Modified Eagle Growth Media
cDNA	Complementary DNA
cm	Centimeter
MEM-α	Minimum Essential Medium Alpha
CSC	Cancer stem cells
Ct	Cycle threshold
DEAB	Diethylaminobenzaldehyde
DNA	Deoxyribonucleic acid
ECM	Extracellular matrix
EGF	Epidermal growth factor
EMT	Epithelial to Mesenchymal Transition
ER	Estrogen receptor
FACS	Fluorescence-activated Cell Sorting
FBS	Fetal Bovine Serum
FDR	False Discovery Rate
FOXP2	Forkhead box P2
g	Gram
GFP	Green Fluorescent Protein
GM	Growth Medium
GOI	Gene of interest
h	Hour
hBM-MSC	Human bone-marrow derived mesenchymal stem cells
HER2	Human epidermal growth factor 2
IL-6	Interleukin 6
IRB	Institutional Review Board
L	Liter
lncRNA	Long non-coding RNA
LOX	Lysyl oxidase
MCP-1	Monocyte chemotactic protein 1
MET	Mesenchymal to Epithelial Transition
Min	Minutes
miRNA	MicroRNA
ml	Milliliter
mM	Millimolar
mRNA	Messenger RNA
MSC	Mesenchymal stem cells
NC	Non-contact
ng	Nanogram
nm	Nanometer
nt	Nucleotides
NTC	No template control
PBS	Phosphate buffered saline
PCR	Polymerase Chain Reaction
PR	Progesterone receptors
PRC2	Polycomb Repressive Complex 2
qRT-PCR	Quantitative Real Time - PCR
RNA	Ribonucleic acid
rpm	Rotations per minute
rRNA	Ribosomal RNA
RT	Reverse transcription
S.O.C	Super Optimal broth with Catabolite repression
SD	Standard deviation
SEM	Standard error of the mean
shRNA	Small hairpin RNA
siRNA	Small interfering RNA
snoRNA	Small nucleolar RNA
snRNA	Small nuclear RNA
TCGA	The Cancer Genome Atlas
TGF-β	Transforming Growth Factor Beta
tRNA	Transfer RNA
TWIST	Twist-related protein 1
U	Units
V	Volt
α	Alpha
µg	Microgram
µl	Microliter

List of Illustrations

List of Tables

1 Introduction

1.1 Breast Cancer

Breast cancer is the most common cancer disease in women worldwide and the most likely cause for women to die from cancer (BOYLE et al., 2008). It accounts for approximately 23 % (1.38 million) of all diagnosed cancers in women and is responsible for over 400,000 deaths per year (JEMAL et al., 2011). The cancer incidence rate between developed and developing countries is considered fairly similar. However, the mortality rate in developing countries is much higher, which can be explained due to inefficient screening, late diagnosis and inadequate treatment facilities (NAROD, 2012; FERLAY et al., 2010).

1.1.1 Risk Factors

The risk for women to develop breast cancer increases with age and cumulative number of ovarian cycles (BOYLE et al., 2008). Breast cancer is almost non-existent in women below the age of 25, but women above the age of 40 have the highest risk to develop this malignant disease (NAROD, 2012). Pregnancy is thought to temporarily increase the risk of breast cancer due to the mitogenic effect of the high free estrogen levels, that's why women who have given birth have generally a lower risk than women have not (LYONS et al., 2009). Additionally, nursing and lactation is suggested to have a protective effect against breast cancer development, due to the suppression of ovulatory function. The use of exogenous hormones in form of oral contraceptives or hormonal replacement therapy, in contrast, is postulated to increase the breast cancer risk in a time-dependent manner (LA VECCHIA, 2004; 1996; COLLABORATIVE GROUP ON HORMONAL FACTORS IN BREAST, 2002).

Women who have a family history of breast cancer have a 2-3 fold higher risk of developing breast cancer themselves (BOYLE et al., 2008). Although the great majority of breast cancer cases occurs sporadically, about 5-10 % are suggested to be of hereditary origin and mostly linked to mutations in high-penetrance genes, including BRCA1, BRCA2 and TP53 (LUX et al., 2006; BOYLE et al., 2008). Depending on the specific mutations within the BRCA1 and the BRCA2 gene, the

estimated risk for a woman to develop breast cancer is between 40-85 % (FRANK et al., 1998). In the past years, the breast cancer mortality rate could be reduced, which is attributed to increased awareness and better accessibility of screening methods (e.g. mammography) that allow early detection. Prognosis upon breast cancer diagnosis strongly depends on tumor staging and the molecular cancer subtype. Especially breast cancer in younger women remain very challenging and are associated with a high mortality rate, due to the lack of screening at this age and the often very aggressive tumor subtypes (NAROD, 2012).

1.1.2 Classification of Breast Cancer

Malignant tumors that originate from breast epithelial cells are by definition adenocarcinomas. The identification of specific cytological and tissue architectural patterns that are associated with certain clinical presentations and outcomes led to the traditional histological classification (WEIGELT and REIS-FILHO, 2009). It categorizes invasive breast cancers into subgroups based on histological patterns (e.g. ductal, lobular etc.), tumor grade, presence of lymphovascular invasion and distant metastasis. The tumor grade describes the aggressiveness of the cancer cells, by evaluating the degree of differentiation, indicating their ability to build normal-breast like structures and by determining their proliferative activity (ELSTON and ELLIS, 1991). Additionally, the expression of a few predictive markers including hormone receptors for estrogen (ER) and progesterone (PR) as well as amplification of the human epidermal growth factor receptor 2 (HER2) are considered in this traditional classification (WEIGELT and REIS-FILHO, 2009; REDIG and MCALLISTER, 2013).

Histological tumor typing alone, however, has proven to be of limited use in individual treatment, due to the high complexity and heterogeneity of the disease. Better methods are required to help assess prognosis and decide for the most appropriate treatment. Recent advances moved from the exclusive use of the morphological phenotype of breast cancer to molecular biological approaches, in which phenotypes are determined on the basis of gene expression patterns (WEIGELT and REIS-FILHO, 2009).

Recently, microarray-based gene expression analysis has been broadly used to study breast cancer and to determine metastatic potential and features associated with prognosis and therapy response (VAN DE VIJVER et al., 2002; WEIGELT and REIS-FILHO, 2009). These studies led to the development of a new molecular classification of breast cancer that goes beyond the traditional histological and hormone-receptor positive and hormone-receptor negative types. It consists of at least four molecular subtypes including, luminal A (ER-positive, low tumor grade), luminal B (ER positive, high tumor grade), HER2-postive and basal-like breast cancer. The clinical features of each subtype are summarized in Figure 1. These molecular subtypes differ in their gene expression pattern, clinical features, response to treatment and prognosis (SCHNITT, 2010; WEIGELT and REIS-FILHO, 2009; PEROU et al., 2000). The characterization of basal-like breast cancers as an individual subtype is one of the most novel findings. Frequently, this specific type is negative for ER, PR and HER2 expression and is therefore referred to as “triple-negative”. This specific subtype is often highly aggressive and associated with poor prognosis, due to the lack of possible targeted receptor therapy (SCHNITT, 2010; CRISCITIELLO et al., 2012).

	Molecular subtype		
	Luminal	***HER2***	***Basal***
Gene expression pattern	High expression of hormone receptors and associated genes (luminal A>luminal B)	High expression of HER2 and other genes in amplicon Low expression of ER and associated genes	High expression of basal epithelial genes, basal cytokeratins Low expression of ER and associated genes Low expression of HER2
Clinical features	~70% of invasive breast cancers ER/PR positive Luminal B tend to be higher histological grade than luminal A Some overexpress HER2 (luminal B)	~15% of invasive breast cancers ER/PR negative More likely to be high grade and node positive	~15% of invasive breast cancers Most ER/PR/HER2 negative (‘triple negative’) *BRCA1* dysfunction (germline, sporadic) Particularly common in African-American women
Treatment response and outcome	Respond to endocrine therapy (but response to tamoxifen and aromatase inhibitors may be different for luminal A and luminal B) Response to chemotherapy variable (greater in luminal B than in luminal A) Prognosis better for luminal A than luminal B	Respond to trastuzumab (Herceptin) Respond to anthracycline-based chemotherapy Generally poor prognosis	No response to endocrine therapy or trastuzumab (Herceptin) Appear to be sensitive to platinum-based chemotherapy and PARP inhibitors Generally poor prognosis (but not uniformly poor)

Figure 1 Summary - Molecular Subtypes of Breast Cancer
Adapted from (SCHNITT, 2010)

1.2 Breast Cancer Metastasis

In the past years, the rate of breast cancer mortality could be reduced due to prophylactic screening of high-risk patients, which led to earlier detection of cancer. Additionally, the development of targeted therapies against HER2 positive and estrogen-receptor positive breast cancers helped reduce the number of breast cancer related deaths (BERRY et al., 2005). As highlighted earlier, triple-negative breast cancer is often very aggressive, characterized by an early reoccurrence upon initial treatment and metastases are frequently found within the lung or brain tissue (CRISCITIELLO et al., 2012). Most breast cancer related deaths are due to distant tumor metastases and therefore, metastatic patients are often considered as incurable (REDIG and MCALLISTER, 2013). Although only few women are initially diagnosed with late stage breast cancer, it is estimated that in approximately 30 % of all breast cancer cases, women will develop metastatic lesions, eventually (O'SHAUGHNESSY, 2005).

The development of metastases is a multistep process that involves, the loss of cellular adhesion, invasiveness, entry and survival in the circulation, extravasation and ultimately results in the colonization of cancer cells at a foreign tissue (GUPTA and MASSAGUE, 2006). During this process, malignant cancer cells have to undergo several different dynamic phenotypical changes in order to successfully metastasize into foreign tissue. Carcinoma is per definition a malignant solid tumor that derives from transformed epithelial cells. However, in order to metastasize, these neoplastic epithelial cancer cells have to change phenotypically to a more mesenchymal state to acquire new characteristics that aid in their progression (TAM and WEINBERG, 2013). The phenotypical shift of epithelial cells to a mesenchymal state is called “Epithelial to Mesenchymal Transition” (EMT).

1.2.1 Epithelial to Mesenchymal Transition

During the process of EMT epithelial cells that normally maintain close contact to neighboring cells due to tight junctions, desmosomes and adherence junctions, loose their traits and shift to a mesenchymal cell phenotype. This is characterized by loss of cell-cell contact, loss of cell polarity, and the gain of migratory and invasive properties (THIERY et al., 2009; POLYAK and WEINBERG, 2009). However, EMT is not an irreversible process, hence the mesenchymal cells can revert back to an epithelial phenotype during mesenchymal-to-epithelial transition (MET) (LEE et al., 2006). Both, EMT and MET are physiological processes that are crucial during embryonic development of most internal organs. However, during carcinoma progression, cancer cells are able to mimic these processes as they contribute to their spread throughout the human body (THIERY et al., 2009).

The loss of E-cadherin, a crucial cell-adhesion protein, is considered one major event in the progression of the EMT program (THIERY et al., 2009). Transcription factors that induce EMT can be divided into two different classes, both result in the repression of E-cadherin. The first class of transcription factors (e.g. SNAIL, SLUG, ZEB1, ZEB2, and KLF8) directly represses E-cadherin expression by binding to the E-cadherin promoter, whereas the second class (e.g. TWIST and FoxC2) indirectly represses transcription (THIERY et al., 2009; PEINADO et al., 2007).

However, these EMT-inducing transcription factors do not only target E-cadherin, but rather pleiotropically lead to the repression of many other junctional proteins. These include claudins and desmosomes that ultimately lead to the phenotypical shift of epithelial to mesenchymal cells (DE CRAENE and BERX, 2013). Indeed, endogenous or forced expression of these EMT-inducing transcription factors in carcinoma cells has been linked to loss of E-cadherin and the up regulation of the mesenchymal genes including Vimentin and N-cadherin (DE CRAENE and BERX, 2013). The diversity of transcription factors that may induce EMT and their differential expression within the cancer cell is postulated to be responsible for the phenotype of some cancer cells that reside in an intermediate state between epithelial and mesenchymal. This is referred to as partial EMT (DE CRAENE and BERX, 2013) (Figure 2).

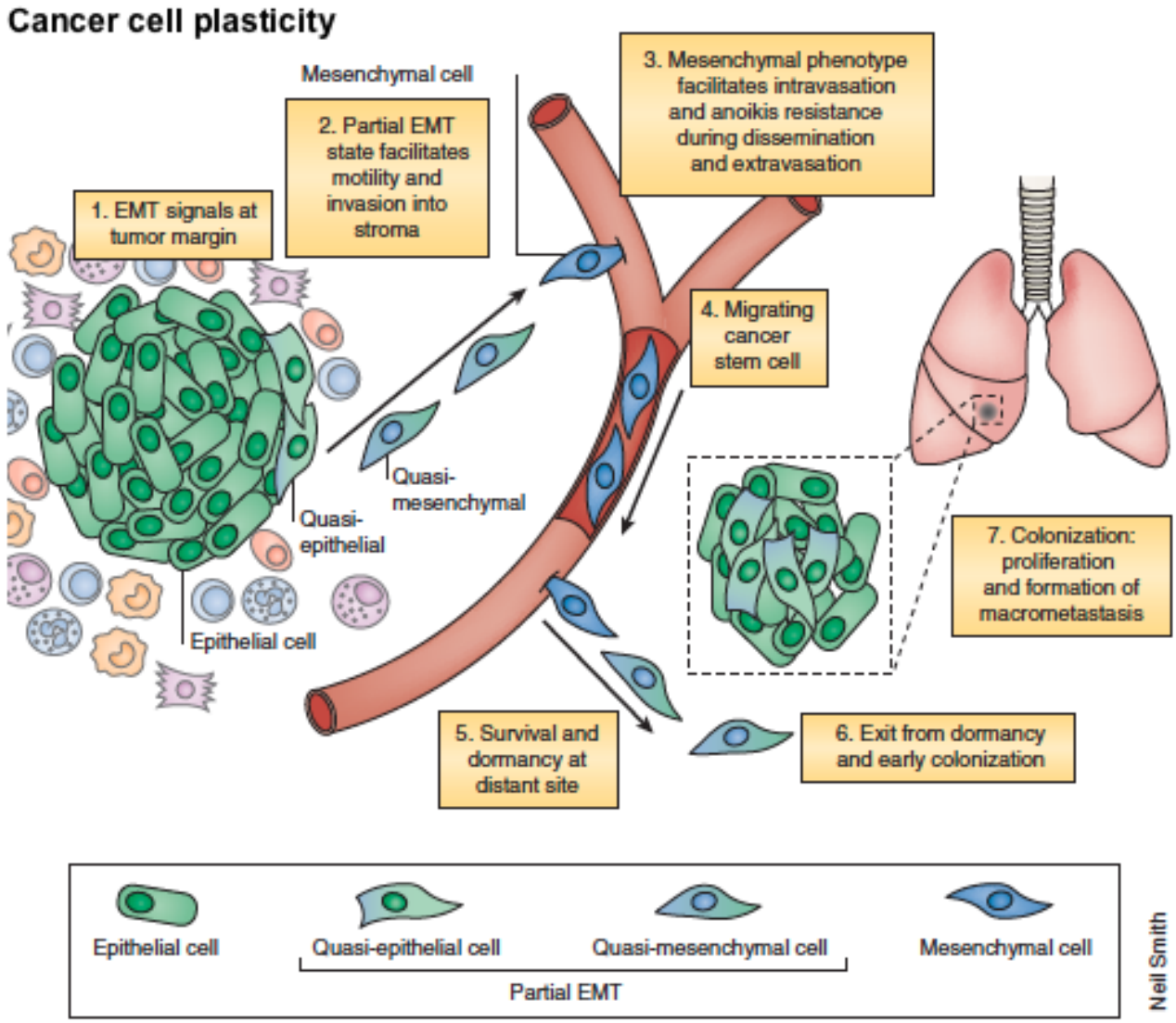

Figure 2 EMT in Metastasis
Adapted from (TAM and WEINBERG, 2013)
Cancer cells at the invasive front undergo partial EMT by factors that may be induced by the surrounding tumor stroma. During partial EMT, the cancer cells acquire traits that help them to further migrate and invade into the blood stream. When completely depleted of the signaling derived from the primary tumor environment, these cells may become completely mesenchymal and survive within the circulation and invade foreign tissue. Upon arrival at the distant side, cancer cells may undergo MET and re-acquire epithelial traits induced by signaling from new environment (TAM and WEINBERG, 2013).

Cancer cells undergo different stages of the EMT program, thereby acquiring new characteristics that aid in their progression. Because of this so-called partial EMT, invading cancer cells rarely lose all epithelial traits, but rather co-express them with new acquired mesenchymal markers (TAM and WEINBERG, 2013). Often, these partial transitions of epithelial cancer cells are induced by contextual signals from

surrounding stromal cells of the tumor microenvironment (TAM and WEINBERG, 2013).

1.2.2 Influence of the Tumor Microenvironment

Breast cancer is not a homogeneous cluster of genetically altered tumor cells, but rather a construct of complex interplay between cancer cells and their microenvironment. This environment consists of different cell types including cancer-associated fibroblasts (CAF), pericytes, endothelial cells, inflammatory cells, mesenchymal stem cells (MSC) and extracellular matrix (QUAIL and JOYCE, 2013) (Figure 3). During tumor development, many of these cell types are attracted by the cancer cells, this process significantly alters the composition of the tumor milieu and thereby favors cancer progression (CUIFFO and KARNOUB, 2012).

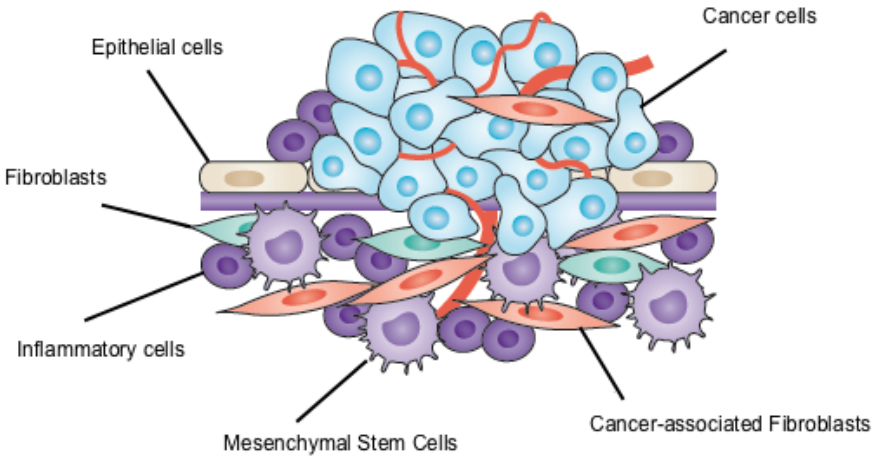

Figure 3 The Tumor Microenvironment
Adapted from (BARCELLOS-HOFF et al., 2013)
The breast cancer microenvironment consists of different cell types, including inflammatory cells, cancer-associated fibroblasts and mesenchymal stem cells. The developing inflammatory milieu due to the cancer lesion is suggested to promote tumor progression (GRIVENNIKOV et al., 2010). Crosstalk between cancer cells and various stromal-derived cells contributes to EMT at the primary tumor site, thereby facilitating cancer metastasis (QUAIL and JOYCE, 2013).

During the last years it became obvious that the crosstalk between stromal cells and malignant cancer cells is crucial for tumor growth and progression (EGEBLAD et al., 2010). For example, CAF, which have a mesenchymal phenotype, are the predominant cell type within the breast cancer microenvironment and were shown to

promote metastasis (DUMONT et al., 2013). The true origin of CAF is still unknown, but it is suspected that they may arise from endothelial cells through endothelial to mesenchymal transition (EMT) or alternatively, might differentiate from MSC (ZEISBERG et al., 2007; MISHRA et al., 2008).

Importantly, stromal-derived signaling within the tumor microenvironment actively contributes to the activation and acquisition of the malignant phenotypes of cancer cells. It is postulated that extracellular ligands maintain the EMT program of cancer cells (ACLOQUE et al., 2009). The transforming growth factor beta (TGF-β) is a multifunctional cytokine that normally exhibits tumor-suppressing functions and regulates homeostasis of mammary epithelial cells (TAYLOR et al., 2010). Paradoxically, the same cytokine leads to oncogenic cell behaviors when exposed to neoplastic cells (TAYLOR et al., 2010). TGF-β is a potent inducer of EMT during embryonic development and plays a major role in cancer progression (TAM and WEINBERG, 2013). CAF are suggested to participate in EMT progression by TGF-β mediated crosstalk with cancer cells (VAN ZIJL et al., 2009).

1.2.3 Mesenchymal Stem Cells within the Tumor Stroma

Of particular interest, and currently under extensive investigations are the contributions of MSC to cancer progression. These multipotent, self-renewing progenitor cells give rise to different cell types, including adipocytes, osteocytes and chondrocytes. Under physiological conditions these stem cells home to sites of tissue injury and inflammation and promote repair and tissue regeneration (EL-HAIBI and KARNOUB, 2010). The biggest population of MSC can be found within the bone marrow but they can also be isolated from other tissues including fat, muscle, skin, cartilage, amniotic fluid and the umbilical cord (BERNARDO et al., 2009). As these stem cells seem to have diverse origins, they do not express a single uniform expression marker that would allow specific discrimination. Instead, they share a common set of surface antigens, including CD166, CD90, CD105, CD147, CD49c and CD29 (EL-HAIBI and KARNOUB, 2010). Importantly, MSC derived from different origins are not functionally identically. This was shown by their varying differentiation propensities, indicating that their physiological impact might be tissue dependent (EL-HAIBI and KARNOUB, 2010).

During cancer progression, it is believed that MSC migrate into the tumor microenvironment, similarly to their migration into injured tissue (SPAETH et al., 2008). Many soluble factors secreted by cancer or stromal cells in the surrounding inflammatory environment are responsible for the recruitment of bone-marrow-derived MSC, including monocyte chemotactic protein 1 (MCP-1), interleukin 6 (IL-6), TGF-β, vascular endothelial growth factor (VEGF) (CUIFFO and KARNOUB, 2012) (Figure 4). Upon recruitment, these bone-marrow-derived MSC activate cancer cells promote metastasis, which most likely happens through paracrine and contact-dependent mechanisms (KARNOUB et al., 2007; MANDEL et al., 2013).

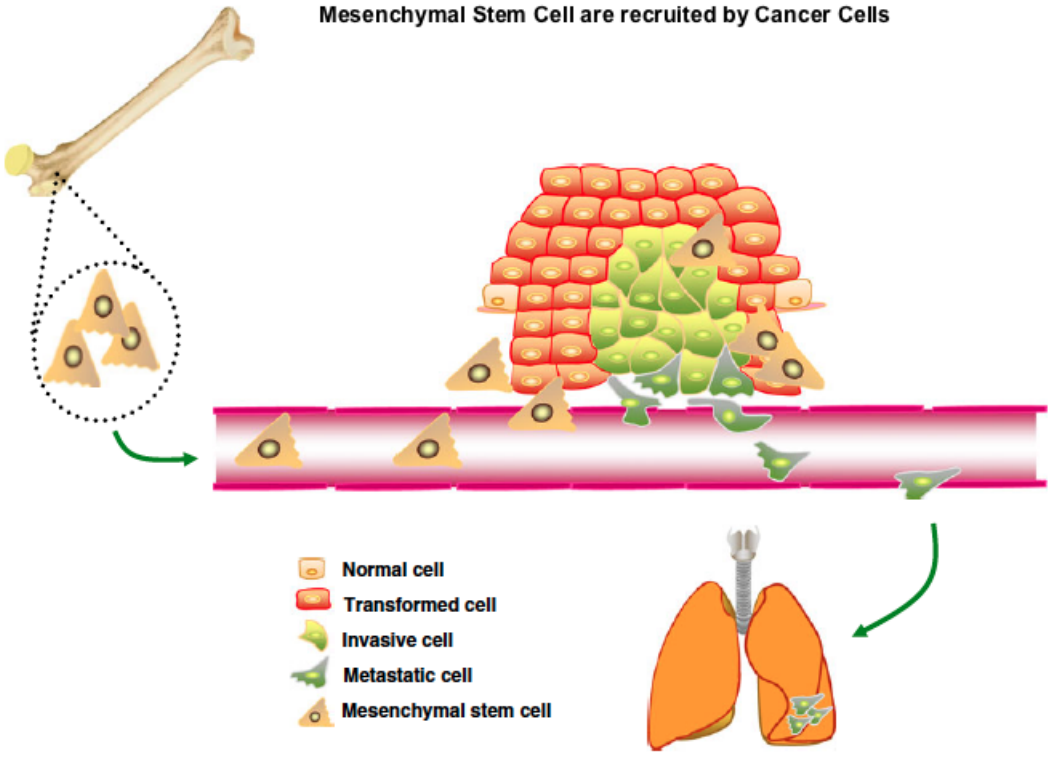

Figure 4 Migration of MSC to the Tumor Microenvironment
Adapted from (EL-HAIBI and KARNOUB, 2010)
MSC migrate to the tumor site as a response to chemoattracting signals, including TGF- β, IL-6, MCP-1, VEGF, derived from the tumor microenvironment. Upon interaction with cancer cells, MSC produce factors themselves and activate cancer cells that lead to their increased metastatic phenotype (EL-HAIBI and KARNOUB, 2010).

The exact mechanism underlying this pathology is not completely understood, however it was shown that certain MSC-derived chemokines, like CCL5, increase motility, invasion and metastasis of cancer cells (KARNOUB et al., 2007). Moreover, it was shown that upon physical interaction of MSC with cancer cells, the malignant

cells at the tumor margin may up-regulate expression of lysyl oxidase (LOX). This is a copper-dependent amine oxidase that triggers TWIST expression and results in EMT of cancer cells (EL-HAIBI et al., 2012). Additionally, very recent data suggests that MSC trigger the expression of a microRNA network within breast cancer cells, which results in the formation of cancer stem cells (CSC), therefore both, MSC-induced EMT and CSC formation are hypothesized to contribute to breast cancer metastasis (Cuiffo BG, unpublished data).

1.2.4 Breast Cancer Stem Cells

Cancer Stem Cells are suggested to be a rare subset of cancer cells with stem-cell-like abilities (e.g. self-renewal and multi-lineage differentiation potential). They are assumed to be responsible for tumor heterogeneity and cancer relapse upon chemotherapy resistance (MEDEMA, 2013). In support of this theory it was shown that remaining tumor cells after breast cancer therapy seem to be enriched for CSC (LI et al., 2008). One explanation for this might be that most cytotoxic substances, which are designed to kill proliferative cells, destroy most of the tumor mass but are assumed to spare CSC, which therefore might lead to cancer relapse (RICCI-VITIANI et al., 2009) (Figure 5).

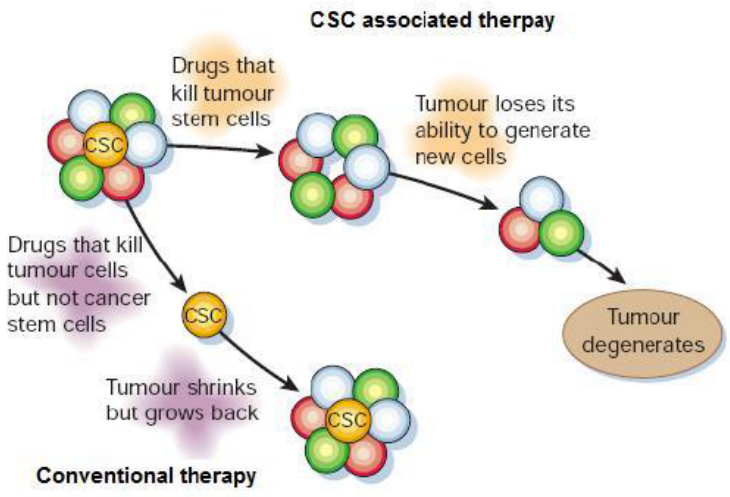

Figure 5 Treatment of CSC
Adapted from (REYA et al., 2001)
Conventional therapy might spare the CSC population within the tumor population, which might result in cancer reoccurrence. CSC-associated therapy might be able to inhibit the self-renewing capacity of cancer stem cells forcing them into differentiation, which would result in tumor degradation.

The current model suggests that breast cancer stem cells express high levels of the surface antigen CD44 and low levels of CD24. This finding is based on the observation that a small population of $CD44^{+}/CD24^{low}$ cancer stem cells was found to

have a higher tumor initiating capacity than the heterogeneous bulk of breast cancer cells when injected into immunosuppressed mice (AL-HAJJ et al., 2003).
Recently, Aldehyde Dehydrogenase 1 (ALDH1) was identified as novel marker for cancer stem cells. ALDH1 activity is suggested to play a role in stem cell maintenance and might even be involved in protecting CSC from chemotherapy (MEDEMA, 2013). Importantly, the abundance of the CSC population within cancer might indicate the potential to metastasize (VERMEULEN et al., 2012).

1.3 Long Non-Coding RNAs in Cancer Progression

1.3.1 Long Non-Coding RNAs

The human genome contains approximately 20,000 protein-coding genes. It is estimated that about 70 % of the human genome is transcribed, however protein-coding genes account only for about 2 % in total (GUTSCHNER and DIEDERICHS, 2012). Historically, these non-coding transcripts were dismissed as background noise, but this view shifted greatly over the last years. Classical non-coding transcripts that have important cellular function are structural RNAs e.g. ribosomal RNAs (rRNA), which build the ribosomal subunits, or transfer RNAs (tRNA), which transport the nascent polypeptide chain (WAHLESTEDT, 2013).

In general, non-coding RNAs can be discriminated based on their size into two classes. Transcripts that are smaller than 200 nt are commonly referred to as small non-coding RNA. The class of small non-coding RNA comprises various subtypes, including small nuclear RNA (snRNA), small nucleolar RNA (snoRNA), PIWI-interacting RNA, small interfering RNA (siRNA) and microRNA (miRNA) (WAHLESTEDT, 2013). The other class, long non-coding RNAs (lncRNA), are longer than 200 nt and are loosely defined as RNA transcripts that lack an open reading frame of significant length, thus do not translate into protein (GUTSCHNER and DIEDERICHS, 2012).

Although lncRNAs are the largest class of non-coding RNAs and account for a great part of the transcriptome, research in the past focused on small non-coding RNAs. Interestingly, many miRNAs were shown to play an important role in development

and progression of breast cancer (MA et al., 2007; VALASTYAN et al., 2009; MA et al., 2010). However, due to advances in technology that allowed large-scale detection of lncRNAs, these transcripts recently gained more attention. Many lncRNAs were discovered by sequencing of cDNA libraries or by using tilling arrays. Most recently, RNA-sequencing became the method of choice for high throughput screening (ULITSKY and BARTEL, 2013).

1.3.2 Biological Function of Long non-coding RNAs

It is estimated that the human genome generates thousands of different lncRNA transcripts. Many lncRNA sequences seem to be transcribed by RNA-polymerase II and undergo similar post-transcriptional processing as mature mRNA, including capping, splicing and polyadenylation (ULITSKY and BARTEL, 2013). Nevertheless, the vast majority of these transcripts has not yet been functionally annotated. The biological relevance of lncRNAs in the cell is not clear and many are suspected to be not functional. However, in a hypothetical scenario it is estimated that if only 10 % of all lncRNAs are functionally relevant, this would still result in more than 1000 transcripts with biological roles within the cell (ULITSKY and BARTEL, 2013).
Data that might indicate the functional relevance of lncRNAs are conservation and transcriptional regulation. Although the sequence of most lncRNAs is not conserved between species, it is argued that their folding and the resulting secondary structure might be highly conserved and of higher importance for this RNA transcripts (MATHEWS et al., 2010). As lncRNAs contain a large class of diverse molecules, it is not possible to describe their biological function in a uniform way. However, different models have been proposed that indicate their potential mode of action (Figure 6).

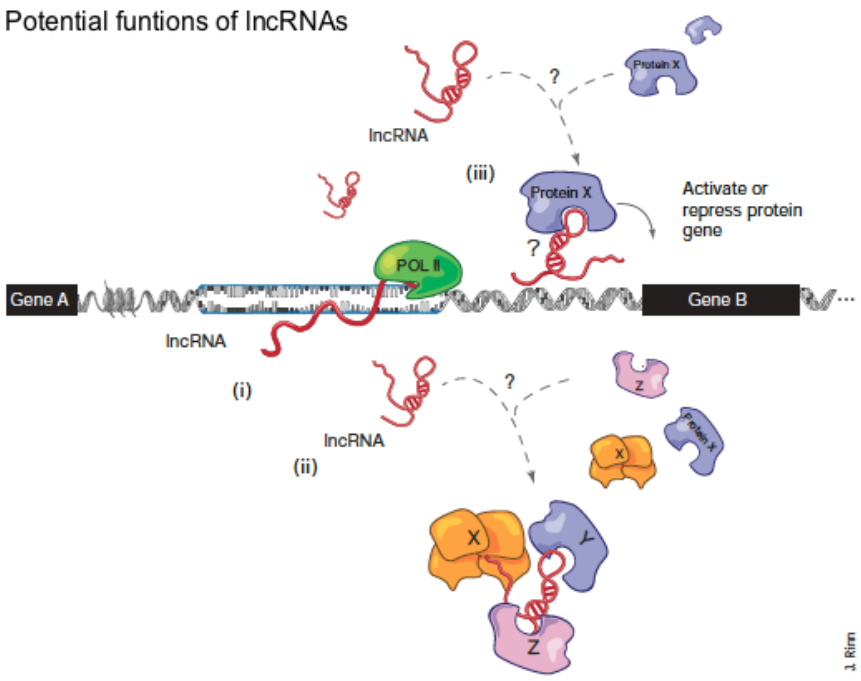

Figure 6 Potential Functions of LncRNA
Adapted from (BAKER, 2011)
LncRNAs may have diverse functions within the cell. While they could just be a byproduct of transcription (i), they could act as a scaffold to link different proteins together (ii) or guide proteins to certain locations within the genome (iii).

LncRNAs may have diverse functions and are suggested to alter gene expression at different levels within the cell, including chromatin modification, transcription and post-transcriptional processing (GUTSCHNER and DIEDERICHS, 2012). Many lncRNAs were shown to act as a guide, which associates with chromatin-modifying complexes to affect gene expression (KHALIL et al., 2009). However, by providing binding surfaces, they can also act as a molecular scaffold. They can assemble for example histone modification enzymes to alter the modification pattern on target genes and thereby their transcription (TSAI et al., 2010).

It is also suggested that some lncRNAs may act as effectors that bind to proteins, thereby changing their structure and their function (WANG et al., 2011).

1.3.3 LncRNAs in Cancer Progression

One common approach to determine the function of lncRNAs within the cell is to evaluate their expression pattern under different conditions and thus to identify a potential regulation mechanism (GUTSCHNER and DIEDERICHS, 2012). By systematic knockdown it could be shown that specific lncRNAs affect global gene expression of mouse embryonic stem cells, regulate maintenance of their

pluripotency and are functionally down regulated upon their differentiation (GUTTMAN et al., 2011).
Similar to protein-coding transcripts, it was shown that de-regulated lncRNAs contribute to cancer progression. Therefore, it is not surprising that many lncRNAs were shown to be differently expressed in solid tumors and leukemia (CALIN et al., 2007). The lncRNA p21 is induced upon p53 expression and leads to the repression of various genes due to the guidance of transcriptional repressor proteins (HUARTE et al., 2010).

Moreover, certain lncRNAs, including MALAT-1 and HOTAIR, are overexpressed in many malignant cancer diseases and were shown to strongly contribute to their progression (GUPTA et al., 2010; SCHMIDT et al., 2011). Also, high expression levels of these lncRNAs indicate an increased potential to develop metastases and strongly correlate with poor patient's prognosis (GUPTA et al., 2010; SCHMIDT et al., 2011).
Most importantly, by functional knockdown of these lncRNAs, cancer cells displayed decreased migration and invasiveness, thereby providing a new potential therapeutic target (GUPTA et al., 2010; REN et al., 2013).

The expression pattern of lncRNAs in many malignant diseases is currently under extensive investigation and will most likely lead to the detection of novel biomarkers for disease progression, prognosis and may even provide potential targets for future therapy approaches (TSAI et al., 2011).

2 Aims of the Thesis

The thesis is divided into two inter-related projects, both describing the involvement of long non-coding RNA induction in breast cancer cells following their interactions with certain tumor-associated stromal cells, called mesenchymal stem cells (MSCs).

The first project describes the identity and role of novel and previously un-described long non-coding RNAs in MSC-primed and highly metastatic breast cancer cells. Microarray screening approaches were utilized to probe for deregulated lncRNAs in MSC-activated cancer cells, and candidate hits were then investigated in terms of their regulation and their mode-of-action. To describe their potential function in breast cancer progression, the novel sequences were cloned into a lentiviral shuttle construct and activities and phenotypes of breast cancer cells overexpressing such constructs were probed using a variety of *in vitro* assays.

The aim of the second project was to describe mechanisms that lead to the expression of the lncRNA HOTAIR in breast cancer cells and to investigate if it plays roles in mesenchymal-stem-cell-catalyzed breast cancer progression. HOTAIR is known to promote cancer metastasis and is associated with poor prognosis in a variety of malignant diseases. However, the induction in cancer cells and the molecular mechanisms by which HOTAIR promotes its malignant function are poorly understood. Therefore, the aims of the second project were to elucidate mechanisms by which MSC may induce HOTAIR expression in breast cancer cells and how this particular lncRNA promotes cancer progression.

3 Materials and Methods

Specifics about all reagents, equipment and other commercial products, including company and catalogue number can be found in the appendix (Chapter 10.3).

3.1 Cell lines, Media and Cell Culture Techniques

All cell culture procedures were performed in a laminar flow tissue-culture hood under sterile conditions. The cells were incubated and maintained at 37 degree Celsius (°C) under 5 % CO_2.

3.1.1 Tissue Culture Growth Media

Dulbecco's Modified Eagle Growth Medium (DMEM)

DMEM with 4.5 g/L glucose (including L-glutamine + sodium pyruvate) was supplemented with:

- 10 % Fetal bovine serum (regular, not heat inactivated)
- 50 U/ml Penicillin
- 50 µg/ml Streptomycin

DMEM/F12 Growth Medium (DMEM/F12)

DMEM with 4.5 g/L glucose (including L-glutamine + sodium pyruvate) and
Ham's F12 medium (including L-glutamine) were mixed in a 1:1 ratio and supplemented with:

- 5 % Horse donor serum
- 20 ng/ml Epidermal growth factor (EGF)
- 0.5 µg/ml Hydrocortisone
- 125 ng/ml Cholera Toxin
- 10 µg/ml Insulin
- 50 U/ml Penicillin
- 50 µg/ml Streptomycin

Minimum Essential Medium Alpha (MEM-α) Growth Medium

The tissue culture medium MEM-α was supplemented with

- 10 % FBS premium-select (not heat inactivated)
- 1 % L-Glutamine 200 mM
- 50 U/ml Penicillin
- 50 µg/ml Streptomycin

3.1.2 Culture of Normal and Breast Cancer Cell Lines

Growth Medium (GM)	Cell Line
DMEM	MDA-MB-231 *(Human epithelial breast cancer cell line)* 293T/17 *(Human epithelial breast cancer cell line)* WI-38 (*Human embryonic lung fibroblasts)*
DMEM/F12	MCF10A *(Human breast epithelial cell line)*
MEM-α	Ad-MSC *(Human adipose-derived MSC)* hBM-MSC *(Human bone-marrow-derived MSC)*

Table 1 Growth Media for respective Cell Lines

MSC were cultured and handled as described in (EL-HAIBI et al., 2012).

3.2 Co-culture Experiments and Cell Sorting

3.2.1 Co-culture of Mesenchymal Stem Cells and Breast Cancer Cells

Human bone marrow derived MSC were cultured together with Green Fluorescent Protein (GFP) expressing MDA-MB-231 breast cancer cells in a ratio of 3:1 in cell culture dishes with a diameter of 15 cm. Briefly, MSC were thawed, re-suspended in MEM-alpha culture media and counted. Cultured MDA-MB-231 cells were trypsinized and counted as well. Then, 1×10^6 MSC and 3.33×10^5 MDA-MB-231 cells were cultured together in a 15 cm diameter dish in 25 ml MEM-α media for 72 hours at 37°C and 5 % CO_2 in a standard tissue culture incubator. As a negative control experiment, 1×10^5 GFP expressing MDA-MB-231 cells were cultured alone in a 15 cm diameter dish in 25 ml complete MEM-α media for 72 hours in the same conditions as described above. Co-culture and negative control culture experiments were conducted in triplicates.

3.2.2 Fluorescence-Activated Cell Sorting

Following co-culture, the tissue culture media were collected and cells were washed with 1x phosphate buffered saline (PBS). Cells were detached using 5 ml trypsin and transferred into a 50 ml Falcon tube. The cell suspension was centrifuged for 3 minutes at 1200 rotations per minute (rpm) and the supernatant was removed. The cells were washed at least two times with 4 ml PBS, then re-suspended in 0.5 ml PBS in FACS tubes and stored on ice.

Cell sorting was performed by using a FACS Aria II sorting machine at the Flow Cytometry Core Facility (Beth Israel Deaconess Medical Center, Boston, USA). Only the strongest GFP-positive MSC-activated MDA-MB-231 cancer cell population from the co-culture experiment was collected. The negative control GFP-positive MDA-MB-231 cells were used to determine sorting gates based on cell size and fluorescence intensity. Sorted MSC-activated MDA-MB-231 cells (231*MSC) and negative control cells (MDA-MB-231 Non-Contact) were centrifuged and washed twice with PBS. Cell pellets were re-suspended in 700 microliter (µl) QIAzol lysis reagent and stored at -80°C until RNA isolation.

3.3 Transwell Assays

For the transwell assays, a 24-well plate with 0.4 µm pore size cell culture inserts was incubated with 700 µl MEM-α per well for 20 min at 37°C for pH equilibration. During the incubation MSC and MDA-MB-231 were counted. Approximately 5 x 10^3 MDA-MB-231 cells were seeded into each well of the 24-well plate. The cell culture inserts of each well were filled with 250 µl MEM-α containing either 5 x 10^3 MDA-MB-231 cells, 1.5 x 10^4 MSC or both. The plates were incubated for 72 h at 37°C and 5 % CO_2 before the cell-culture inserts were removed. Then, MDA-MB-231 cells on the bottom of the 24-well plate were used for RNA isolation.

3.4 Thin Layer Collagen Coating of Cell Culture Dishes

Human Collagen I was used to coat the surface of 12-well plates for subsequent culture of MDA-MB-231 breast cancer cells. The thin layer coating was performed with either 0.75 µg/cm^2 or 0.25 µg/cm^2 Collagen I per well. Briefly, the Collagen I solution with was thawed on ice and 2.85 µg or 0.95 µg, respectively, were suspended in 500 µl 2 mM HCL to cover the 3.8 cm^2 surface area of a 12-well plate. Once the growth surface area was covered, the tissue culture plates were incubated for 2 hours at room temperature. Following incubation, the excess Collagen I was removed using a sterile Pasteur pipette under the tissue culture hood and plates were allowed to air dry. Then, 1x10^4 MDA-MB-231 were re-suspended in 1 ml cDMEM and added to Collagen I coated 12-well plates or as a control to uncoated plates. Following incubation for 72 h, cells were washed twice with 1x PBS, lysed using 700 µl QIAzol and RNA was isolated as described earlier.

3.5 Development of Overexpression Cell Lines

3.5.1 Transfection of 293T/17 Cells

For the production of lentiviral particles that were used to generate stable cell lines, 293T/17 had to be transfected with packing, envelope and a transfer plasmid. In this context the transfer plasmids refer to either PLJM1-GFP, PLJM1-HOTAIR or PLJM1-lnc-KCNJ9-2:2. First, three tissue culture plates (diameter of 6 cm) were seeded with 1 x 10^6 293T/17 cells in 4 ml GM and incubated overnight. The next day, transfection mixtures were prepared using 5 µg transfer plasmid, 0.1 µg VSVG envelope plasmid, 1 µg PAX packaging plasmid and 20 µl of FUGENE transfection reagent. Serum-free DMEM was added up to a total volume of 250 µl. The transfection mixtures were incubated for 5 min at room temperature. In the meantime, the 293T/17 cells were washed with 1x PBS and 3.75 ml GM was added to each plate. Following incubation, the transfection mixtures were added in a dropwise manner and the cells were incubated for 72 hours at 37°C and 5% CO_2. Viral supernatant was collected from each plate and filtered using a 0.45 µm micron filter.

3.5.2 Infection of Recipient Cell Lines

For each cell line three tissue culture plates (diameter of 6 cm) were seeded with 1 x 10^6 recipient cells and incubated overnight. On the next day, the tissue culture medium was removed and cells were given 1 ml fresh GM. Then, 1 ml of filtered viral supernatant was used to infect each recipient cell line, respectively. Also, 10 µg/ml polybrene was added to each plate to increase infection efficiency. The cells were incubated for 24 hours at 37°C and 5 % CO_2, Then, the tissue culture medium was replaced with 4 ml fresh GM and the cells were incubated for another 24 hours. Finally, the infected cells were selected using 0.5 µg/ml puromycin.

3.5.3 Verification of Overexpression Cell Lines

As the control vector (PLJM1-GFP) leads to the expression of GFP, the efficiency of the transfection into 293/17 cells and the infection of recipient cells were verified by fluorescence microscopy. The expression levels of HOTAIR and lnc-KCNJ9-2:2 were verified by qRT-PCR after puromycin selection of the cells.

3.6 ALDH Stem Cell Identification Assay

3.6.1 Principle

Aldehyde dehydrogenase 1 (ALDH-1) is highly expressed in embryonic tissues and can be used as a marker for stem cells (MOREB, 2008). The ALDEFLUOR stem cell identification assay is based on the fact that these cell populations, which have a high ALDH enzyme activity, will metabolize BODIPY-amino-acetaldehyde (BAAA). This fluorescent substrate diffuses into viable cells, will be converted into BODIPY-amino-acetate (BAA) and retained. Because the amount of fluorescent product is proportional to ALDH enzyme activity, highly expressing cells can be identified (MOREB, 2008).

3.6.2 ALDEFLUOR Assay

Enzyme activity was measured using the ALDEFLUOR Kit (Stemcell Technologies). The control and test cells were detached using trypsin, re-suspended into a 15 ml Falcon tube using cDMEM and counted. For both, the control and the test cell line, four samples were tested, each using 1 x 10^5 cells. One sample of each cell line was used as internal control. The cells were transferred into 5 ml polystyrene round-bottom tubes, washed twice with 1x PBS and re-suspended into 0.5 ml ALDH buffer. Into each sample, 5 µl fluorescent ALDH substrate was added. Additionally, the internal control sample of each cell line received 5 µl diethylaminobenzaldehyde (DEAB). This is a specific inhibitor of ALDH enzymatic activity and was used to determine the level of background fluorescence.

All samples were incubated for 30 min in a water bath at 37°C. Then tubes were centrifuged shortly to pellet the cells, and the supernatant was removed. All samples were re-suspended into 0.5 ml ice-cold ALDH buffer and the fluorescence was measured using the FITC channel of the BD FACS Canto II flow cytometer. Analysis was performed with FlowJo (Version 10.0.6).

3.7 Microarray Probe Set Verification

The Affymetrix HT-HG U133 2.0 Plus Array is used for analyzing whole human genome expression and is not restricted to lncRNA sequences. Therefore, it may not be as specific as dedicated non-coding RNA microarrays. Each Affymetrix probe set that is used for the detection of transcripts consists of 11 short oligonucleotides that are usually 25 nucleotides (nt) in length. Ideally, each of these sequences should be specific for the transcript they were designed to detect.

For the purpose of verification, each probe set that led to the discovery of a lncRNA transcript was analyzed for its specificity using the BLAST-like alignment tool (BLAT) (KENT, 2002). If the alignment led to more than 1 specific result, the oligonucleotide sequence was considered non-specific.

Also if more than 50 percent of the 11 sequences of one probe set were non-specific then the whole probe set was considered non-specific and the associated lncRNA not further tested. The full list of probe sets used to determine the nine lncRNAs that were found to be significantly up-regulated by microarray screening and their verification can be found in the appendix (Chapter 10.1).

3.8 Gene Expression Analysis

Gene expression analysis based on real-time PCR was performed to verify the levels of lncRNA expression in MSC-induced cancer cells or to quantitate the levels of differently expressed genes when treated under different conditions.

3.8.1 RNA Isolation

Adherent cells were washed twice with 1x PBS and lysed using 700 µl QIAzol reagent. RNA was extracted using miRNeasy Mini Kit (Qiagen) following the manufacturer's protocol. The concentration and quality of each purified RNA sample was determined using the NanoDrop1000 Spectrophotometer. The quality of each sample was evaluated by calculating the ratio of absorbance at 260 nanometer (nm) and 280 nm and a ratio of around 2.0 was considered as pure RNA.

3.8.2 Reverse Transcription

Before the reverse transcription (RT) reaction was performed, all RNA samples were measured, analyzed for their quality and diluted to the same concentration.
100 ng of RNA were used to produce cDNA, with the QuantiTect Reverse Transcription Kit according to the manufacturer's protocol. This kit includes one step of genomic DNA elimination, which is necessary to further reduce the risk of false positive results due to DNA contamination when gene expression analysis is performed later on. Again, the concentration and quality of each produced cDNA sample was determined using the NanoDrop1000 Spectrophotometer. An absorbance ratio at 260/280 nm between 1.8-2.0 was desired and considered as pure DNA.

3.8.3 Primers Design for qRT-PCR

Primers were designed using the Primer3Plus web interface (UNTERGASSER et al., 2007). For the purpose of real-time PCR analysis, the amplicon was designed to be between 200 and 300 bp. If possible, forward and reverse primers were placed to span over more than one exon to avoid the risk of false positive results due to DNA contamination. Primer specificity was evaluated using the BLAT sequence alignment tool (KENT, 2002).

Only primers that were shown to align specifically to the sequence of interest were used and further tested by *In-Silico* PCR (http://genome.ucsc.edu/cgi-bin/hgPcr). With this algorithm, it could be tested if the designed primers would lead to the expected amplicon. Ultimately, primer oligonucleotide sequences were ordered from Eurofins MWG Operon (Huntsville, AL). A complete list of all primers used can be found in the appendix (Chapter 10.2).

3.8.4 Quantitative Real-Time PCR

In general, every PCR involved at least three biological replicates of each condition ($n \geq 3$). Real-time PCR was performed using the QuantiTect SYBR Green PCR Kit. Every master mix that was prepared included a non-template control (NTC) to detect potential DNA contamination. All samples, including the NTC were measured in triplicates (technical replicates). For normalization, expression levels of the housekeeping gene 18S ribosomal RNA were analyzed from all samples.

qRT-PCR Master Mix Preparation for sample triplicate

Primer mix (Forward+Reverse) 20 µM	3.4 µl
Nuclease-free H_20	12.6 µl
2X SYBR Green PCR MM	17 µl
Template cDNA (340 ng/µl)	1 µl
Total volume:	**34 µl**

For each triplicate reaction 10 µl of master mix were used and pipetted into one well of a 384-well plate.

qRT-PCR Final concentration / reaction

Primer mix (Forward+Reverse)	2 µM
2X SYBR Green PCR MM	1X
Template cDNA	100 ng

The 384-well plates were sealed with a polypropylene adhesive film and centrifuged 3 min at 1000 rpm at 4°C. The real time PCR reaction was set up using Sequence Detection Software 2.4 and performed with the 7900HT Fast Real-Time PCR System using the following conditions:

Thermal Profile

Stage 1	**Stage 2 (40 cycles)**			**Stage 3 - Dissociation Stage**		
95°C	95°C	60°C	70°C	95°C	60°C	95°C
15 min	15 sec	30 sec	30 sec	15 sec	15 sec	15 sec

3.8.5 qRT-PCR Data Analysis

Melting Curve Analysis

First, the specific amplification of each PCR reaction was evaluated by melting curve analysis. During the dissociation stage, the generated double-stranded DNA product will denature at a certain temperature, resulting in the emission of the fluorescent SYBR green dye. A specific amplification will result in the generation of one single peak within the melting curve. However, unspecific products will differ in melting temperature and therefore result in multiple peaks within the melting curve. Only specific PCR reactions were further analyzed to determine relative gene expression.

Determination of Relative Gene Expression Levels

Gene expression can be quantified by the number of PCR cycles needed to produce a fluorescence signal that exceeds a certain threshold. However, the number of cycles that are needed depends on the initial amount of cDNA within the sample. Importantly, samples that express the gene of interest (GOI) at high levels will need fewer cycles to exceed the threshold than samples with a lower expression levels. The Cycle threshold (Ct) value of a sample determines the PCR cycle number at which the threshold is exceeded.

The threshold line was evaluated for each experiment individually and set to a level that distinguished unspecific background from specific amplification in the exponential phase of the PCR reaction.

In summary:

1.) Ct values of sample triplicates were averaged for each gene.
2.) The averaged Ct value of the housekeeping gene was subtracted from the averaged Ct value of the GOI to calculate Delta Ct values for each sample.
3.) The Delta Ct values between biological replicates within control and test samples were averaged.
4.) Averaged Delta Ct value of control samples was subtracted from averaged Delta Ct values of experimental samples (Delta Delta Ct).
5.) The relative fold change between experimental and control samples was calculated $2^{(-\text{Delta Delta Ct})}$.

Statistical Analysis of Relative Gene Expression

The variability of gene expression levels within the experimental and control samples was determined by evaluating the standard deviation (SD) and the standard error of the mean (SEM).

To evaluate if the levels of GOI expression between experimental and control samples are significantly different, a two-tailed Student's t-test was performed. The significance level was set at 0.05.

3.8.6 Real-Time PCR Product Verification

To further ensure primer specificity, real-time PCR products that were generated using newly designed primers were combined with 6x loading dye and loaded onto a 2 % agarose gel containing 0.6 µg/ml ethidium bromide. The 1kB DNA ladder was used as a reference to identify the size of the PCR products and loaded on the gel as well. The gel electrophoresis was performed using 1x TAE buffer and run for ca. 40 min at 100 Volt. PCR products were visualized using a gel Illumination system and analyzed using Quantity One 1-D Analysis Software (both Bio-Rad). By comparing the expected size of the PCR product to the actual size on the gel, amplification

specificity could be determined. Further, DNA was extracted from the gel and purified PCR products were sent to Genewiz Inc. (Cambridge, MA) and sequenced for verification.

3.9 Clinical Samples of Patients

The RNA of clinical samples from breast cancer and healthy patients was kindly provided by the Curie Institute (Paris, FRANCE). There, the samples were collected by macro dissection of breast cancer or healthy tissue and further processed according to Institutional Review Board (IRB) approved protocols and ethical policy. The provided, anonymized RNA samples were reverse transcribed into cDNA and tested by qRT-PCR to determine lncRNA expression, as described previously.

3.10 Analysis of Gene Promoter Region

The non-coding transcript lnc-KCNJ9-2:2, which is derived from the RefSeq gene LOC100505633, is transcribed from the chromosomal location between chr1:159,931,014-159,948,876. The 2 kb immediate up-stream region of lnc-KCNJ9-2:2 was screened for potential FOXP2 binding motifs. Briefly, the up-stream sequence was retrieved using the USCS Genome browser and compared to known and potential FOXP2 binding sites.

3.11 Comparing Large Scale Genomic Data Sets

The cBioPortal for cancer genomics was used to compare large-scale genomic data sets (CERAMI et al., 2012; GAO et al., 2013a). This portal stores cancer genomic data sets and allows exploratory data analysis. Alterations of the RefSeq gene LOC100505633 (lnc-KCNJ9-2:2) in various malignant cancer diseases were looked up in the database.

3.12 Molecular Cloning of LncRNAs

3.12.1 Cloning Strategy

The PLJM1-GFP plasmid was kindly provided by Dr. Sabatini (Whitehead Institute, Cambridge, USA) and served as plasmid backbone for the development of lncRNA overexpression constructs. It is a lentiviral vector that allows mammalian expression and is based on pLKO.1 (SANCAK et al., 2008). However, due to its CMV promoter any cDNA is expressed very efficiently. Due to this backbone, the green fluorescent protein (GFP) can be expressed at very high levels. Moreover, it confers puromycin resistance, which can be used to select infected cells. By using AgeI and EcoRI restriction enzyme digestion, the inserted GFP can be removed and replaced by other cDNA (e.g. HOTAIR or lnc-KCNJ9-2:2) that has been amplified with the respective restriction sites on the 5' and 3' end and digested in the same way.

Development of the PLJM1-HOTAIR Construct

The LZRS-HOTAIR vector is a retroviral construct that contains the complete HOTAIR sequence and was purchased from Addgene (GUPTA et al., 2010). This plasmid, however, as it is retroviral, will not lead to the same infection efficiency as any lentiviral vector. Moreover, LZRS-HOTAIR infected cells are not puromycin selectable. Therefore, this vector was not used directly for the development of overexpression cell lines but served as a template for the amplification of the complete HOTAIR sequence.

First, cloning primers were designed that allowed specific amplification of the complete HOTAIR sequence and included the artificial AgeI and EcoRI restriction sites on the 5' and the 3' end, respectively.

The full-length HOTAIR sequence was amplified by regular PCR using Phusion High-Fidelity PCR Master Mix with GC Buffer. In short, the cloning primers were mixed with the Phusion Master Mix, LZRS-HOTAIR plasmid DNA and water. The PCR reaction was performed according to the manufacturer's protocol. PCR products were separated by agarose gel electrophoresis and the specific HOTAIR product was extracted with the QIAquick Gel extraction Kit. Then, the PLJM1-GFP plasmid and

the HOTAIR PCR product were digested with AgeI and EcoRI. Digested backbone and HOTAIR insert were run on a 1 % agarose gel and extracted. Finally, the PLJM1 backbone and digested HOTAIR were ligated resulting in the development of the PLJM1-HOTAIR overexpression construct. The resulting plasmid was confirmed by AgeI and EcoRI double digest and subsequent evaluation of the expected pattern. Additionally, the new plasmid was verified by sequencing, to proof that the lentiviral construct contains the full-length HOTAIR sequence.

Development of the PLJM1-lnc-KCNJ9-2:2 Construct

First, the full-length lncRNA KCNJ9-2:2 was amplified from MSC-induced MDA-MB-231 cDNA. However, sequence information concerning this novel lncRNA exists only due to high-throughput RNA-seq data (LOC100505633) that can be found on the LNCipedia website or the UCSC Genome browser.

To determine the complete sequence from 5’ to the 3’ end a PCR-primer based assay had to be performed. This was necessary because data based on RNA-sequencing is not reliable and often leads to transcripts that are longer, because of a 5’ and 3’ end shift. Starting from the RefSeq based 5’ end of lnc-KCNJ9-2:2, multiple forward primers were designed in approximately 20 bp intervals. Additionally, 2 different reverse primers were designed to identify the real 3’ end. Forward and reverse primers were cross-combined in regular PCR reactions to determine the real starting and end point of lnc-KCNJ9-2:2. PCR products were run on a 2 % agarose gel to distinguish expected from unspecific products and extracted from the gel, subsequently. Then, DNA of the expected products was purified and sequenced.

The full-length lncRNA was amplified with cloning primers including artificial AgeI and EcoRI restriction sites. Similarly to cloning of HOTAIR, the PCR product was digested and run on a 2% agarose gel. Upon extraction, the lnc-KCNJ9-2:2 insert was ligated with the PLJM1 backbone, resulting in the development of the PLJM1-lnc-KCNJ9-2:2 overexpression construct. This new vector was evaluated by AgeI and EcoRI double digest to confirm the proper ligation. Additionally, the plasmid was sequenced using the CMV-Forward primer, to verify that it contains the complete lncRNA transcript (1040 bp).

3.12.2 Cloning Primer Development

Cloning primers were designed based on the following concept (5' – 3'):

- 6 nt that are not specific, but improve restriction enzyme digestion
- 6 nt that generate the artificial restriction site (AgeI or EcoRI)
- 21 nt that allow specific recognition of the sequence.

HOTAIR-cloning Forward primer (AgeI restriction site):
5' ATTAGTACCGGTGACTCGCCTGTGCTCTGGAGC 3'

HOTAIR-cloning Reverse primer (EcoRI restriction site)
5' TGCTTAGAATTCTTTTTTTTTTGAAAATGCATC 3'

lnc-KCNJ9-2:2-cloning Forward primer (AgeI restriction site)
5' CGATTCACCGGTATAATAAAAGGCCAAACCTTTGC 3'

lnc-KCNJ9-2:2-cloning Reverse (EcoRI restriction site)
5' TGATTGGAATTCTTGACGAGACACATTTAATAA 3'

3.12.3 PCR-based Cloning

The PCR primer assay to determine 5' and 3' ends, as well as amplification of the full-length HOTAIR and lnc-KCNJ9-2:2 sequences were done using the Phusion High-Fidelity PCR Master Mix with GC Buffer following manufacturer's protocol.

Reaction setup

Component	**Per reaction**	**Final Concentration**
10 µM Forward	1.25 µl	0.5 µM
10 µM Reverse	1.25 µl	0.5 µM
2X Phusion Master Mix	12.5 µl	1X
Template DNA	1.0 µl	150 ng (cDNA) 50 ng (Plasmid DNA)
Nuclease-Free Water	9.0 µl	
Total Volume	**25 µl**	

Thermocycling Conditions

Step	Temperature	Time
Initial Denaturation	98°C	30sec
35x	98°C	10 sec
	60°C	30 sec
	72°C	35 sec
Final extension	72°C	7min
Hold	4°C	

PCR products were mixed with 6x loading dye and run onto a 2% containing 0.6 µg/ml ethidium bromide as described earlier.

3.12.4 DNA Gel Extraction

Agarose gels were evaluated under a hand-held UV-lamp to visualize DNA. Desired bands within the agarose gel were identified based on their size and cut out. DNA was extracted from the gel using the QIAquick Gel extraction Kit, following manufacturer's protocol. Concentration and quality were determined using NanoDrop1000 Spectrophotometer as described earlier.

3.12.5 Restriction Enzyme Digestion and Plasmid Verification

In general, 1 µg DNA was mixed with 5 µl 10x Buffer and 1 µl of each restriction enzyme. Then, nuclease-free water was added to a final volume of 50 µl. To perform the restriction enzyme digest, the mix was incubated at 37°C for 60 min in a heating block. For verification, plasmids were digested with restriction enzymes that resulted in the generation of a unique pattern that could only be observed in successfully cloned constructs. To analyze the pattern, samples were run on a 1 % agarose gel. Digested backbone or insert DNA was isolated from the gel as described before.

3.12.6 Ligation and Transformation

Ligation of PLJM1 empty vector (backbone) and either HOTAIR or lnc-KCNJ9-2:2 insert was performed using the T4 DNA Ligase. Briefly, ligation reactions were set up in a 1.5 ml tube using a molar ratio of 1:3 of vector to insert DNA. Additionally, a control reaction was set up for every ligation performed that contained no insert DNA. This was done to see to which extend self-ligation of the vector itself was possible. Then, 2 µl ligase and nuclease-free water were added to a final volume of 20 µl. The mixture was incubated 10 min at room temperature and chilled on ice, subsequently.

For transformation, vials containing One-Shot Stbl3 chemically component E. coli cells were thawed on ice. Then, 5 µl DNA of each ligation reaction was gently added to the cells and the vials were incubated for 30 min on ice. Following incubation, heat-shock was performed in a 42°C water bath for 45 sec. Afterwards, the tubes were placed on ice for another 2 min, before 250 µl of pre-warmed Super Optimal broth with Catabolite repression (S.O.C) medium was added. The tubes were incubated at 37°C for 1 h at 250 rpm in an orbital shaker, before 50 µl from each transformation was distributed onto a LB-agar plate containing 100 µg/ml ampicillin. Finally, the plates were incubated overnight at 37°C.

3.12.7 Plasmid Isolation

Colony formation on the selective LB-agar plates was evaluated after overnight incubation. No growth on the control plate (no insert DNA) was observed. This indicated that the digested vector did not self-ligate and colonies observed in the other plates are due to an efficient vector and insert ligation. At least 5 single colonies of each plate were inoculated in 3 ml LB-media containing 100 µg/ml ampicillin and incubated at 37°C and 250 rpm, overnight. The next day, plasmid DNA was isolated using the QIAprep Spin Miniprep Kit following manufacturer's protocol. To verify the constructs a restriction enzyme digest was performed, as described earlier. Once successfully cloned constructs had been identified, their bacterial liquid culture was used to inoculate another 150 ml LB/ampicillin media. Following

overnight incubation at 37°C and 250 rpm, plasmid DNA was isolated using the HP Plasmid Maxiprep System, following manufacturer's protocol. This was done to generate a plasmid DNA stock solution.

3.10.8 Verification by Sequencing

Backbone, inserts (HOTAIR and lnc-KCNJ9-2:2) and resulting expression plasmids were verified by sequencing. Briefly, 10 µl plasmid DNA with a concentration of 80 ng/µl or 10 µl of purified PCR products with a concentration of 4 ng/µl (HOTAIR or lnc-KCNJ9-2:2 inserts) were mixed with 5 µl of 5 µM sequencing primer. The samples were sent to Genewiz Inc. (Cambridge, MA) and sequenced. Sequence analysis was performed using the ApE Plasmid Editor and the SDSC Biology WorkBench. By using Clustal W multiple sequence alignment the plasmid constructs were verified.

4 Results

Human bone marrow derived mesenchymal stem cells (MSCs) were shown to travel to the site of primary tumor lesion and to promote breast cancer metastasis (KARNOUB et al., 2007; MANDEL et al., 2013). The physical interactions that take place between these stromal progenitors and the (primary tumor) cancer cells induce transcriptional alterations in the MSC-activated cancer cells. These lead to changes in the expression levels of a number of genes regulating cancer progression (EL-HAIBI et al., 2012). Indeed, cancer cells that have been educated by MSCs were shown to be highly metastatic, express higher levels of mesenchymal markers and display an expanded cancer stem cell population (EL-HAIBI et al., 2012). However, the detailed molecular mechanisms that lead to these phenotypes still remain elusive. Evidence suggests that long non-coding RNAs play a critical role in cancer progression (GUPTA et al., 2010; LU et al., 2013; ZHAO et al., 2014). The aim of this study was to determine their contributions to MSC-triggered metastasis of breast cancer.

Background

To investigate the transcriptional fluctuations exhibited by highly metastatic breast cancer cells, comparative microarray expression analyses were previously conducted on total RNA derived from MSC-activated MDA-MB-231 cells. MDA-MB-231 cells cultured alone were used as controls (EL-HAIBI et al., 2012). The resulting data, analyzed by Dr. Karnoub (BIDMC), Dr. Beck (BIDMC), and Dr. Rinn (Broad Institute), revealed significant >1.5-fold enrichment in the expression of 9 distinct lncRNAs in MSC-primed cancer cells, shown in Table 2.

Transcript ID	Fold Change	q-value(%)	LNCipedia ID
TCONS_00005559_1	4.000	0.000	lnc-ST3GAL6-2:1
TCONS_00004205_1	2.166	0.000	lnc-AC007401.2.1-1
TCONS_00026813_1	1.992	7.162	lnc-UQCRFS1-9
TCONS_00002647_1	1.903	2.317	lnc-AC016722.1.1-1
TCONS_00014854_1	1.896	4.355	lnc-MYC-2
TCONS_00013598_1	1.763	3.014	lnc-MKLN1-1
TCONS_00017736_1	1.631	2.317	lnc-MBL2-4
TCONS_00000659_2	1.625	8.664	lnc-KCNJ9-2
TCONS_00019082_1	1.543	0.000	lnc-SBF2-2

Table 2 MSC-induced LncRNAs detected by Microarray Screening
List of lncRNAs that were induced in breast cancer cells upon co-cultivation with MSC. Ranking was based on fold change of lncRNA expression between MSC-induced MDA-MB-231 cells versus MDA-MB-231 cells alone. The q-value threshold (false-discovery rate (FDR) adjusted p-value) was set < 10 %, only significant hits were considered in the analysis (Karnoub AE, unpublished data).

4.1 Verification of the Microarray Analysis

To validate the identity of the hits, and to ascertain the elimination of any potential false-positives detected, the 9 MSC-induced lncRNAs were vetted using two approaches: (1) the microarray probe sets utilized to denote the respective lncRNAs were analyzed for specificity using bioinformatic algorithms (BLAT); (2) The remaining lncRNAs were verified by qRT-PCR to determine their actual fold change in MSC-activated MDA-MB-231 cells as compared to ‘resting’ MDA-MB-231 cells cultured alone.

4.1.1 Microarray Probe Set Verification

To confirm that the lncRNAs, which were significantly up regulated by the microarray analysis, (shown in Table 2) are not false positive results, the probe set of each one was verified by BLAT sequence alignment. Upon analysis (shown in appendix Chapter 10.1), it was found that the probe sets of lnc-UQCRFS1-9 and lnc-MYC-2 are not specific and therefore, these two lncRNAs were not further tested. All the probe sets for the remaining lncRNAs exhibited >98% specificity to the designated transcripts, and the respective lncRNAs were retained for further analyses below.

4.1.2 Identification of Two Novel MSC-induced LncRNA Transcripts

To test the remaining lncRNA candidates, a co-culture of human bone marrow derived MSCs with GFP-labeled MDA-MB-231 breast cancer cells was performed. The MSC-activated MDA-MB-231 cells (231*hBM-MSC) were recovered by cell sorting.

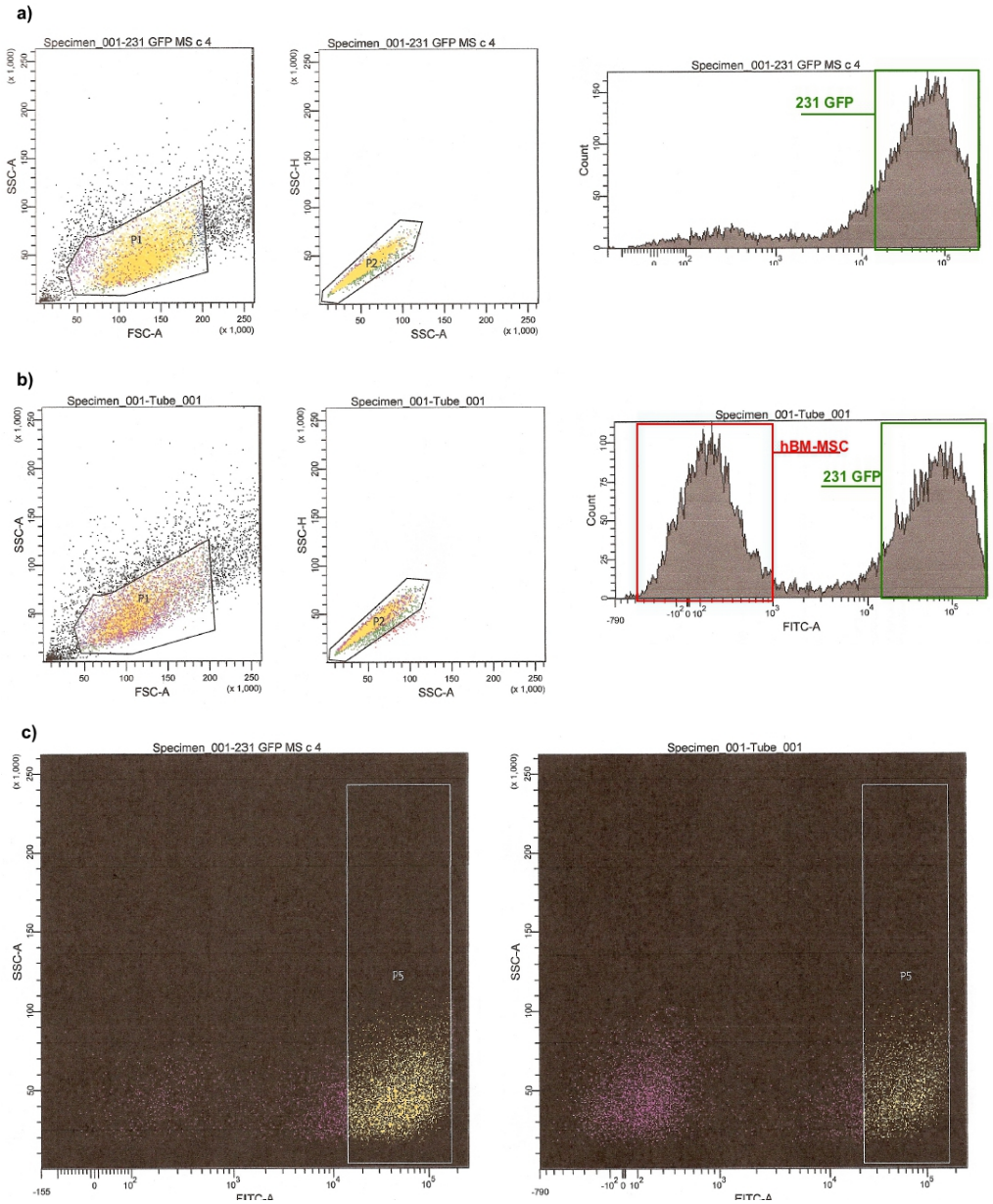

Figure 7 Gating and Cell Sorting of Co-culture Experiments
*a) Sorting of GFP-labeled MDA-MB-231 cells only***.** Used to determine gates. Cells were distinguished based on size to separate living cells from debris and to avoid cell aggregates. From the remaining cells only the strongest GFP-expressing ones were selected. *b) Sorting of MDA-MB-231 + MSC co-cultures***.** The same gates used before were applied on co-cultures to isolate MSC-activated MDA-MB-231 cells only. Both cell populations can be clearly discriminated. *c) Collection of sorted cells.* MSC-activated MDA-MB-231 cells or control cells were selected and collected in 1.5 ml MEM-α media.

As shown in Figure 7, gates were set to select the main population of cells and to exclude small cell debris. Ultimately, only the MDA-MB-231 cancer cells with the strongest GFP expression were selected in order to avoid MSC contamination. The gates that were determined by using GFP-expressing MDA-MB-231 cells alone were used for sorting of the co-culture samples.

Then, lncRNA expression levels in MSC-activated MDA-MB-231 cells (231*hBM-MSC) were tested compared to resting MDA-MB-231 cells cultured alone (231 Non-Contact) using qRT-PCR.

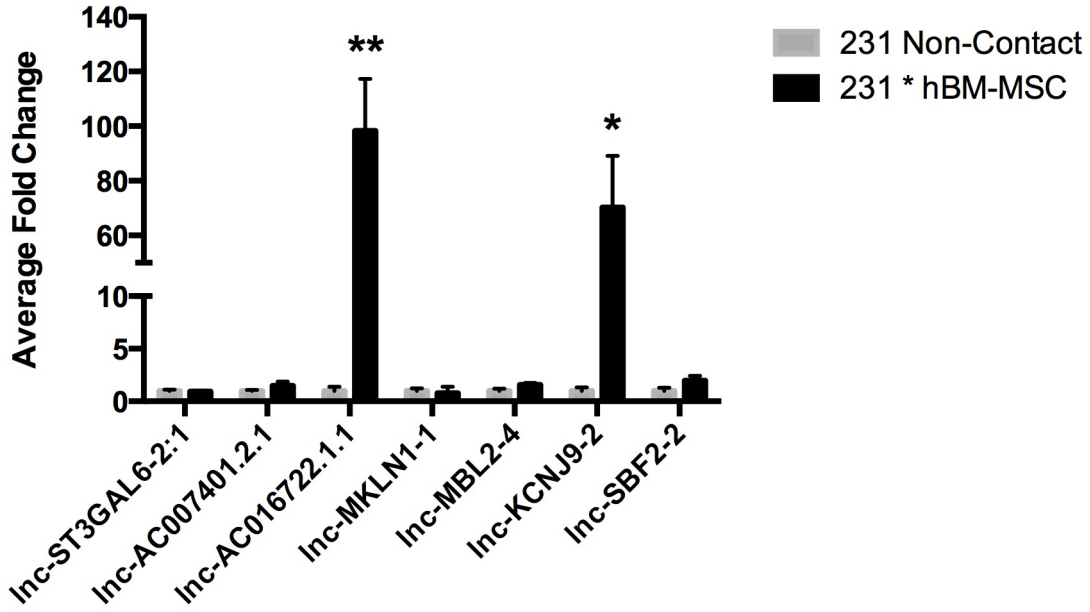

Figure 8 Verification of LncRNA Expression Levels by qRT-PCR
Expression levels of the lncRNAs discovered by microarray screening were tested by qRT-PCR in MDA-MB-231 cells cultured with or without hBM-MSC for 72 h. Data was normalized to the levels of 18S housekeeping gene expression and is expressed as fold induction ± SEM (n=3). Statistical significance was evaluated by two-sided student's t-test (* p-value < 0.05, ** p-value < 0.01). Two lncRNAs (lnc-AC016722.1.1 and lnc-KCNJ9-2) were found to be elevated.

As shown in Figure 8, the expression levels of the novel, non-coding transcripts lnc-AC016722.1.1 and lnc-KCN9-2 were significantly up regulated in MSC-activated MDA-MB-231 cells. In both cases a very strong induction (lnc-AC016722.1.1 > 100 fold and lnc-KCNJ9-2 > 70 fold) could be determined. However, while these two lncRNAs showed even a 40-50 times higher induction than suggested by microarray

screening, no significant difference in the expression levels of the other lncRNAs could be seen.

The detailed results of the verification of lncRNA expression in MSC-activated MDA-MB-231 cells by qRT-PCR are shown in Table 3. The microarray analysis suggested that lnc-ST3GAL6-2:1 would be 4 fold higher expressed and that lnc-MKLN1-1 would be approximately 1.7 fold higher expressed in MSC-activated MDA-MB-231 cells. However, this could not be confirmed by qRT-PCR analysis.
Although the analysis by qRT-PCR of lnc-AC007401.2.1-1, lnc-MBL2-4:2 and lnc-SBF2-2 indicated a similar expression as suggested by microarray screening, the results were not statistically significant. For this reason and due to their significant induction in the MDA-MB-231 cancer cells by MSC, the project focused on the characterization of the novel lncRNAs ln-AC016722.1.1-1 and lnc-KCNJ9-2.

Transcript ID	Fold Change	p-value	LNCipedia-ID	Verification
TCONS_00005559_1	0.924	0.65	lnc-ST3GAL6-2:1	X
TCONS_00004205_1	1.491	0.275	lnc-AC007401.2.1-1	X
TCONS_00002647_1	**98.316**	**0.007**	**lnc-AC016722.1.1-1**	✓
TCONS_00013598_1	0.787	0.759	lnc-MKLN1-1	X
TCONS_00017736_1	1.556	0.125	lnc-MBL2-4	X
TCONS_00000659_2	**70.274**	**0.021**	**lnc-KCNJ9-2**	✓
TCONS_00019082_1	1.977	0.135	lnc-SBF2-2	X

Table 3 MSC-Induced lncRNAs Verified by qRT-PCR Analysis
Summary of the expression analysis of novel lncRNAs in MSC-activated MDA-MB-231 cancer cells compared to MDA-MA-231 cells cultured alone. The overexpression of lnc-ST3GAL-6-2:1 and lnc-MKLN1-1 could not be confirmed by qRT-PCR. The differences in expression of lnc-AC007401.2.1-1, lnc-MBL2-4 and lnc-SBF2-2 between MSC-activated and control MDA-MB-231 cells were not significant. However, qRT-PCR revealed that lnc-AC016722.1.1-1 and lnc-KCNJ9-2 were 40-50 times more induced than predicted by the microarray analysis.

4.1.3 Determination of Specific Isoforms of MSC-induced LncRNA Transcripts

LncRNAs are often expressed from loci that generate multiple closely related isoforms that share substantial sequence homologies. Due to the fact that lncRNA isoforms are very similar, the microarray probe sets that designated lnc-AC16722.1.1-1 and lnc-KCNJ9-2 cannot indicate the expression of a specific isoform. More importantly, the oligonucleotide sequences that were used on the microarray for the detection of the expression levels of these lncRNAs align at the same exon sequence that is shared by all isoforms. Therefore, discrimination by microarray analysis was not possible.

To determine the specific isoforms of lncRNAs AC16722.1.1-1 and lnc-KCNJ9-2, the UCSC genome browser was used to retrieve sequence information regarding the different isoforms and to illustrate how and where the respective Affymetrix probe sets align.

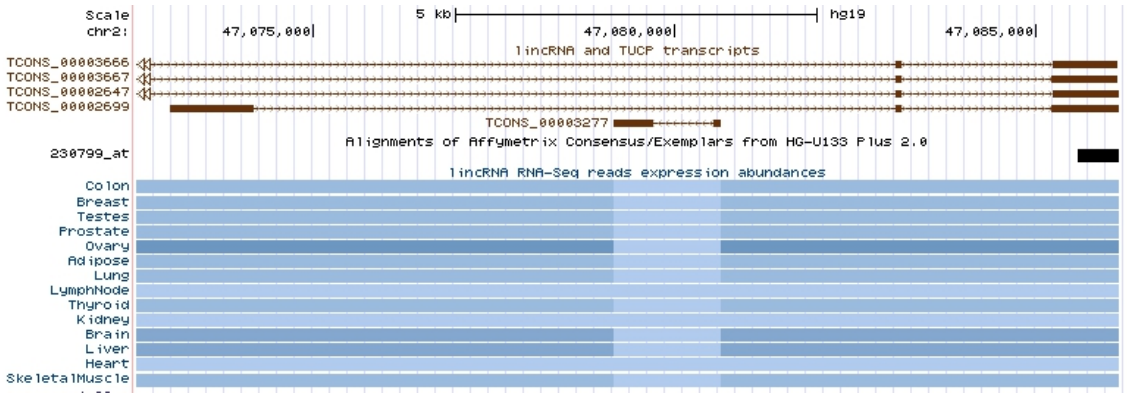

Figure 9 Lnc-AC016722.1.1-1 Sequences and Probe Set Alignment
Based on the chromosomal location of the lncRNA derived by microarray screening, information on the different isoforms of lnc-AC016722.1.1-1 was retrieved using UCSC genome browser. The Affymetrix probe set (230799_at) that led to the discovery of this lncRNA can be aligned to all 4 isoforms, indicating that no discrimination among the different isoforms based on microarray analysis was possible. LncRNA expression abundances based on RNA-seq data in different human tissues are displayed from light blue (very low abundance) to dark blue (very high abundance) (TRAPNELL et al., 2010; CABILI et al., 2011).

As a result, 4 isoforms (TCONS_00003666, TCONS_00003667, TCONS_00002647 and TCONS_00002699) of lnc-AC016722.1.1-1 were found based on RNA-seq data using the UCSC genome browser (CABILI et al., 2011; TRAPNELL et al., 2010), as shown in Figure 9. Additionally, this data provided an insight into the expression of

this lncRNA isoforms in different human tissues. Expression abundance scores range from 0 to 1000 fragments per kilobase of exon per million fragments mapped (FPKM), which approximates relative lncRNA transcript abundance, and are displayed from light blue to dark blue, respectively (TRAPNELL et al., 2010; CABILI et al., 2011). The data suggested a high enrichment within ovary, brain and liver tissues.

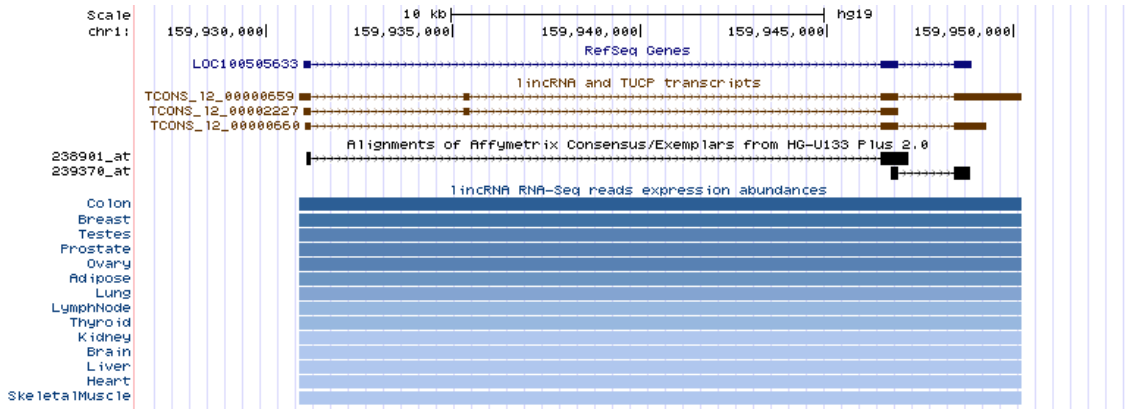

Figure 10 Expression of Lnc-KCNJ9-2
Data retrieved from the UCSC genome browser on the chromosomal location of lnc-KCNJ9-2 indicated 3 different expressed isoforms. Additionally, the RefSeq gene (LOC100505633) is identical to isoform 3 (TCONS_I2_00000660). LncRNA expression abundances based on RNA-seq data in different human tissues are displayed from light blue (very low) to dark blue (very high) (TRAPNELL et al., 2010; CABILI et al., 2011).

The lncRNA KCNJ9-2 exists in 3 different isoforms (TCONS_I2_00000659, TCONS_I2_0000227 and TCONS_I2_0000660) (Figure 10). The RefSeq gene LOC100505633 is identical to TCONS_I2_0000660 and refers to isoform 3. Moreover, RNA-seq data indicated that this non-coding transcript seems to be expressed ubiquitously, although the highest abundances are found in colon, breast, testes and prostate tissue (TRAPNELL et al., 2010; CABILI et al., 2011). This data suggests that its expression is not restricted to a specific location.

Further, isoform-specific primer sets were designed to discriminate and display abundance levels of lncRNA isoforms. Then, expression levels of the specific isoforms in MSC-induced MDA-MB-231 cells were compared to control MDA-MB-231 cells by qRT-PCR analysis.

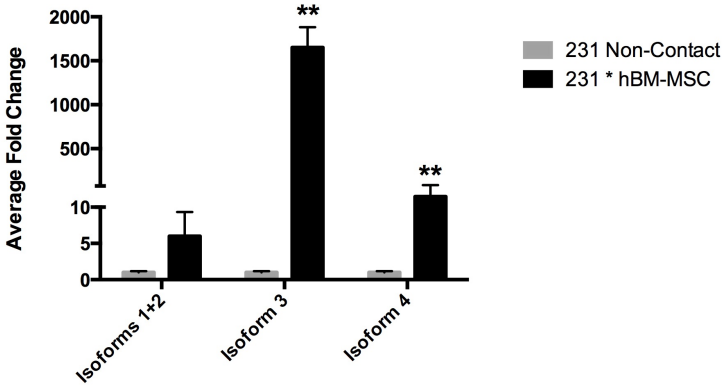

Figure 11 Isoforms of Lnc-AC016722.1.1 in MSC-activated MDA-MB-231 Cells
Three primer pairs were designed to discriminate between isoforms 1+2, isoform 3 and isoform 4 of lnc-AC016722.1.1. qRT-PCR analysis was performed to show expression levels in MDA-MB-231 cells cultured with or without hBM-MSC. Data was normalized to 18S and expressed as fold induction ± SEM (n=3). ** p-value < 0.01 (two-sided student's t-test)

It was not possible to design primers that could discriminate between isoforms 1 and 2 (TCONS_00003666 and TCONS_00003667) of lnc-AC016722.1.1 because they differ only by a few base pairs. However, specific primers could be designed for isoform 3 (TCONS_00002647) and isoform 4 (TCONS_00002699). Figure 11 shows the results of the determination of the expression levels of the different isoforms of lnc-AC016722.1.1. It could be demonstrated that isoform 3 is the most abundant one. However, the difference in expression of isoforms 1+2 between control and MSC-activated MDA-MB-231 cells was not statistically significant. Moreover, upon testing of expression levels in MDA-MB-231 cells, isoform 4 was also significantly up regulated by MSC induction. Since isoform 3 seemed to be most induced following MSC stimulation (> 1500 fold change), all subsequent experiments concentrated on this particular isoform.

Next, the expression abundances of lnc-KCNJ9 isoforms were tested in MSC-activated MDA-MB-231 cells and compared to MDA-MB-231 control.

Because the isoforms of lnc-KCNJ9 are very similar among each other, it was not possible to design isoform-specific primers for all. Only expression levels of isoform 1 (TCONS_I2_00000659) could be determined specifically. Therefore, primer sets were developed that allowed evaluation of the expression profile of these isoform variants through different primer combinations. Isoform 1 of lnc-KCNJ9 was not significantly increased in 231*hBM-MSC cells as compared to control cells. However, when the primer combination that detects isoform 1 + isoform 3 (TCONS_I2_0000659 + TCONS_I2_0000660) was tested, a significant difference (>1000 fold) could be determined (Figure 12). Also, when the primer combination was used that detects Isoform 1+2+3 together, a significant difference (>600 fold) could be determined in MSC-activated MDA-MB-231 cells compared to control MDA-MB-231 cells.

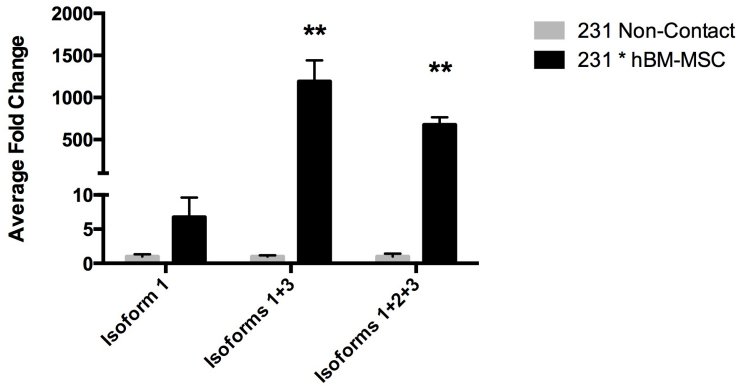

Figure 12 Isoforms of lnc-KCNJ9-2 in MSC-activated MDA-MB-231 Cells
Three primers were designed to discriminate between isoforms 1, isoform 1+3 and isoform 1+2+3 of lnc-KCNJ9-2. qRT-PCR was performed to show expression levels in MDA-MB-231 cultured with or without hBM-MSC. Data was normalized to 18S and expressed as fold induction ± SEM (n=6). ** p-value < 0.01 (two-sided student's t-test)

Expression levels of Isoform 1+2+3 displayed a lower fold change of induction in MSC-activated MDA-MB-231 cells versus control when directly compared to expression levels of Isoform 1+3. This result suggested that isoform 2 (TCONS_I2_0000227) of lnc-KCNJ9 might not be highly abundant in MSC-activated cell, thereby diminishing the overall fold change.

To further prove this hypothesis, the PCR products of the qRT-PCR that used the primer set that detects Isoform 1+2+3, of MSC-activated MDA-MB-231 or control were analysis by gel electrophoresis. It is important to mention that the primer pair

that detects isoforms 1+2+3 will result in the generation of two bands as indicated by figure 13.

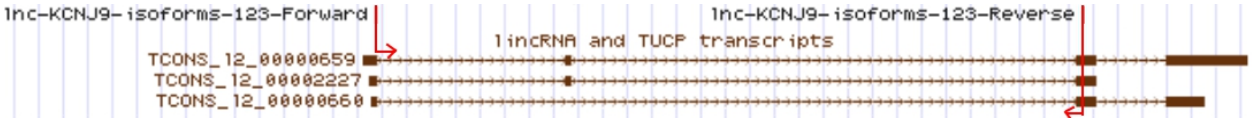

Figure 13 Primer Set for the Detection of Isoforms 1+2+3 of Lnc-KCNJ9-2
By using a primer set that detects all 3 isoforms at the same time, 2 bands can be expected. Isoform 1+2 (TCONS_l2_0000659 + TCONS_l2_0000227) will result in the generation of a PCR product that is 150 bp larger than the PCR product of isoform 3 (TCONS_l2_0000660). This is due to an additional exon that does not exist in isoform 3.

The primer set for detection of Isoform 1+2+3 resulted into 2 bands. Isoform 1+2 was shown by the band at 375 bp and Isoform 3 displayed by the band at 228 bp.

Isoform determination in MSC-activated MDA-MB-231 cells (here displayed as 231*MSC) resulted in the detection of a very strong band for isoform 3, however, the band indicating isoforms 1+2 was very weak (Figure 14). Upon gel extraction and subsequent sequencing, the identity of the bands corresponding to the respective isoforms could be confirmed. Moreover, isoform determination of lnc-KCNJ9-2 in MDA-MB-231 control cells (here shown as 231 NC) displayed no amplification of Isoform 3 and very weak amplification of Isoform 1+2 and unspecific background amplification.

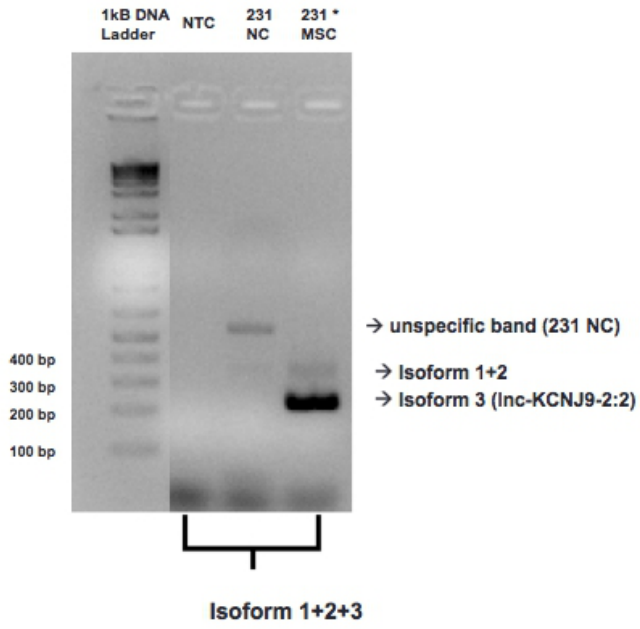

Figure 14 Isoform Determination of Lnc-KCNJ9-2
PCR products of the qRT-PCR using the primer set 1+2+3 to determine isoform abundance levels of lnc-KCNJ9-2-2 of control MDA-MB-231 cells (231 NC), MSC-activated MDA-MB-231 cells (231*MSC) or no template control (NTC) were run on a 2 % agarose gel. A strong band of Isoform 3 (228 bp) and a weak band of Isoform 1+2 (375 bp) were determined in 231*MSC. In 231 NC unspecific amplification and only weak amplification of Isoform 1+2 and no amplification of Isoform 3 can be seen.

In summary, these results suggest that isoform 3 of lnc-KCNJ9-2 is the transcriptional variant that is affected the most by MSC-activation. Isoform 3 (TCONS_I2_00000660 / RefSeq gene LOC100505633) will be further referred to as lnc-KCNJ9-2:2, because this is the transcript ID given by the LNCipedia database (VOLDERS et al., 2013).

It was not possible to design primers that allow a specific detection of lnc-KCNJ9-2:2 because its complete sequence is shared with isoform 1. However, specific primers could be designed for isoform 1, because it is longer, but indicated no significant difference between control and MSC-activated MDA-MB-231 cells.

Therefore, all further experiments that concerned the expression levels of lnc-KCNJ9-2:2 were performed using the primer set that detected lnc-KCNJ9-2:2 and isoform 1 at the same time.

The putatively most abundant isoforms (lnc-AC016722.1.1:3 and lnc-KCNJ9-2:2) were determined and their expression levels were tested by qRT-PCR analysis (Figure 15). Both lncRNAs were vastly (> 1000 fold) induced by MSC in MDA-MB-231 cancer cells.

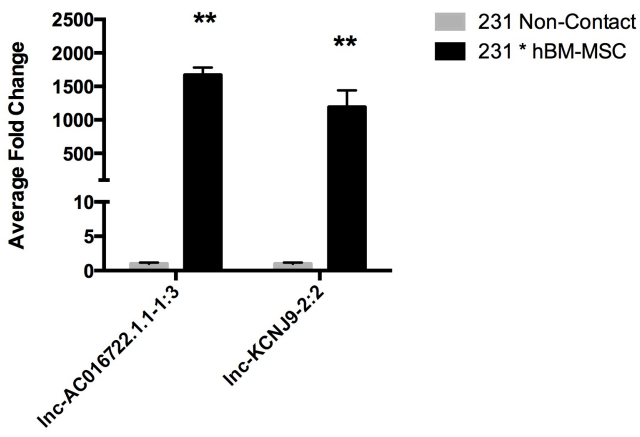

Figure 15 Summary - Isoforms of MSC-induced lncRNAs in MDA-MB-231 Cells
Two lncRNAs were found that are significantly up regulated in MSC-activated MDA-MB-231 cells compared to MDA-MB-23 cell cultured alone. The suggestive most abundant isoforms of each lncRNA were determined and expression levels were tested by qRT-PCR. Data was normalized to 18S and expressed as fold induction ± SEM (n=6). ** p-value < 0.01 (two-sided student's t-test)

4.2 Regulation of LncRNA Expression

In the current model, bone-marrow-derived MSC enter the bloodstream and lodge to the site of tumor formation where they interact and activate breast cancer cells (KARNOUB et al., 2007). Upon this MSC-activation of cancer cells, two different pathways were shown to play a major role in regulating the enhanced metastatic properties of these cells (summarized in Figure 16). MSC-activated cancer cells express increased levels of the transcription factor lysyl oxidase (LOX), which triggers a signaling cascade that results in the overexpression of Twist-related protein 1 (TWIST), a key regulator of epithelial-to-mesenchymal transition (EMT) (EL-HAIBI et al., 2012). However, simultaneously, MSC-activated cancer cells induce expression of several microRNAs, including miR-A2 and miR-214 that result in knockdown of the transcriptional repressor protein Forkhead box P2 (FOXP2) (Cuiffo BG, unpublished data). Interestingly, it has been shown that knockdown of this repressor protein in cancer cells, results in an expansion of the cancer stem cell population (Cuiffo BG, personal communication). The current model therefore suggests that MSC activation shapes the breast cancer cells that ultimately, results in their increased metastatic potential, through enhancement of the EMT program together with the induction of cancer stem cell formation.

Therefore, one aim of the present study was to investigate the mechanisms of MSC induction and the potential role of the novel lncRNAs lnc-AC016722.1.1-1:1 and lnc-KCNJ9-2:2 in either the LOX or microRNA pathway.

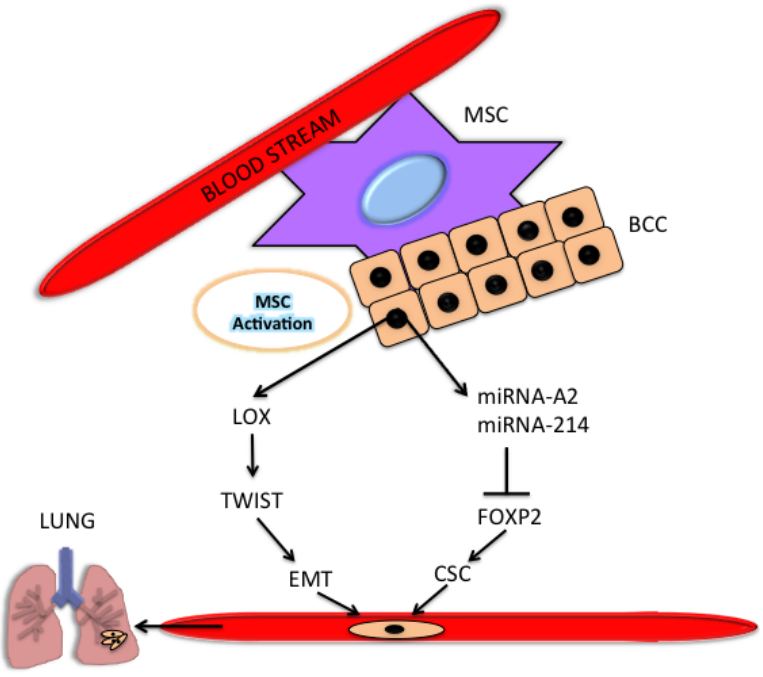

Figure 16 The Current Model of MSC-Induced Metastasis
MSC migrate from the bone marrow to the site of primary tumor lesion (KARNOUB et al., 2007). Upon interaction with cancer cells, two separate pathways become activated.
LOX overexpression is triggered by MSC in BCC and results in TWIST up-regulation and EMT (EL-HAIBI et al., 2012). MSC-induced microRNAs (A2 + 214) lead to knockdown of FOXP2, resulting in an increased cancer stem cell population.
Together, both pathways lead to the promotion of breast cancer cell metastasis (Cuiffo BG, personal communication).

4.2.1 Induction of lncRNAs by MSCs is Contact-Dependent

Next it was investigated if the up-regulation of lncRNAs found in MSC-induced cancer cells is caused by paracrine activity of soluble factors or is cell-contact dependent, transwell assays have been performed.

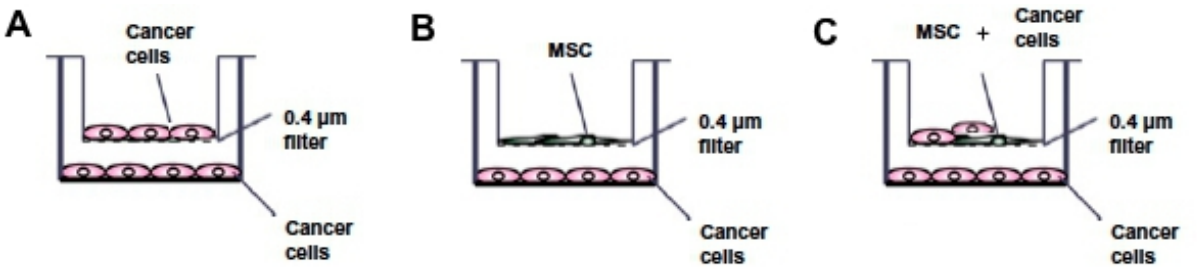

Figure 17 Transwell Assays
Adapted from (EL-HAIBI et al., 2012)
Transwell assays were performed under three different conditions to determine the mechanism of MSC-induced lncRNA expression. The 0.4 µm pore size cell culture insert physically separates cells from the bottom of the 24-well plate but allows the exchange of soluble factors. (A) Negative control experiment. No MSC involved. (B) MSC on top of BCC to probe for paracrine effects of MSC-derived factors. (C) MSC and BCC on top of BCC. To determine if initial cell-contact is needed to trigger release of paracrine factors that lead to induction of lncRNAs.

MDA-MB 231 cells were cultured in transwell inserts along with MDA-MB-231, MSC or both (Figure 17). If transcription of the lncRNAs is induced in a paracrine way, it was expect that BCC cultured with transwell inserts of MSC show increased levels of their expression. However, to test if an initial physical contact between MSC and BCC is needed to trigger the expression of paracrine factors, which then can activate lncRNA expression, MSC+BCC were co-cultured in a transwell insert on top of BCC. Additionally, as a negative control experiment, BCC were cultured on top of BCC, therefore no lncRNA expression was expected in this condition. After 72 hours of incubation, the RNA of the bottom BCC was isolated and the expression levels of lnc-AC016722.1.1-1:3 and lnc-KCNJ9-2:2 were tested by qRT-PCR .

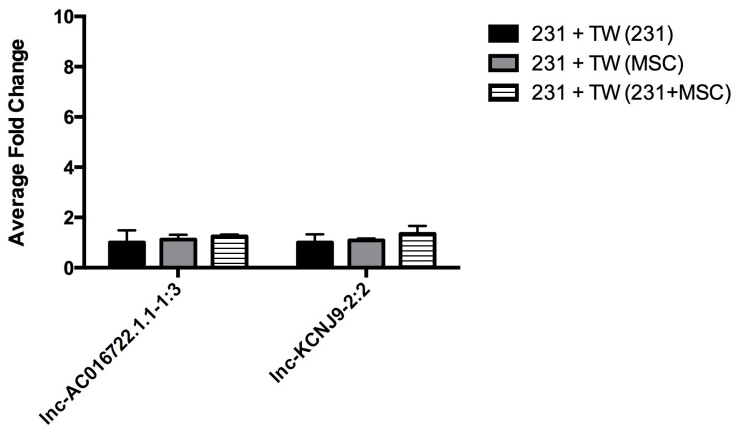

Figure 18 Mechanism of LncRNA Induction by MSC
MDA-MB-231 cells cultured with transwell inserts along with MDA-MB-231, MSC or MSC+MDA-MB-231. After 72h, inserts were removed and RNA isolated. Expression levels of lnc-AC016722.1.1-1:3 and lnc-KCNJ9-2:2 were tested by qRT-PCR. Data was normalized to 18S and expressed as fold induction ± SEM (n=3).

No significant induction of lnc-AC016722.1.1-1:3 and lnc-KCNJ9-2:2 was seen in any of the conditions tested in the transwell assay, excluding that lncRNA induction is due to paracrine effects (Figure 18). Therefore, it might be concluded that their expression is dependent on physical contact with MSC.

4.2.2 Induction of Lnc-KCNJ9-2:2 Is Not Induced by Ad-MSC and WI-38 cells

Next, it was investigated if the induction of lncRNA expression in cancer cells might be specific to bone marrow derived MSC or can be induced by similar stromal cells. Therefore, lncRNA expression levels were tested upon induction of BCC with other stromal cells. Human adipose-derived MSC (Ad-MSC) and the embryonic lung fibroblast cell line WI-38 share a similar spindle-like morphology and are positive for MSC markers CD44, CD73 and CD105 (ALT et al., 2011). Therefore, these cell lines were used to activate MDA-MB-231 upon co-culture and expression levels of lnc-KCNJ9-2:2 were determined, subsequently.

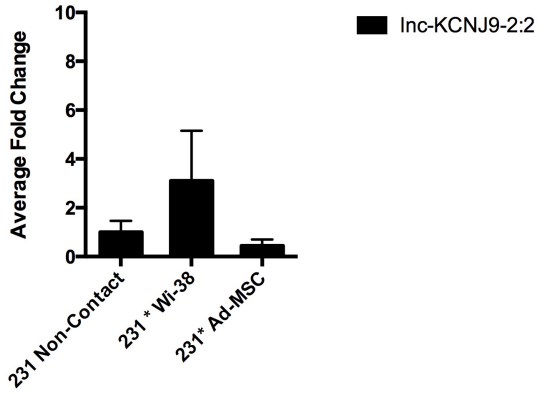

Figure 19 Lnc-KCNJ9-2:2 Expression Is Not Induced by Ad-MSC or WI-38
GFP-labeled MDA-MB-231 cells were cultured alone, or co-cultured with either Ad-MSC or WI-38 cells for 72 hours and recovered by cell sorting. RNA was isolated and provided by Dr. Cuiffo (BIDMC) to test for expression levels of lnc-KCNJ9-2:2 by qRT-PCR. Data was normalized to 18S and expressed as fold induction ± SEM (n=3).

It was found that MDA-MD-231 which were co-cultured with WI-38 cells exhibited a 3 fold increase of lnc-KCNJ9-2:2 compared to control cells. However, it has to be mentioned that the differences in expression were not statistically significant. MDA-MB-231 cells that were cultured together with Ad-MSC displayed no increase in lncRNA expression (Figure 19). In summary, the strong induction of lnc-KCNJ9-2:2 that was seen upon co-culture with bone-marrow derived MSC could not be reproduced using similar stromal cells and therefore appears to be a specific feature of this cell type.

4.2.3 Expression of MSC-induced lncRNAs is Independent of LOX

LOX expression is activated by MSC in MDA-MB-231 cells, as described previously (EL-HAIBI et al., 2012).

To investigate if lnc-AC016722.1.1-1:3 and lnc-KCNJ9-2:2 might be induced downstream of the LOX pathway, their expression levels were tested in LOX overexpressing BCC.

The pCMV LOX MYC overexpression cell line displayed 6 fold more LOX as the MDA-MB-231-control cell line. However, no significant changes in lnc-AC016722.1.1-1:3 or lnc-KCNJ9-2:2 expression were found (Figure 20). This data suggests that these lncRNAs are not downstream targets of LOX and therefore regulated differently.

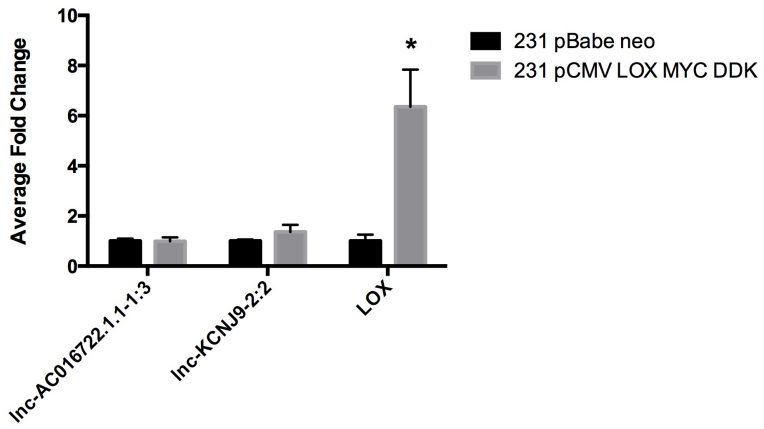

Figure 20 Expression of MSC-Induced lncRNAs Is Independent of LOX
LOX overexpressing 231 cells (developed by Dr. El-Haibi, BIDMC) were cultured for 72 hours and lncRNA expression levels were compared to controls by qRT-PCR. Data was normalized to 18S and expressed as fold induction ± SEM (n=3). * p-value < 0.05 (two-sided student's t-test)

4.2.4 MSC-induced MicroRNAs Influence Lnc-KCNJ9-2:2 Expression

Next, it had to be determined if expression of both lncRNAs could be associated with the microRNA pathway that is also induced in MDA-MB-231 cells upon MSC-activation (Cuiffo BG, personal communication). Therefore, the expression levels of lnc-AC016722.1.1-1:3 and lnc-KCNJ9-2:2 were tested in a miR-A2 and miR-214 overexpressing MDA-MB-231 cell line developed by Dr. Cuiffo (BIDMC). Additionally, the mRNA levels of the transcriptional repressor protein FOXP2 were tested, since it is a direct target of these microRNAs and served as an expression control (Cuiffo BG, personal communication). As shown in Figure 21, knockdown of FOXP2 was determined in MDA-MB-231 miR-A2 and miR-214 overexpressing cells, indicating that these miRNAs are indeed expressed and lead to knockdown of FOXP2.

However, the lncRNA lnc-AC016722.1.1-1:3 was not detected in either control or the miR-A2/214 overexpression cell lines, thereby indicating that its expression might be regulated though a different pathway (data not shown). In contrast, lnc-KCNJ9-2:2 was approximately 13 fold overexpressed in the miR-A2/214 cell line. Although the difference observed was statistically less significant than desired (p-value = 0.079), the results still indicates that a connection exists between lnc-KCNJ9-2:2 expression, miR-A2 and miR-214 overexpression and FOXP2 knockdown.

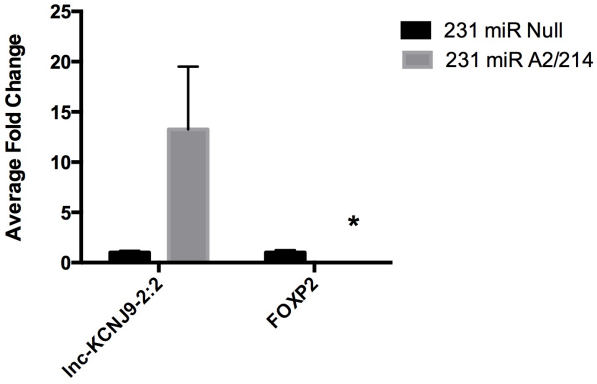

Figure 21 Expression of Lnc-KCNJ9-2:2 Is Influenced by miR A2/214
Expression levels of lnc-KCNJ9-2:2 and FOXP2 were tested in miR-A2 and miR-214 overexpressing MDA-MB-231 cells (provided by Dr. Cuiffo, BIDMC) and compared to empty vector control cells (231 miR-Null) by qRT-PCR. Data was normalized to 18S and expressed as fold induction ± SEM (n=6). * p-value < 0.05 (two-sided student's t-test)

4.2.5 The Transcriptional Regulator FOXP2 Alters Lnc-KCNJ9-2:2 Expression

To test if the lncRNA KCNJ9-2:2 is regulated by FOXP2, its expression levels were tested in MDA-MB-231 cells after FOXP2 knocked down through stable expression of a specific small hairpin RNA (shRNA) (Figure 22). Expression levels of lnc-KCNJ9-2:2 were compared to MDA-MB-231 cells with stable expression of shRNA that have no target in the genome (231 shRNA scramble).

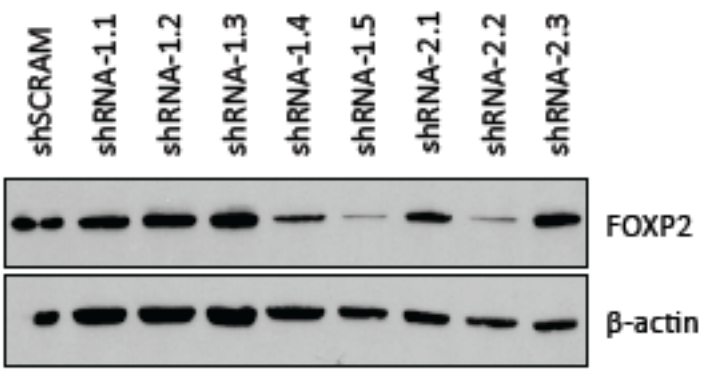

Figure 22 Knockdown of FOXP2 in shRNA Cell Lines
(from Cuiffo BG, unpublished data)
Western blot analysis was performed by Dr. Cuiffo (BIDMC). Protein levels of FOXP2 in MDA-MB-231 shRNA scramble or MDA-MB-231 cells stably overexpressing shRNA against FOXP2 were tested. Beta-actin was used as loading control (Cuffio BG, unpublished data).

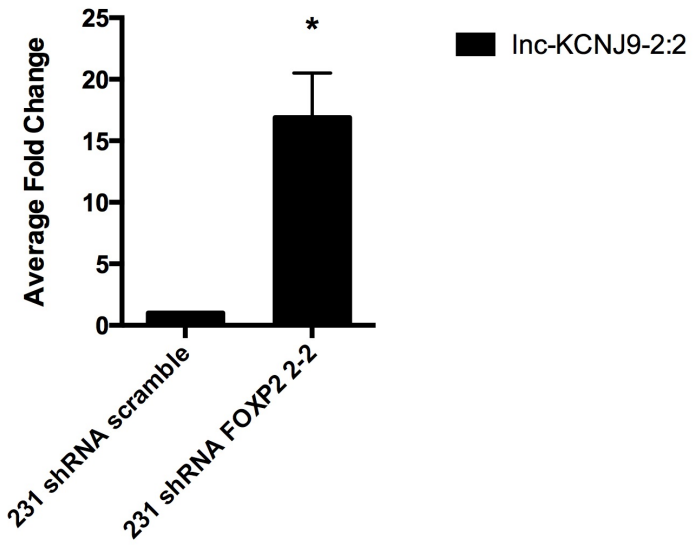

Figure 23 FOXP2 Regulates Lnc-KCNJ9-2:2 Expression
Expression levels of lnc-KCNJ9-2:2 were tested in MDA-MB-231 shRNA-FOXP2 cells (provided by Dr. Cuiffo, BIDMC) and compared to expression in MDA-MB-231 shRNA scramble control by qRT-PCR. Data was normalized to 18S and expressed as fold induction ± SEM (n=3).* p-value < 0.05 (two-sided student's t-test)

Upon stable knockdown of FOXP2 in MDA-MB-231 cells (Figure 22), the expression of lncRNA KCNJ9-2:2 was significantly increased (> 15 fold) as compared to the controls (Figure 23).

Taken together, this result suggests that lnc-KCNJ9-2:2, which is repressed in regular MDA-MB-231 cells, can be induced by MSC due to knockdown of the FOXP2 repressor protein, which is a target of miR-A2 and miR-214).

4.2.6 Promoter Analysis of Lnc-KCNJ9-2:2 Reveals FOXP2 Binding Motifs

The genomic sequence 2 kb immediate up-stream of lnc-KCNJ9-2:2 was screened for potential FOXP2 binding motifs. This was done because it is very likely that within this sequence binding sites for FOXP2 could be found (Figure 24).

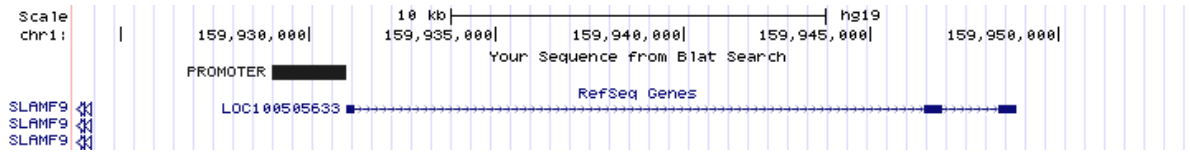

Figure 24 Screening of lnc-KCNJ9-2:2 Promoter Sequence
The sequence 2 kb up-stream of lnc-KCNJ9-2:2 (which derives from the RefSeq Gene LOC100505633) was retrieved using UCSC Genome browser.

Sequence was screened for the FOXP2 binding site (CAAATT) and the FOXP2 core binding motif (AAAT) (STROUD et al., 2006).

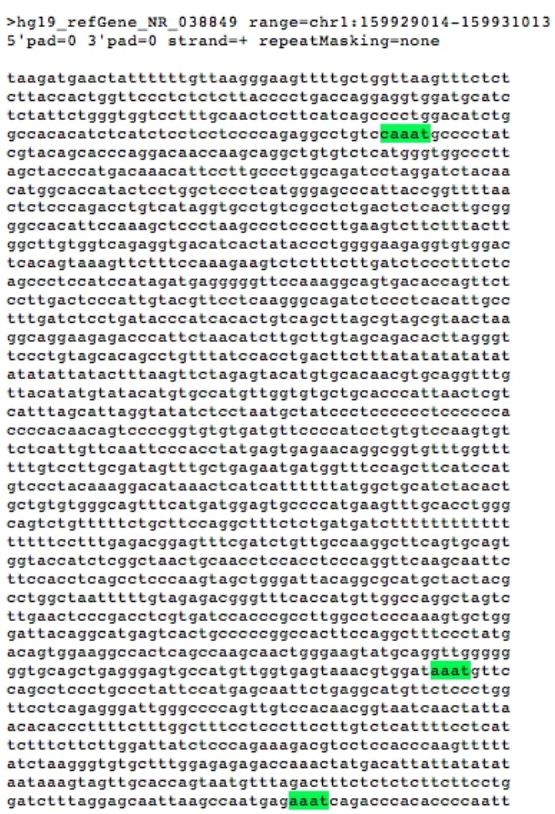

Figure 25 FOXP2 Binding Sites within Lnc-KCNJ9-2:2 Promoter

Screening of the 2 kb promoter region of lnc-KCNJ9-2:2 revealed three potential binding sites for FOXP2. The potential binding motifs are highlighted in green.

As shown in Figure 25, screening of the promoter region of lnc-KCNJ9-2:2 revealed multiple potential FOXP2 binding sites (highlighted in green).

Taken together, the results suggest a direct proximal association between the transcriptional regulator protein FOXP2 and expression of lnc-KCNJ9-2:2 in MDA-MB-231 cells. Whether the identified putative FOXP2 motifs present in the analyzed promoter are functional awaits further investigation.

4.3 MSC-induced LncRNAs in Clinical Breast Cancer Samples

4.3.1 Overexpression of Lnc-KCNJ9-2:2 in Breast Cancer

To investigate the relevance of the MSC-induced lncRNAs in a clinical setting, their expression was analyzed in 73 tumor samples from clinical breast cancer patients. These samples consist of different breast cancer subtypes, including HER2-positive, Luminal A, Luminal B and the triple-negative/basal-like breast (BLC) carcinoma. Expression levels of lnc-KCNJ9-2:2 and lnc-AC016722.1.1-1:3 were determined and compared to 5 control samples of normal breast tissue from patients that harbor no cancer.

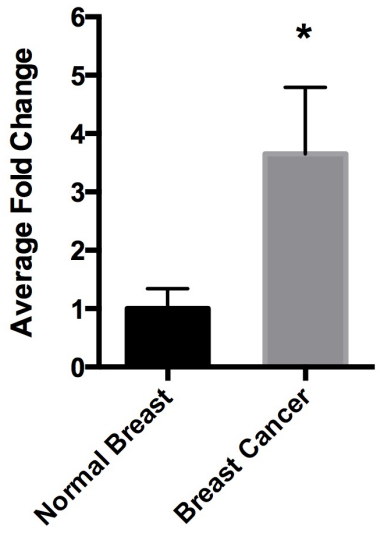

Figure 26 Lnc-KCNJ9-2:2 Expression In Breast Cancer Samples
Expression levels of lnc-KCNJ9-2:2 were tested by qRT-PCR in breast cancer samples from different origin (n=73) and normal breast control samples (n=3). Data was normalized to 18S and expressed as fold induction ± SEM. Statistical significance was tested using two-sided Welch's corrected t-test.
* p-value < 0.05

Indeed, lnc-KCNJ9-2:2 was found to be on average 3.7 fold overexpressed in clinical breast cancer samples as compared to normal breast patient samples (Figure 26). However, lnc-AC016722.1.1-1:3 was not detectable in breast cancer and control samples. While the findings indicated that lnc-KCNJ9-2:2 might play an important role in MSC-driven breast cancer malignancy, the clinical relevance of lnc-AC016722.1.1-1:3 remained unclear.

4.3.2 Overexpression of Lnc-KCNJ9-2:2 in Basal-Like Breast Cancer

Next, the overexpression of lnc-KCNJ9-2:2 in different specific breast cancer subsets was determined.

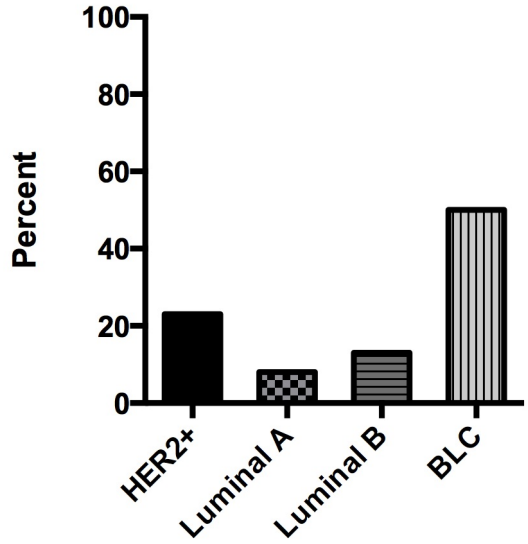

Figure 27 Breast Cancer Samples with Lnc-KCNJ9-2:2 Expression > 2 Fold
Fold changes of HER2+ (n=13), Luminal A (n=25), Luminal B (n=23), BLC (n=12) cancer samples were compared. Samples with at least 2 fold more lnc-KCNJ9-2:2, as compared to normal breast tissue control samples were counted.

It can be shown that 50% of all tested BLC/triple-negative cancer samples have at least 2 two times higher levels of lnc-KCNJ9-2:2 as compared to the normal control samples (Figure 27). Therefore, it can be assumed that overexpression of this lncRNA can be found more frequently in triple-negative/BLC breast cancer than in the other cancer subtypes.

4.3.3 Expression of Lnc-KCNJ9-2:2 in Various Malignant Diseases

After lnc-KCNJ9-2:2 was found to be significantly overexpressed in clinical breast cancer samples, its expression levels were analyzed in different types of malignant diseases by comparing large-scale genomic datasets. For this approach the cBioPortal for Cancer Genomics was used (CERAMI et al., 2012; GAO et al., 2013a) to compare data provided by the Cancer Genome Atlas (TCGA). As shown in Figure 28, lnc-KCNJ9-2:2 is amplified in various malignant diseases, besides breast cancer. Interestingly, the data suggested that lnc-KCNJ9-2:2 might be found to an even higher extent in hepatocellular carcinoma (16% of all cases) and lung adenocarcinoma (14% of all cases) as compared to invasive breast carcinoma (13% of all cases).

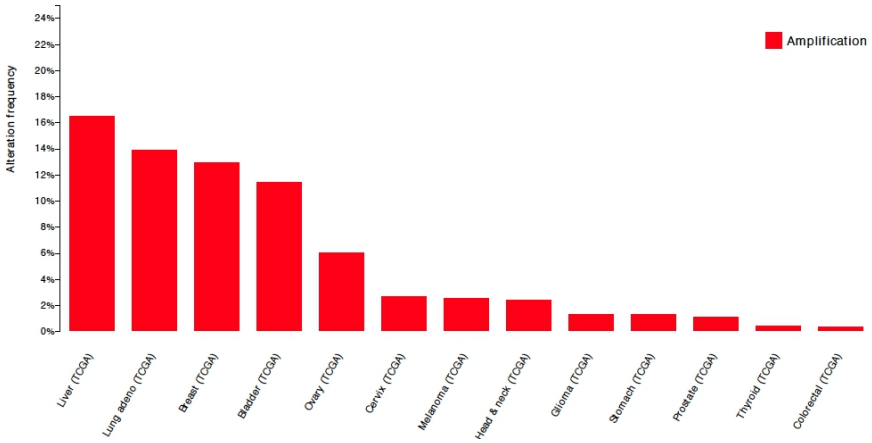

Figure 28 Lnc-KCNJ9-2:2 Is Amplified In Various Malignant Diseases
Alterations in the expression of the RefSeq gene LOC100505633 from which the lncRNA KCNJ9-2:2 derives were compared between different malignant diseases. Number of cases with a gene set alteration was expressed in percent. Liver Hepatocellular Carcinoma (Liver, n=139), Lung Adenocarcinoma (Lung adeno, n=129), Breast Invasive Carcinoma (Breast, n=760), Bladder Urothelial Carcinoma (n=26), Ovarian Serous Cystadenocarcinoma (Ovary, n=311), Cervical Squamous Cell Carcinoma and Endocervical Adenocarcinoma (Cervix, n=36), Skin Cutaneous Melanoma (Melanoma, n=228), Head and Neck Squamous Cell Carcinoma (Head & neck, n=279), Brain Lower Grade Glioma (Glioma, n=218), Stomach Adenocarcinoma (Stomach, n=219), Prostate Adenocarcinoma (Prostate, n=82), Thyroid Carcinoma (Thyroid, n=399), Colon and Rectum Adenocarcinoma (Colorectal, n=220).

To summarize, it can be stated that the novel lncRNA KCNJ9-2:2 is overexpressed in breast cancer samples and found to be particularly enriched in the triple-negative/basal-like subtype. Additionally, it was observed that lnc-KCNJ9-2:2 is amplified in a variety of malignant diseases, whereby it is most frequently found in hepatocellular carcinoma and lung adenocarcinoma.

As lnc-AC016722.1.1-1:3 was not detectable in any of the analyzed clinical breast cancer samples, the project further concentrated only on lnc-KCNJ9-2:2.

4.4 Cloning of Lnc-KCNJ9-2:2

4.4.1 Determination of 5’ and 3’ Ends of Lnc-KCN9-2:2

To elucidate the biological function of lnc-KCNJ9-2:2, an overexpression construct was developed. However, as a first step, the full-length lncRNA had to be amplified by regular PCR. To this end, a primer assay was conducted to determine the 5’ and 3’ ends of this novel lncRNA (Figure 29).

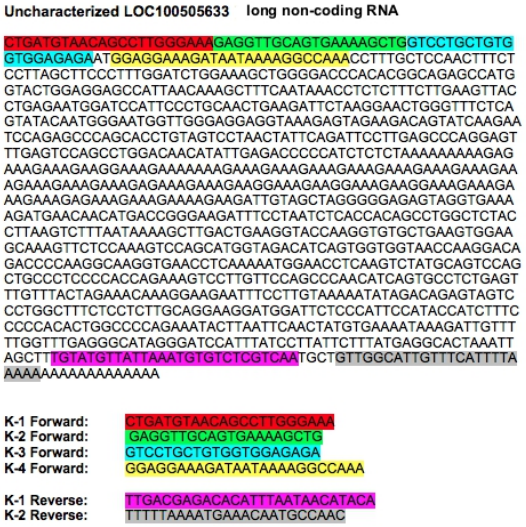

Figure 29 Determination of 5’ and 3’ Ends of Lnc-KCNJ9-2:2

To determine the real 5’ and 3’ ends of lnc-KCNJ9-2:2 a primer assay was performed. The lncRNA sequence based on the RefSeq Gene LOC100505633 was downloaded from the UCSC Genome browser and multiple forward and reverse primers were designed. Upon cross-combining them in multiple PCR reactions, the real 5’ and 3’ ends could be determined.

Only one primer combination (K4-FWD+K1-REV) resulted in the expected PCR product and, upon sequencing, could be confirmed to be lnc-KCNJ9-2:2 (Figure 30).

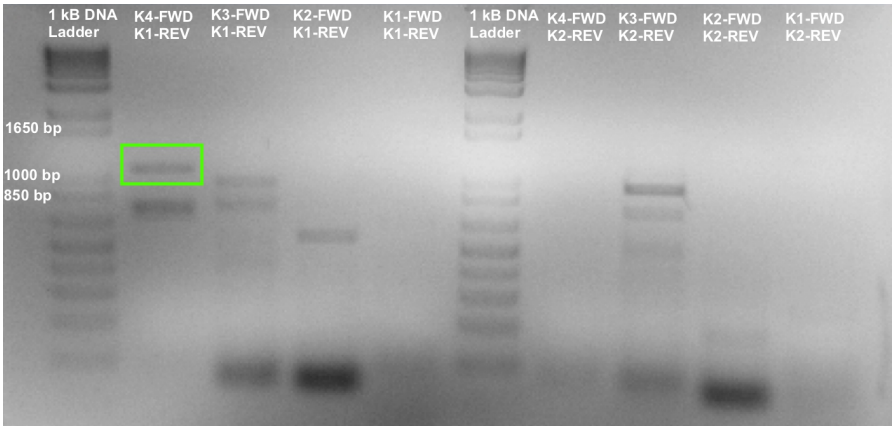

Figure 30 PCR-Primerassay To Determine 5' and 3' Ends of Lnc-KCNJ9-2:2
Forward and reverse primers were designed based on RNA-seq data of lnc-KCNJ9-2:2. Four forward primers, binding every 20 bp and starting on the suspected 5' end (K1-FWD to K4-FWD), were combined with 2 different 3' reverse primers (K-2-REV, the suggested 3' end based on RNA-seq, and 20 bp up-stream, K1-REV). PCR products that had the expected size (1040-1100 bp, depending on the primer combination) were extracted, purified and sequenced.

4.4.2 Molecular Cloning of Lnc-KCN9-2:2

The results of the PCR-primerassay identified the 5' and 3' ends of lnc-KCNJ9-2:2. The sequence of lnc-KCNJ9-2:2 can be found in the appendix (Chapter 10.4). This data was used to amplify the full-length lncRNA, including artificial restriction enzyme sites AgeI (5'-end) and EcoRI (3'-end) (Figure 31).

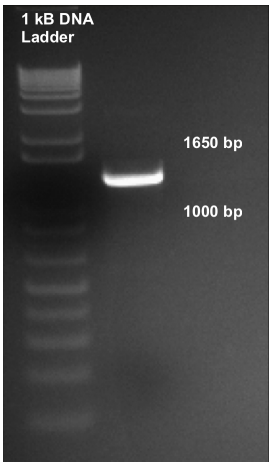

Figure 31 Amplification of Lnc-KCNJ9-2:2
The lncRNA was amplified by PCR using cloning primers that resulted in lnc-KCNJ9-2:2 with artificial AgeI (5'end) and EcoRI (3' end) restriction sites (ca. 1040 bp). The product was run on a 2 % agarose gel and then purified.

For cloning of the lncRNA overexpression construct, the backbone of the lentiviral mammalian expression vector PLJM1-GFP was used. This plasmid, due to its CMV promoter, allows very efficient expression of any cDNA (SANCAK et al., 2008).

Next, PLJM1-GFP and lnc-KCNJ9-2:2 were AgeI and EcoRI double digested, run on a 1% agarose gel and subsequently purified. This was done to remove the GFP insert and to allow an exchange with lnc-KCNJ9-2:2. Then, the purified lncRNA insert and the empty PLJM1-backbone were ligated, to develop the new plasmid PLJM1-lnc-KCNJ9-2:2 (Figure 32). The construct was transformed into Stbl3 cells and single bacterial colonies were inoculated into LB/ampicillin-media. Upon incubation, the plasmid DNA was isolated.

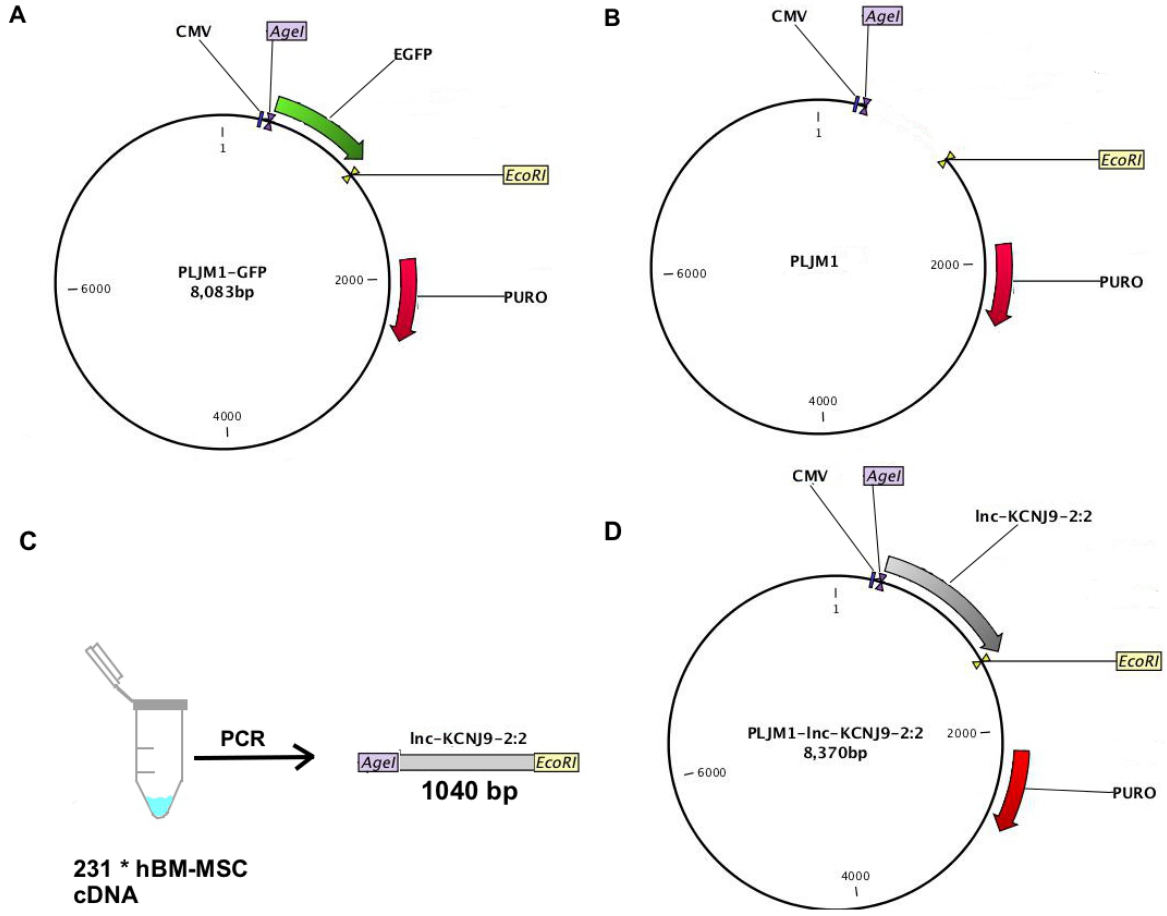

Figure 32 Cloning of the Lnc-KCNJ9-2:2 Overexpression Construct
The parental lentiviral CMV-driven vector PLJM1-GFP (A) was digested with AgeI and EcoRI to cut out the GFP insert (B). Upon the determination of the real 5' and 3' end of lnc-KCNJ9-2:2, the full-length transcript, including artificial cloning sites (AgeI and EcoRI), was amplified from cDNA of MSC-induced MDA-MB-231 cancer cells (C). After AgeI and EcoRI double digestion, the lnc-KCNJ9-2:2 insert and the PLJM1 backbone were ligated and the PLJM1-lnc-KCNJ9-2:2 construct was built (D).

To verify the new construct an AgeI and EcoRI double digest was performed to determine the band pattern. As shown in Figure 33, the expected unique digestion pattern was observed and, therefore indicated the successful swapping of the lnc-KCNJ9-2:2 insert into the PLJM1 plasmid backbone. Additionally, the new construct was verified by sequencing, which further confirmed the successful cloning of PLJM1-lnc-KCNJ9-2:2.

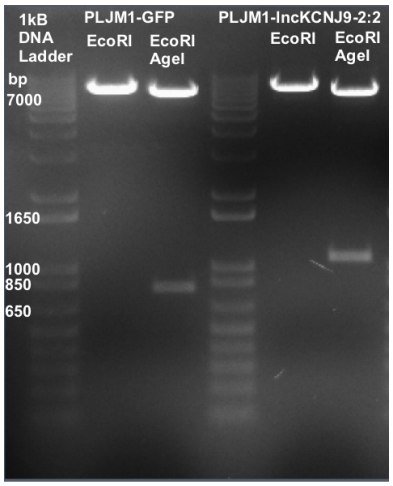

Figure 33 Plasmid Verification
PLJM1-GFP control and PLJM1-lnc-KCNJ9-2:2 plasmid were digested with EcoRI to linearize the constructs and determine their size. As shown, linearized PLJM1-GFP (8083 bp) and PLJM1-lnc-KCNJ9-2:2 (8370 bp) have the expected size. Further, AgeI and EcoRI double digest of both constructs led to the extraction of GFP (753 bp) or lnc-KCNJ9-2:2 (1040 bp) insert, respectively. The pattern shown indicates successful cloning of the lnc-KCNJ9-2:2 construct.

4.5 Development of LncRNA Overexpression Cell Lines

The highly metastatic MDA-MB-231 breast cancer cells were used for overexpression of lnc-KCNJ9-2:2 or GFP as a control. Thereby, the MSC-induction that leads to overexpression of this particular lncRNA could be phenocopied.
Additionally, lnc-KCNJ9-2:2 or GFP control were overexpressed in MCF10A cells. This cell line is an immortalized, normal breast epithelial cell line and was used to determine potential effects of this lncRNA in untransformed cells. However, the base levels of lnc-KCNJ9-2:2 in MCF10A had to be tested first and were compared to the expression levels in MDA-MB-231 cancer cells.

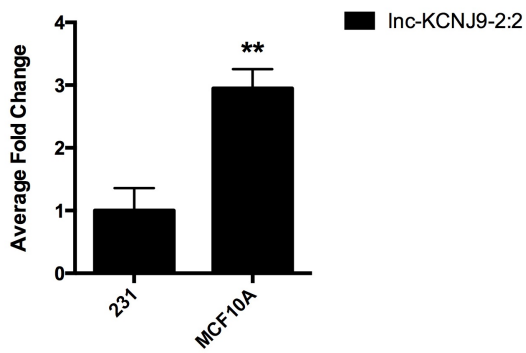

Figure 34 Base Level of Lnc-KCNJ9-2:2 in MCF10A cells
Expression levels of lnc-KCNJ9-2:2 were tested by qRT-PCR in MDA-MB-231 cells and compared to levels in MCF10A cells. Data was normalized to 18S and expressed as fold induction ± SEM (n=3). ** p-value < 0.01 (two-sided student's t-test)

Interestingly, MCF10A cells had approximately 3 fold higher levels of lnc-KCNJ9-2:2 as compared to MDA-MB-231 cells (Figure 34), indicating that expression of this lncRNA is not a restricted feature of metastatic cancer cells.

Then, the new PLJM1-lnc-KCNJ9-2:2 and the PLJM1-GFP control transfer plasmids were used to generate overexpression cell lines. First, 293T/17 cells were transfected with each transfer plasmid and additional lentiviral packaging and envelope plasmids. After 72 hours, viral supernatant was used to infect recipient cells (Figure 35). The efficiency of infection was evaluated after 48 hours by determining GFP-positive cells using fluorescence microscopy.

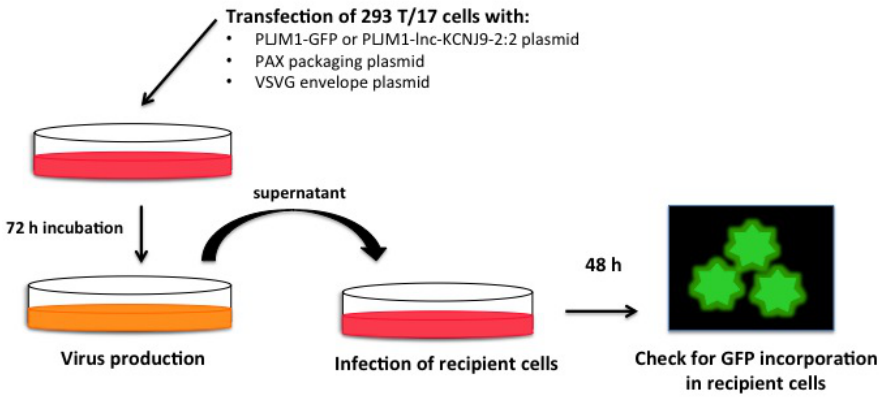

Figure 35 Infection of MCF10A and MDA-MB-231 Cells
293T/17 cells were transfected with transfer plasmids, lentiviral envelope and packaging plasmids. After 72 hours of incubation, viral supernatant was used to infect MDA-MB-231 and MCF10A cells. Control cell lines were tested for GFP incorporation.

Nearly all cells that were infected with the PLJM1-GFP lentivirus were GFP positive when screened by fluorescence microscopy, indicating the functionality of the plasmid and a very high infection efficiency (Figure 36).

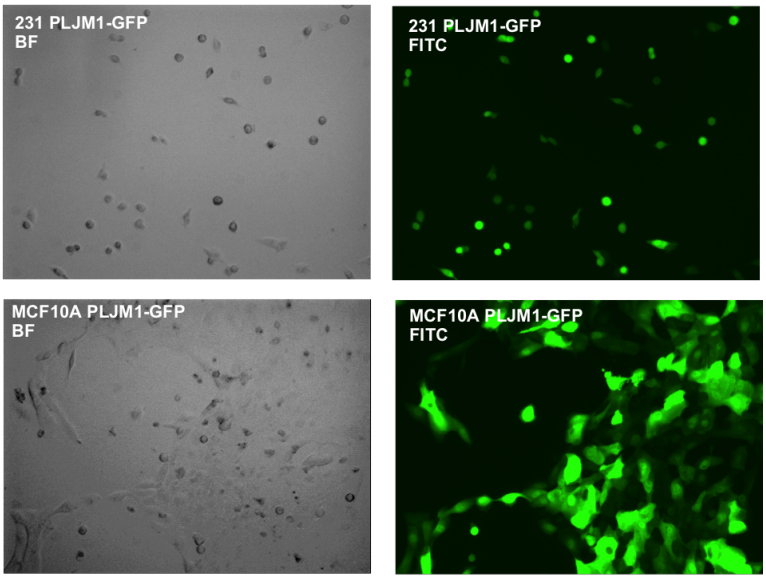

Figure 36 Verification of GFP Expression
The PLJM1-GFP infected MDA-MB-231 (upper panels) and MCF10A cell lines (bottom panels) were screened for green fluorescent cells 48 hours after infection. Almost all cells were green, therefore very high infection efficiency can be assumed.

Following infection, both cell lines were selected using puromycin. Then, the expression levels of lnc-KCNJ9-2:2 were tested by qRT-PCR.

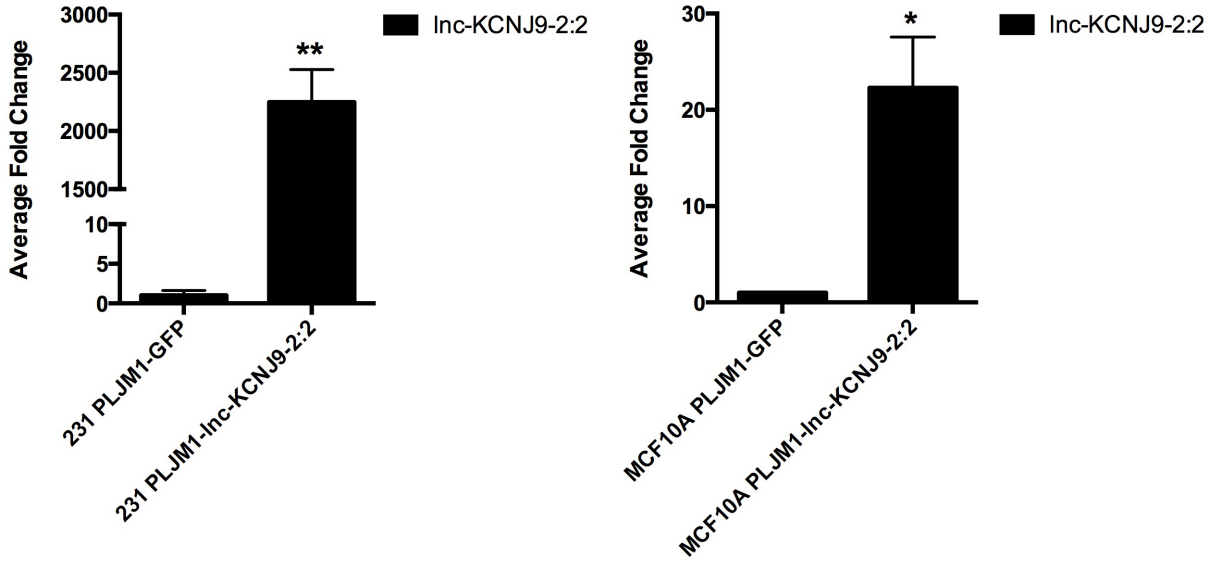

Figure 37 Verification of Lnc-KCNJ9-2:2 Overexpression Cell Lines
After selection, expression levels of lnc-KCNJ9-2:2 were tested by qRT-PCR using the bulk of PLJM1-lnc-KCNJ9-2:2 infected MDA-MB-231 and MCF10A cells. Then the levels were compared to PLJM1-GFP infected control. Data was normalized to 18S and expressed as fold induction ± SEM (n=3). ** p-value < 0.01, * p-value < 0.05 (two-sided student's t-test).

The MDA-MB-231 PLJM1-lnc-KCNJ9-2:2 overexpression cell line had a 2000 fold higher expression level of lnc-KCNJ9-2:2 than the 231 PLJM1-GFP control cell line. This result indicated that the overexpression construct worked well and that it leads to a vast increase of expression in MDA-MB-231 cancer cells. The same construct led to an approximately 20 fold increase in lnc-KCNJ9-2:2 expression in MCF10A cells when compared to the control (Figure 37).

This, however, might be explained due to the endogenous expression levels of lnc-KCNJ9-2:2 in MCF10A cells. Due to their endogenous expression, the overexpression construct may not lead to an enormous increase when compared to control MCF10A cells that exhibit already high amounts of lnc-KCNJ9-2:2. In summary, it was shown that the PLJM1-lnc-KCNJ9-2:2 construct is functional and leads to high expression levels in different cell types.

4.6 Influence of Lnc-KCNJ9-2:2 on EMT and Cancer Stem Cells

As a next step, the lnc-KCNJ9-2:2 overexpression cell lines were tested for EMT markers induction and for the size of the cancer stem cell population. Both characteristics can be found at enhanced levels upon induction of cancer cells with MSC (EL-HAIBI et al., 2012).

4.6.1 Influence of Lnc-KCNJ9-2:2 on EMT

Several common EMT-markers were tested in MDA-MB-231 and MCF10A cell lines that overexpress the lncRNA KCNJ9-2:2. These included EMT-inducting transcription factors SNAIL, TWIST and SLUG as well as markers that describe characteristics of cells that are undergoing the EMT program. The latter involves markers that are expressed as loss of E-cadherin, or increased expression levels of N-cadherin, Fibronectin or alpha-SMA.

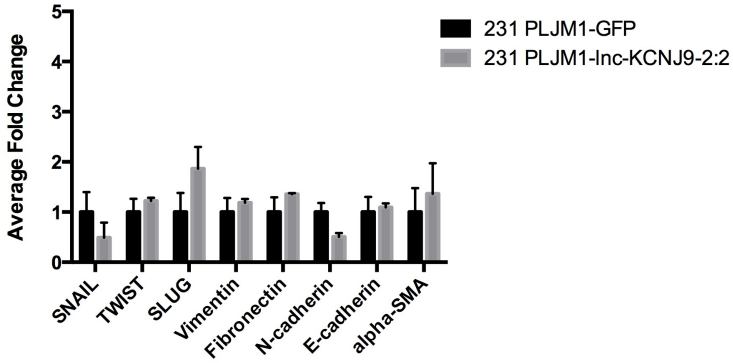

Figure 38 EMT Markers on Lnc-KCNJ9-2:2 overexpressing MDA-MB-231 Cells
Expression levels of common EMT markers were tested by qRT-PCR in the bulk of 231 PLJM1-lnc-KCNJ9-2:2 cells and compared to GFP control. Data was normalized to 18S and expressed as fold induction ± SEM (n=3).

As shown in Figure 38, no significant difference in EMT marker expression was found in MDA-MB-231 PLJM1-lnc-KCNJ9-2:2 cells, indicating that this lncRNA can not enhance the already high expression of mesenchymal markers in this cell line.

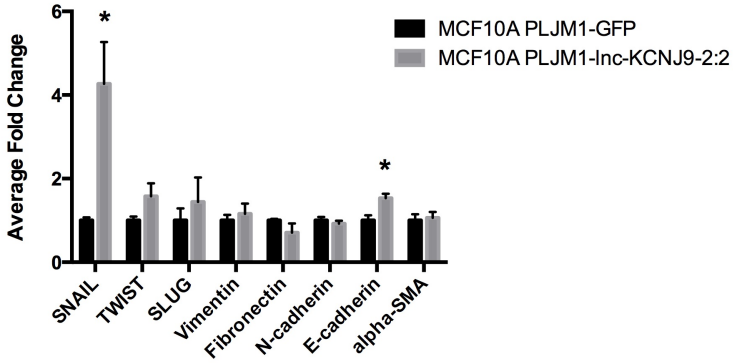

Figure 39 EMT Markers on Lnc-KCNJ9-2:2 overexpressing MCF10A Cells
Expression levels of common EMT markers were tested by qRT-PCR in the bulk of MCF10A PLJM1-lnc-KCNJ9-2:2 cells and compared to GFP control. Data was normalized to 18S and expressed as fold induction ± SEM (n=3). * p-value < 0.05 (two-sided student's t-test).

Testing of EMT markers in MCF10A cells that overexpress lnc-KCNJ9-2:2 led to the discovery of significantly increased levels of SNAIL (4 fold) and E-cadherin (1.5 fold) (Figure 39). Interestingly, these findings are contradictory, as the EMT-inducting transcription factor SNAIL is known to repress E-cadherin expression.

Taken together, the result suggests that lnc-KCNJ9-2:2 may not be a major driver of EMT of normal breast epithelial or transformed cancer cells.

4.6.2 Influence of lnc-KCNJ9-2:2 on Cancer Stem Cell Formation

To describe the potential role of lnc-KCNJ9-2:2 expression in cancer stem cell (CSC) formation, an ALDH stem cell identification assay was performed.
High activity of ALDH-1 is a characteristic found in various embryonic cells and is used a marker for adult stem cells. Moreover, in xenograft studies it was shown that only 200 breast cancer cells with very high ALDH activity are sufficient to initiate tumors in immunosuppressed mice, whereby over $2x10^3$ cancer cells with low ALDH activity failed to do so (MOREB, 2008). This finding is in accordance with the idea of cancer stem cells that have a very high tumor initiating potential. Therefore, it is assumed that high activity of ALDH-1 can be used as a marker for cancer stem cell identification.

For this assay, the puromycin selected bulk of MDA-MB-231 cells that overexpress lnc-KCNJ9-2:2 and empty vector control cells were stained using the ALDEFLUOR fluorescent substrate. Activity levels were determined using FACS. Additionally, one sample of each (test and control cell line) were treated with DEAB, a specific ALDH inhibitor, to determine fluorescence background levels (see Figure 40). The DEAB treated sample, thus indicates the amount of cells in percent that will be detected falsely positive.

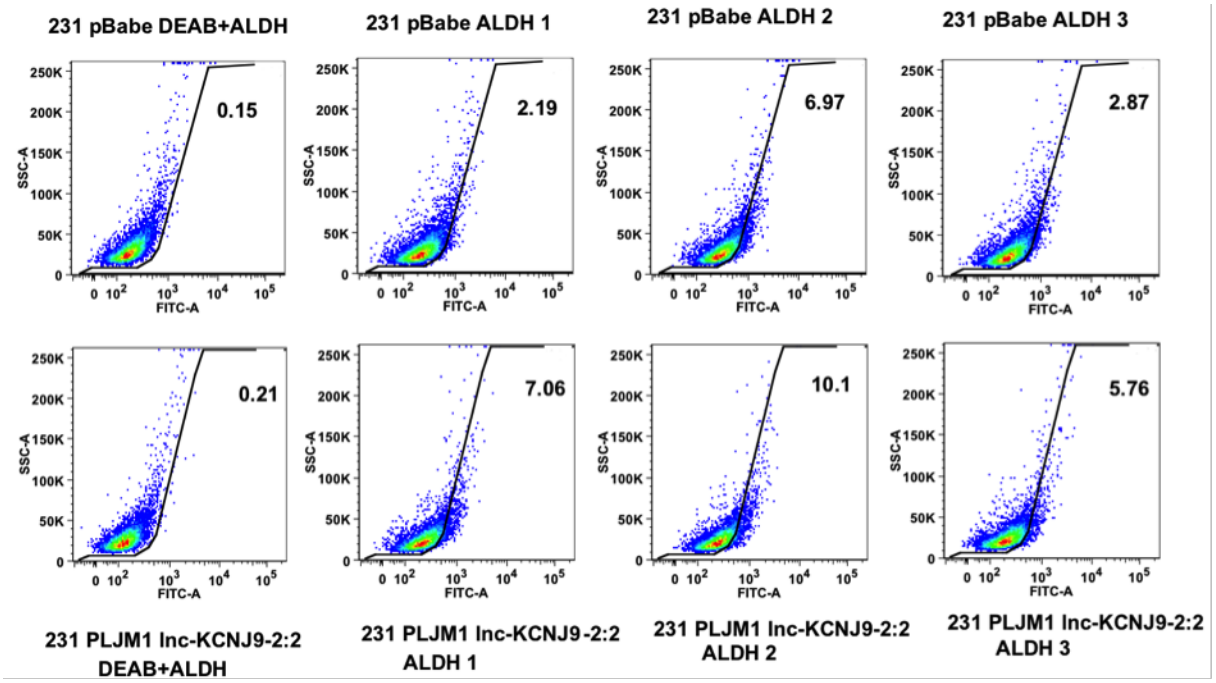

Figure 40 ALDH Stem Cell Identification Assay
Four samples of 231 PLJM1-lnc-KCNJ9-2:2 or 231 pBabe empty control, each using 1 x 10^5 cells, were stained using ALDEFLUOR reagent. One control sample of each cell line was treated with DEAB to determine background fluorescence (231 pBabe DEAB+ALDH, 231 PLJM1 lnc-KCNJ9-2:2 DEAB+ALDH). ALDH-1 activity of all samples was measured using FITC channel of BD FACS Canto II. Gates were set according to the SSC properties of breast cancer stem cells. Highly ALDH+ cells were displayed in percent of the whole population.

The ALDH positive population of 231 lnc-KCNJ9-2:2 overexpressing cells and 231 pBabe control cells was determined in percent (Figure 40). Background fluorescence of each cell line (231 pBabe DEAB+ALDH or 231 PLJM1 lnc-KCNJ9-2:2 DEAB+ALDH) was subtracted from corresponding samples (ALDH 1-3). The amount of ALDH positive cells in percent was averaged of 231 PLJM1-lnc-KCNJ9-2:2 samples and 231 pBabe control samples and compared to each other.

As a preliminary result it was demonstrated that lnc-KCNJ9-2:2 overexpressing 231 cells have an approximately 2 times larger cancer stem cell population than 231 control cells (Figure 41). However, it has to be mentioned that the difference found was statistically not significant. This can be explained by the fact that one control sample (231 pBabe ALDH 2) had abnormally high ALDH-1 activity levels.

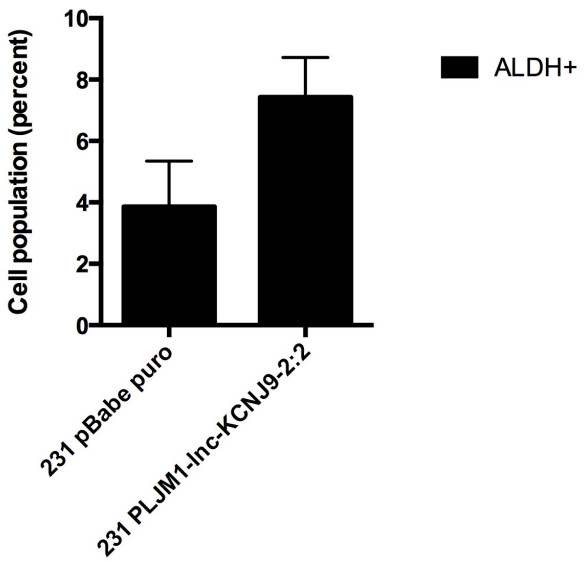

Figure 41 Influence of Lnc-KCNJ9-2:2 on the CSC Population
Averaged CSC populations of 231 PLJM1-lnc-KCNJ9-2:2 and 231 pBabe control were expressed in percent ± SEM (n=3). Statistical significance was tested using two-sided student's t-test.

Taken together, it can be stated that although this experiment needs to be repeated in order to get significant results, this data still suggests a role of the novel MSC-induced lncRNA KCNJ9-2:2 in CSC formation.

4.7 HOTAIR is Up-regulated in MSC-induced MDA-MB-231 Cancer Cells

The lncRNA HOTAIR is a well-studied transcript that was recently shown to promote breast cancer metastasis. It is a potential indicator for poor patient's prognosis in a variety of malignant diseases including breast cancer, lung cancer and pancreatic cancer (GUPTA et al., 2010; GENG et al., 2011; KIM et al., 2013b).

Microarray screening of MSC-induced MDA-MB-231 cancer cells compared to regular MDA-MB-231 cells ultimately lead to the discovery of two novel lncRNAs (lnc-AC016722.1.1-1:3 and lnc-KCNJ9-2:2) that could be verified by qRT-PCR. Since the microarray used in this study was not specifically designed to evaluate lncRNA expression, it was speculated if expression levels of HOTAIR should be tested as a candidate approach.
Interestingly, HOTAIR was significantly up regulated (>200 fold) in MSC-induced MDA-MB-231 cancer cells as compared to regular MDA-MB-231 cells (see Figure 42).

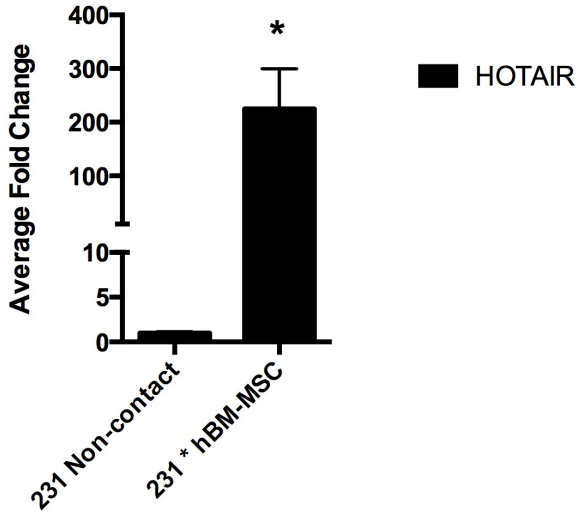

Figure 42 HOTAIR is Up-regulated in MSC-Activated MDA-MB-231 Cancer Cells
Expression levels of HOTAIR were tested by qRT-PCR in MDA-MB-231 cells cultured with or without hBM-MSC for 72 h. Data was normalized to 18S expression levels and indicated as fold induction ± SEM (n=3). Statistical significance was evaluated by two-sided student's t-test (* p-value < 0.05)

Similar to the previously tested lncRNAs it was determined whether the expression of HOTAIR is due to contact-dependent or paracrine effects of bone marrow derived MSC on MDA-MB-231 cancer cells (Figure 43). Also, it was tested if similar cell lines (adipose-derived MSC or WI-38) lead to a comparable induction (Figure 44).

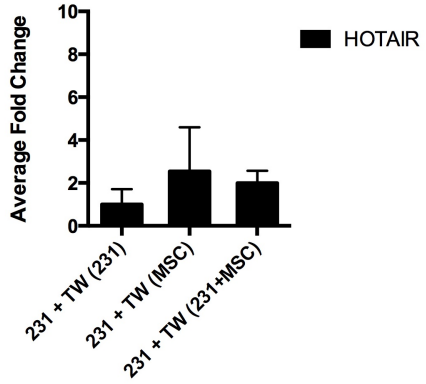

Figure 43 HOTAIR Induction might be Contact-Dependent on MSC
MDA-MB-231 cells cultured with transwell inserts of MDA-MB-231, MSC or MSC+231. After 72h, inserts were removed and RNA isolated. Expression levels of HOTAIR were tested by qRT-PCR. Data was normalized to 18S and expressed as fold induction ± SEM (n=3).

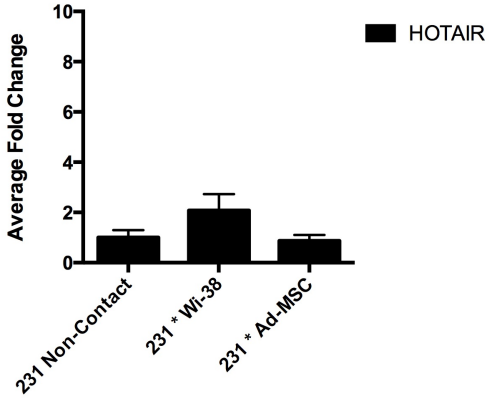

Figure 44 HOTAIR Expression might be Specific to hBM-MSC Induction
GFP-labeled MDA-MB-231 cells were cultured alone, or co-cultured with either Ad-MSC or WI-38 cells for 72 hours and recovered by cell sorting. RNA was isolated and provided by Dr. Cuiffo (BIDMC) to test for expression levels of HOTAIR by qRT-PCR. Data was normalized to 18S and expressed as fold induction ± SEM (n=3).

No significant difference in HOTAIR expression was found in BCC cultured with transwell inserts of MSC or MSC and MDA-MB-231 cancer cells when compared to the controls. This result indicated that the up-regulation of HOTAIR that was found in

MSC-activated MDA-MB-231 cells is not due to paracrine effects and may be dependent on physical contact between BCC and MSC. Moreover, it was shown that neither WI-38 cells nor adipose-derived MSC, despite their morphological similarities, could induce HOTAIR expression in BCC upon co-culture. This suggests that this feature is might be specific to bone marrow derived MSC.

4.7.1 HOTAIR is Not Downstream of MSC-induced LOX or miR-A2/214

Next, it was tested if HOTAIR expression might be a downstream target of the MSC-induced LOX pathway, or regulated by MSC-induced microRNAs that lead to FOXP2 knockdown, similarly to lnc-KCNJ9-2:2.

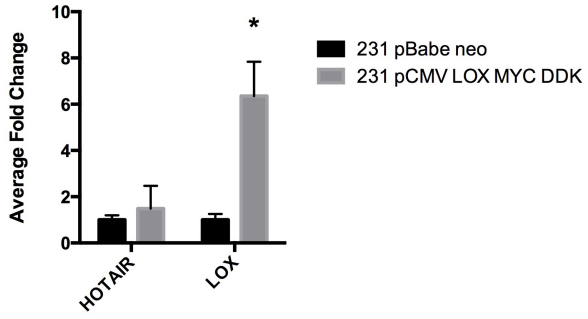

Figure 45 HOTAIR is not Influenced by LOX Overexpression
LOX overexpressing 231 cells (provided by Dr. El-Haibi, BIDMC) were cultured for 72 hours and HOTAIR expression was compared to regular MDA-MB-231 cells by qRT-PCR. Data was normalized to 18S and expressed as fold induction ± SEM (n=3). * p-value < 0.05 (two-sided student's t-test)

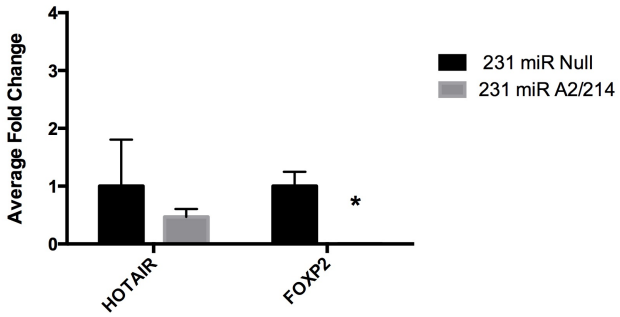

Figure 46 miR-A2 and miR-214 do not Influence HOTAIR Expression
Expression levels of HOTAIR and FOXP2 were tested in miR-A2 and miR-214 overexpressing 231 cells (provided by Dr. Cuiffo, BIDMC) by qRT-PCR and compared to control cells. Data was normalized to 18S and expressed as fold induction ± SEM (n=3). * p-value < 0.05 (two-sided student's t-test)

To determine if HOTAIR is transcriptionally activated by LOX in MSC-activated cancer cells, HOTAIR expression was tested in LOX overexpressing 231 pCMV LOX cells and compared to 231 pBabe control cells. While the overexpression cell line had increased levels of LOX (>6 fold), which proved its functionality, no significant differences were found in HOTAIR expression (Figure 45). Therefore, it can be assumed that LOX does not lead directly to HOTAIR expression in MDA-MB-231 cancer cells. Interestingly, 231 cells overexpressing miR-A2 and miR-214, which lead to FOXP2 knockdown, also did not have increased levels of HOTAIR when

compared to MDA-MB-231 control cells. This result indicates that in contrast to lnc-KCNJ9-2:2, HOTAIR is not regulated by miRNA induced knockdown of the transcriptional repressor FOXP2 (Figure 46).
In summary, these results suggest that HOTAIR is induced upon MSC-activation of MDA-MB-231 cancer cells over pathways that are different from LOX signaling or the microRNA-induced FOXP2 knockdown effects.

4.7.2 MSC Induce Expression of Estrogen Receptors and Collagen I

Recent evidence showed that the lncRNA HOTAIR is transcriptionally induced in lung cancer cells by Collagen type I (ZHANG et al., 2012). This study demonstrated that HOTAIR expression could be up regulated by supplementation of Collagen I to lung cancer cells in three-dimensional basement membrane matrix cultures (ZHANG et al., 2012). Moreover, it was shown that this induction could be diminished by using a neutralizing antibody against the integrin-receptor that mediated Collagen I signaling (ZHANG et al., 2012).

These findings were extremely interesting because, besides the LOX gene, Collagen I was found to be the gene that is up-regulated the most upon MSC-activation of MDA-MB-231 cancer cells by microarray screening (EL-HAIBI et al., 2012). MSC-induced Collagen I production in MDA-MB-231 cancer cells would provide one plausible explanation for the HOTAIR induction. Therefore, it was verified that MSC-activation leads to Collagen I overexpression in MDA-MB-231 cancer cells as suggested by microarray screening (Figure 47).
Another study suggested that HOTAIR is transcriptionally induced by estrogen in estrogen-receptor positive (ER+) breast cancer cells (BHAN et al., 2013). The MDA-MB-231 breast cancer cell line is considered to be triple-negative, indicating that they do not express ER, progesterone receptors (PR) and the human epidermal growth factor 2 (HER2) (TURNER and REIS-FILHO, 2013). As MSC-activation changes the gene expression of MDA-MB-231 cancer cells radically, it was tested if these cells may re-express ER upon MSC-activation (Figure 47).

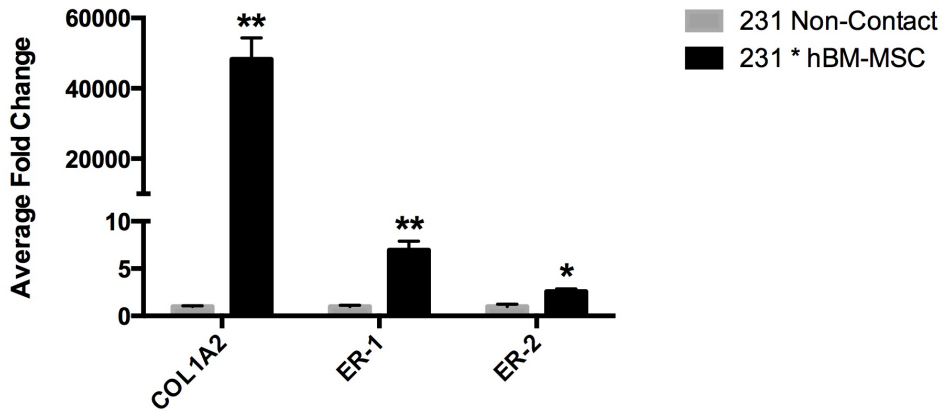

Figure 47 MSC Induce Expression of Collagen I and ER in MDA-MB-231 Cells
Expression levels of Collagen I and ER-1 and ER-2 were tested by qRT-PCR in MDA-MB-231 cells cultured with or without hBM-MSC for 72 h. Data was normalized to 18S levels and expressed as fold induction ± SEM (n=3). Statistical significance was evaluated by two-sided student's t-test (* p-value < 0.05, ** p-value < 0.01).

Not surprisingly, Collagen I was highly (>40,000 fold) up regulated in MSC-activated cancer cells as suggested by microarray screening, providing one potential mechanism of HOTAIR induction. Unexpectedly, both estrogen receptors, ER-1 and ER-2, were up regulated in MSC-activated MDA-MB-231 cancer cells. Although, the induction was relatively weak (2.5 - 7 fold), the findings were statistically significant.
This finding might play an important role in future therapy approaches of MSC-driven triple negative breast cancer, providing a potential treatment option with estrogen inhibitors.

4.7.3 Collagen I Alone is Not Sufficient for HOTAIR Induction

As a next step, it was tested if Collagen I alone is sufficient to induce HOTAIR expression in MDA-MB-231 cancer cells. Therefore, cancer cells were cultured in plates that were coated with different concentrations of human Collagen I and HOTAIR expression levels were compared (Figure 48).

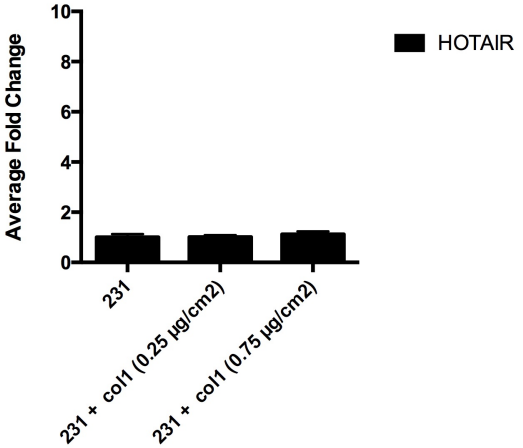

Figure 48 Collagen I alone does not Induce HOTAIR Expression
HOTAIR expression was tested by qRT-PCR in MDA-MB-231 cells cultured on plates that were thin layer-coated with varying concentrations of Collagen I. Data was normalized to 18S and expressed as fold induction ± SEM (n=3).

No significant difference was observed between MDA-MB-231 cancer cells cultured on top of 0.25 µg/cm^2 and 0.75 µg/cm^2 Collagen I coated plates as compared to MDA-MB-231 cultured in uncoated dishes. This result suggests that Collagen I alone is not sufficient to induce HOTAIR expression to levels that are comparable to MSC-induced cells. Therefore, further triggers are needed in order to induce HOTAIR expression in regular MDA-MB-231 cancer cells.

4.8 Development of the HOTAIR Overexpression Construct

To investigate the molecular pathology of HOTAIR, an overexpression construct was developed. The full-length HOTAIR sequence, including artificial AgeI and EcoRI restriction sites on the 5' and 3' end, was amplified by PCR from PLZRS-HOTAIR plasmid. Upon double digestion, PLJM1-empty backbone was ligated with digested HOTAIR insert, resulting in the generation of the new PLJM1-HOTAIR plasmid (Figure 49).

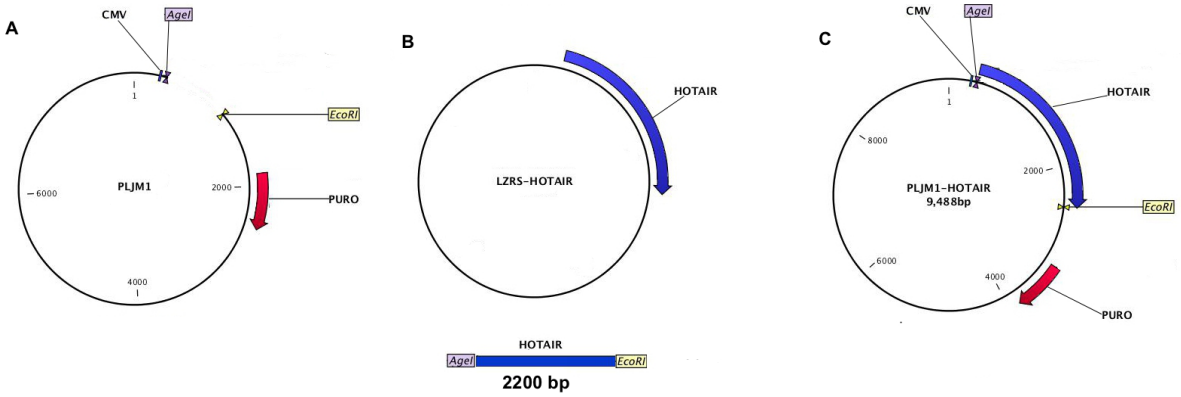

Figure 49 Cloning of the HOTAIR Overexpression Construct
(A) The empty PLJM1 was used as a backbone. (B) For the development of the HOTAIR overexpression construct, the full-length transcript including artificial cloning sites (AgeI and EcoRI) was amplified from LZRS-HOTAIR. (C) Upon digestion with respective restriction enzymes, HOTAIR and PLJM1 backbone were ligated to create the PLJM1-HOTAIR construct.

The new construct was transformed into Stbl3 cells and plasmid DNA was isolated from single bacterial colonies. The construct was verified by restriction enzyme digestion, by evaluating the unique pattern that is only found in PLJM1-HOTAIR, and additionally by sequencing (Figure 50).

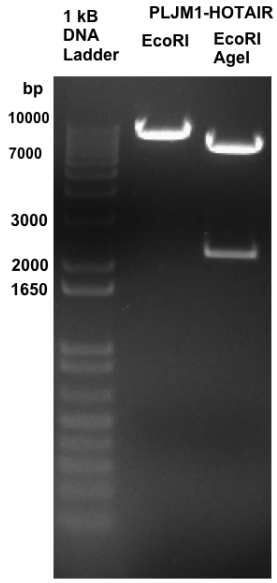

Figure 50 PLJM1-HOTAIR Verification
PLJM1-HOTAIR was digested with EcoRI to linearize the complete construct (~ 9500 bp) or double digested with AgeI and EcoRI to extract HOTAIR (~ 2200 bp). The expected unique pattern is shown.

Before the HOTAIR overexpression cell lines were produced, HOTAIR levels were measured in MCF10A breast epithelial cells and compared to MDA-MB-231 cancer cells (Figure 51).

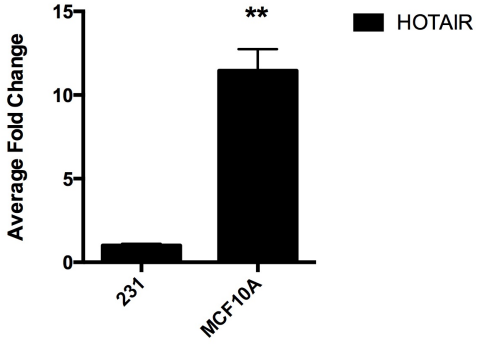

Figure 51 Base Level of HOTAIR in MCF10A
HOTAIR expression levels were tested by qRT-PCR in MCF10A and MDA-MB-231 cells. Data was normalized to 18S and expressed as fold induction ± SEM (n=3). ** p-value < 0.01 (two-sided student's t-test)

Interestingly, MCF10A cells had approximately a 10-fold higher HOTAIR expression compared to MDA-MB-231 cells, thereby indicating that expression is not restricted to malignant cancer cells.

To generate MCF10A and MDA-MB-231 HOTAIR overexpression cell lines, the PLJM1-HOTAIR plasmid was transfected, together with lentiviral packing and envelope plasmids into 293T/17 cells. After 72 hours of incubation viral supernatant was used to infect both MDA-MB-231 and MCF10A recipient cell lines (Figure 35).

Following infection, overexpressing cells were selected with puromycin and HOTAIR expression levels were determined, subsequently.

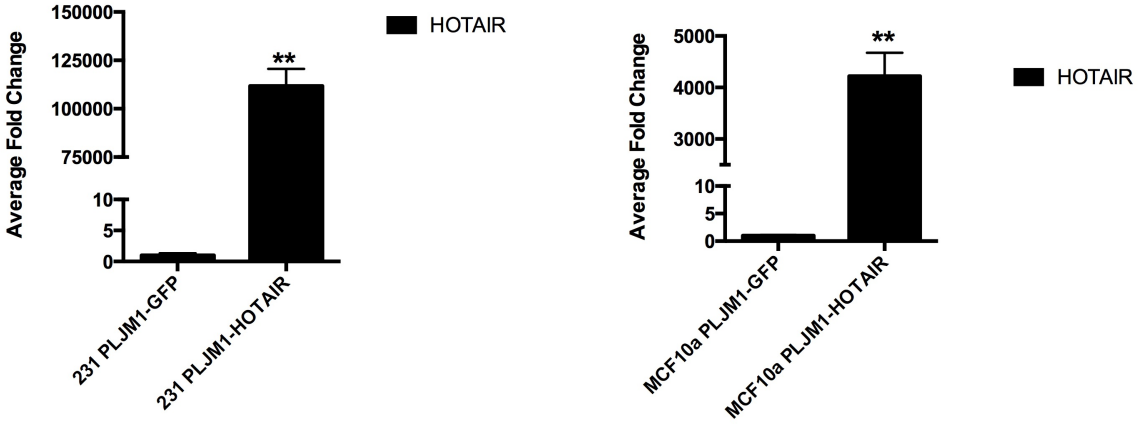

Figure 52 Verification of HOTAIR Overexpression Cell Lines
Expression levels of HOTAIR were tested by qRT-PCR in the puromycin-selected bulk of PLJM1-HOTAIR infected MDA-MB-231 (left figure) and MCF10A cells (right figure) and compared to PLJM1-GFP infected control. Data was normalized to 18S and expressed as fold induction ± SEM (n=3). ** p-value < 0.01 (two-sided student's t-test).

Both, the 231 PLJM1-HOTAIR and the MCF10A PLJM1-HOTAIR cell lines had greatly increased HOTAIR expression levels as compared to GFP control cell lines (Figure 52), indicating the functionality of both overexpression cell lines. HOTAIR expression was increased to >10,000 fold in 231 PLJM1-HOTAIR and >4000 fold in MCF10A PLJM1-HOTAIR cells as compared to control. However, one has to keep in mind that MCF10A cells have a higher endogenous level of HOTAIR as compared to MDA-MB-231 cells, explaining the difference in expression efficiency between these cell lines.

4.9 Influence of HOTAIR on EMT and Cancer Stem Cells

MCF10A and 231 HOTAIR overexpressing cells were tested for different EMT marker expressions and the size of their CSC population was evaluated.

4.9.1 HOTAIR Overexpression Induces EMT in MCF10A

Expression levels of various EMT markers were tested in the puromycin-selected bulk of 231 PLJM1-HOTAIR and MCF10A PLJM1-HOTAIR cells and compared to each GFP control. While no significant difference could be observed in HOTAIR overexpressing 231 cancer cells (Figure 53), MCF10A cells displayed a significant increase of the EMT master transcription factor TWIST (2 fold) and SNAIL (>3 fold) (Figure 54). Interestingly, fibronectin, an extracellular matrix protein that is up regulated in certain EMT phenotypes (KIM et al., 2013a), was found to be significantly down regulated (0.5 fold) when compared to control MCF10A cells.

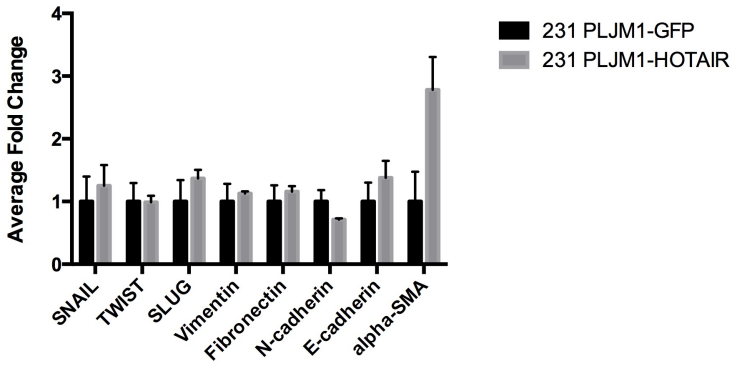

Figure 53 EMT Marker Expression on HOTAIR overexpressing 231 Cells
Expression levels of several EMT markers were tested by qRT-PCR in the puromycin-selected bulk of 231 PLJM1-HOTAIR cells and compared to GFP control. Data was normalized to 18S and expressed as fold induction ± SEM (n=3).

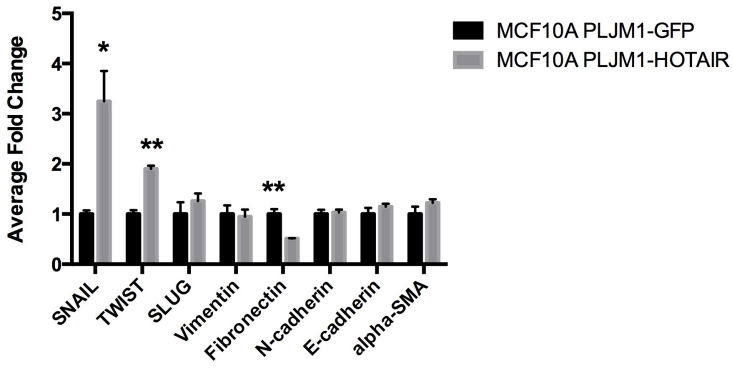

Figure 54 EMT Marker Expression on HOTAIR overexpressing MCF10A Cells
Expression levels of several EMT markers were tested by qRT-PCR in the puromycin-selected bulk of MCF10A PLJM1-lnc-KCNJ9-2:2 cells and compared to GFP control. Data was normalized to 18S and expressed as fold induction ± SEM (n=3). * p-value < 0.05, ** p-value < 0.01 (two-sided student's t-test).

Taken together these results indicate that HOTAIR overexpression may lead to EMT in normal breast epithelial cells. However, because MDA-MB-231 cancer cells already express high levels of mesenchymal markers, HOTAIR may not be able to enhance their transcriptional activation.

4.9.2 HOTAIR Overexpression Does Not Influence the CSC Population

To test the influence of HOTAIR overexpression on the expansion of the CSC (highly ALDH-positive) population within MDA-MB-231 cancer cells, an ALDH assay was performed. Four samples, including one control sample to determine background fluorescence (DEAB), of HOTAIR overexpressing 231 cells or 231 pBabe empty control cells were stained with fluorescent substrate and ALDH activity was monitored by FACS analysis. The amount of CSC within the whole cell population was determined in percent. Background fluorescence levels were subtracted from ALDH samples (Figure 55).

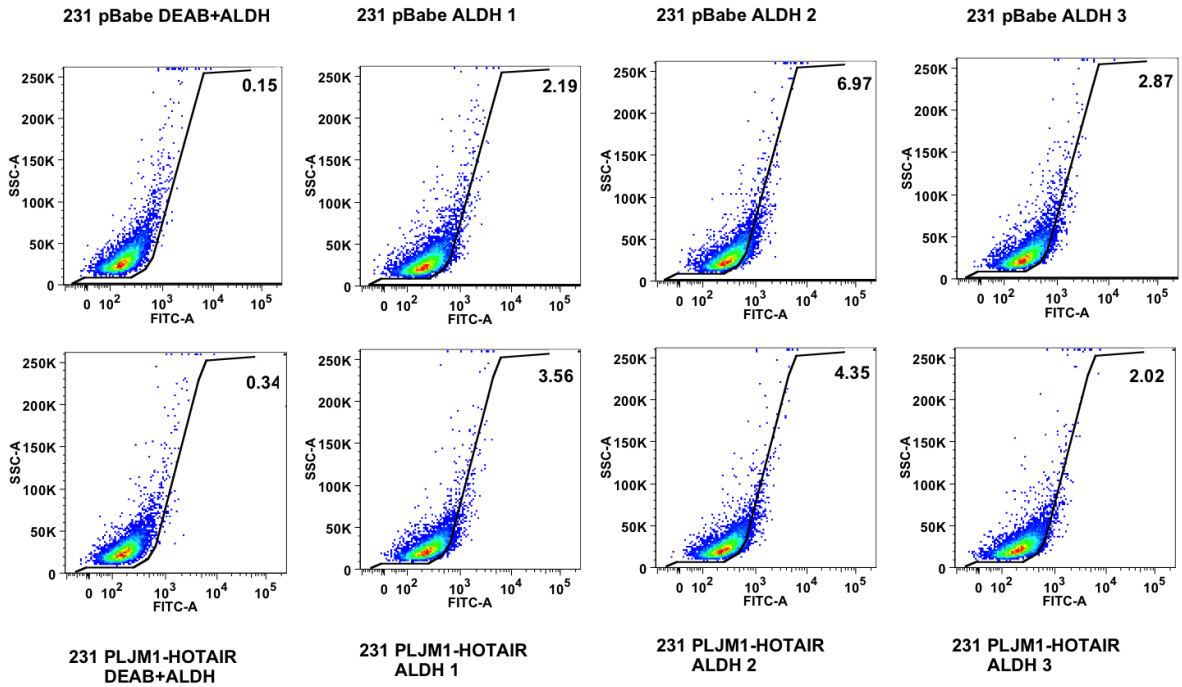

Figure 55 ALDH Assay using HOTAIR overexpressing 231 Cells
Four samples of 231 PLJM1-HOTAIR or 231 pBabe empty control, each using 1 x 10^5 cells, were stained using ALDEFLUOR reagent. One control sample of each cell line was treated with DEAB to determine background fluorescence (231 pBabe DEAB+ALDH, 231 PLJM1 HOTAIR DEAB+ALDH). ALDH-1 activity of all samples was measured using FITC channel of BD FACS Canto II. Gates were set according to the SSC properties of breast cancer stem cells. Highly ALDH+ cells were displayed in percent of the whole population.

The amount of CSC of 231 PLJM1-HOTAIR and 231 pBabe control samples was averaged and compared to each other. HOTAIR overexpressing 231 cells did not have an significantly increased amount of highly ALDH positive cells when compared to MDA-MB-231 control cells, indicating that HOTAIR might have no influence on the CSC population (Figure 56).

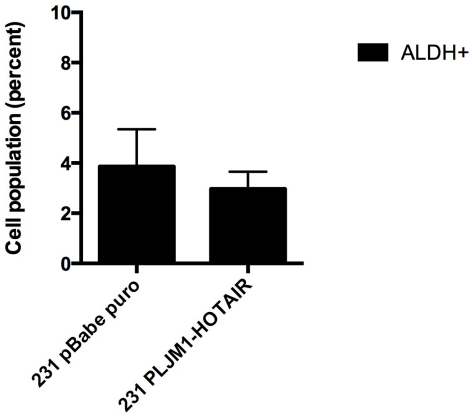

Figure 56 Influence of HOTAIR on CSC
Averaged CSC populations of 231 PLJM1-HOTAIR and 231 pBabe control were expressed in percent ± SEM (n=3). Statistical significance was tested using two-sided student's t-test.

These results indicate that the MSC-induced lncRNA HOTAIR might influence EMT, as shown by overexpression within MCF10A breast epithelial cells. However, HOTAIR does not seem to influence CSC within MDA-MB-231 cancer cells. In contrast, the lncRNA KCNJ9-2:2 did not influence EMT but led to an increased CSC population.

Taken together, it was shown that lncRNAs add a new layer of complexity to the known pathology of MSC-driven breast cancer metastasis. While the phenotypes of MSC-activated cancer cells are well described, it was demonstrated that MSC-induced lncRNAs may contribute to their progression.

5 Discussion

Breast cancer is the most common cancer disease in women worldwide, and the most important cause of tumor related mortality with an estimated number of over 400,000 deaths per year (PARKIN et al., 2005). As with many cancer diseases, most deaths are due to tumor spread via metastasis (FIDLER, 2003). In breast cancer, MSC were shown to promote cancer cell metastasis primarily through two major signaling machineries that trigger EMT and to increase cancer stem cell phenotypes (EL-HAIBI et al., 2012). Although much effort in recent years has shed some light onto the basic concepts of these processes, the exact molecular details that regulate them still remain elusive. In the past years, it became clear that lncRNAs play an important role in developmental processes and, when deregulated, may contribute to cancer progression (GUTTMAN et al., 2011; GUPTA et al., 2010). The focus of this work was to identify MSC-induced lncRNAs and to describe their potential contributions to breast cancer metastasis.

In this study, it was shown that MSC induce the de-novo expression of two novel lncRNAs (lnc-AC016722.1.1-1:3 and lnc-KCNJ9-2:2). It was demonstrated that their up-regulation in highly metastatic MDA-MB-231 cancer cells might be dependent on physical contact to MSC and that the transcriptional repressor FOXP2 might play a role in regulating lnc-KCNJ9-2:2. While the role of lnc-AC016722.1.1-1:3 remains unclear, it was shown that lnc-KCNJ9-2:2 is up-regulated in breast cancer and also amplified in a variety of other malignant diseases. Upon overexpression, this lncRNA leads to an expanded cancer stem cell population and thus might serve as a potential prognostic factor or future therapy target.

HOTAIR is a well-described lncRNA that is up regulated in a variety of cancer diseases and promotes breast cancer metastasis (GUPTA et al., 2010). Although the details of its biological functions are known, its molecular pathology is not completely understood In the second part of the study, it was demonstrated that HOTAIR is up regulated upon MSC-activation of cancer cells, indicating an important function in MSC-driven breast cancer malignancy. Further, it was shown that HOTAIR overexpression might lead to EMT in breast epithelial cells, thus providing a novel insight into its molecular function.

5.1 Identification of Novel MSC-induced LncRNAs in BCC

In 2008 a microarray analysis was performed using a whole genome expression array to compare highly metastatic MSC-induced 231 cells to control cancer cells. As this microarray was not specifically designed for long non-coding RNAs, the probe sets used in this analysis had to be verified. It was determined that many probes were unspecific and therefore, falsely positive detected lncRNAs were excluded from further analysis. To overcome this problem, recent studies started using specialized non-coding arrays to determine lncRNA expression levels (LI et al., 2013a). These newer arrays can specifically detect non-coding transcripts and due to their newer annotation will most likely be able to discriminate between distinct lncRNA isoforms.

Upon verification of the up-regulated MSC-induced lncRNAs by qRT-PCR, two lncRNA loci, namely lnc-AC016722.1.1-1 and lnc-KCNJ9-2 were identified. However, due to its design, the microarray used did not allow specific determination of the lncRNA isoform variants. This was determined by the fact that the sequences of the microarray probes could be aligned to the same exon sequence of all isoforms, when evaluated using the UCSC genome browser. Therefore, it had to be evaluated which isoform(s), of the detected lncRNA were up regulated in MSC-activated cancer cells. It could be shown that although all isoforms of both lncRNAs were transcribed in MSC-activated MDA-MB-231 cancer cells, in all cases one specific variant was more prominent than the other closely-related isoforms. Specifically, lnc-AC016711.1.1-1 isoform 3 (lnc-AC016722.1.1-1:3) and lnc-KCNJ9-2 isoform 3 (lnc-KCNJ9-2:2) were found to be >1000 fold up-regulated in MSC-activated MDA-MB-231 cancer cells, whereas all other variants were much lower, approximately 7 to 70 fold, induced.

This result is particularly interesting, because it indicates a specific regulatory mechanism behind the expression of lnc-AC016722.1.1-1:3 and lnc-KCNJ9-2:2 lncRNA transcripts. This finding is a first convincing argument supporting the idea of their functional relevance. Indeed, others could recently show that a lncRNA that is specifically induced by p53, meditates global gene repression, indicating strict regulation of lncRNAs with important cellular functions (HUARTE et al., 2010).

5.2 Regulation of Lnc-AC016722.1.1-1:3 and Lnc-KCNJ9-2:2

The current model of MSC-driven breast cancer metastasis suggests that these from bone marrow derived cells home to the tumor site and contribute there to cancer progression. MSC were shown to secrete bioactive chemical compounds (e.g. CXCL2, CXCL12/SDF-1) that act on cancer cells in a paracrine fashion, change their phenotype and foster their proliferation (RHODES et al., 2010; HALPERN et al., 2011). Also, a direct contact-dependent interaction between MSC and breast cancer cells was shown to initiate de-novo production of the chemokine CCL5 in MSC, which then acts in a paracrine way on cancer cells to induce metastasis (KARNOUB et al., 2007). These findings suggest a variety of possibilities for MSC to induce transcriptional changes in cancer cells that ultimately result in their highly malignant phenotype.

To investigate the mechanism that may up-regulate the lncRNA transcripts discovered in this study, a transwell-assay was performed. If chemical compounds secreted directly by naïve MSC transcriptionally up-regulated lnc-AC016722.1.1-1:1 or lnc-KCNJ9-2:2, one expects that BCC cultured in the MSC-media supernatant to exhibit elevated lncRNA levels. If MSC that had physical contact to BCC initiate de-novo production of a soluble factor (e.g., a chemokine) that acts in a paracrine way on distant cancer cells, one expects increased lncRNA levels in BCC cultured with conditioned media of MSC and BCC. Interestingly, expression levels of lnc-AC016722.1.1-1:3 and lnc-KCNJ9-2:2 were not significantly different in BCC that were cultured in any of these conditions when compared to regular BCC. This suggests that the transcriptional up-regulation most likely results from direct contact-dependent effects of MSC to BCC.

Although several studies suggest that MSC that are found within the cancer microenvironment and exert pro-malignant functions, are bone marrow derived, it might be possible that they actually originate from local tissues (EL-HAIBI and KARNOUB, 2010). Especially in the breast tissue, local adipose-derived MSC might present a source of stromal cells that are attracted to the tumor site in favor of the distant bone marrow MSC. A study supporting this hypothesis demonstrated that adipose-derived stromal cells favor the tumor cell growth *in vivo* (YU et al., 2008). Therefore, it was tested if other stromal cells, besides bone marrow derived MSC can

induce transcriptional up-regulation of lnc-AC016722.1.1-1:3 and lnc-KCNJ9-2:2. However, analysis in MDA-MB-231 cancer cells that were either stimulated with Ad-MSC or WI-38 embryonic lung fibroblasts indicated no significant difference in lncRNA expression levels. Although both cell types are morphologically very similar and express markers that are frequently found on bone marrow derived MSC (ALT et al., 2011), these cells were not able to induce lncRNA expression. This suggests that induction of these lncRNAs might be a specific feature of bone marrow derived MSC.

Cancer-associated fibroblasts (CAF), play an important role in the cancer microenvironment, can regulate tumor progression, and were recently shown to contribute to cancer cell metastasis (GAO et al., 2013b; LUGA and WRANA, 2013). Mesenchymal stem cells were reported to differentiate into CAF (MISHRA et al., 2008). It will be of great interest in the future, to investigate if CAF can resemble changes in lncRNA expression, similar to MSC.

Analysis of the transcriptome of MSC-activated MDA-MB-231 cancer cells revealed many major changes in the gene expression pattern. Importantly, LOX, a secreted amine oxidase that normally regulates cross-linking of Collagen and elastin in the extracellular matrix (ECM), was found to be one of the most prominent up-regulated genes in MSC-activated cancer cells (EL-HAIBI et al., 2012). In the context of metastasis, it was demonstrated that overexpression of this particular amine oxidase in cancer cells results in induction of EMT, mainly due to transcriptional activation of TWIST (EL-HAIBI et al., 2012). These cells have a very aggressive phenotype and display increased motility and invasion. Besides the enhanced EMT program, MSC-activated cancer cells, however, display an enriched cancer stem cell population, which cannot be explained by LOX overexpression (EL-HAIBI et al., 2012). For this reason, it was suggested that the pathway, that regulates LOX-induced EMT is different from the one that leads to an increased CSC population. Recent evidence proves that a network of microRNAs that is induced by MSC-activation of cancer cells leads to knockdown of the transcription factor FOXP2 (Cuiffo BG, unpublished data). This gene which was originally associated with speech development (VICARIO, 2013) could recently be linked to prostate-specific antigen (PSA) recurrence in prostate cancer (STUMM et al., 2013). In MSC-activated cancer cells, however, microRNA-induced knockdown of FOXP2 results in strong expansion of the

CSC population (Cuiffo BG, unpublished data). Additionally, MDA-MB-231 cancer cells with FOXP2-knockdown were shown to have a very high tumor-initiating potential in immunosuppressed mice, as compared to regular MDA-MB-231 cells (Cuiffo BG, unpublished data).

Therefore, it was tested if MSC-induced lncRNAs (lnc-AC016722.1.1-1:3 or lnc-KCNJ9-2:2) were a target gene of LOX and may result in EMT or of the microRNA network that leads to FOXP2 knockdown and subsequent CSC expansion. Indeed, lnc-KCNJ9-2:2 was found to be >13 fold up-regulated upon overexpression of MSC-induced miR-A2 and miR-214 in MDA-MB-231 cancer cells. The result was not statistically significant, but gave a first hint on the regulation of this particular lncRNA. Both MSC-induced microRNAs lead to FOXP2 knockdown (Cuiffo BG, unpublished data). Therefore, it was tested if lnc-KCNJ9-2:2 may be induced in MDA-MB-231 cells with stable knockdown of FOXP2 due to specific shRNA expression. Surprisingly, lnc-KCNJ9-2:2 was found to be significantly up-regulated (>15 fold) upon knockdown of FOXP2 in MDA-MB-231 cells as compared to controls.

This result demonstrated that the expression of this particular lncRNA, which seems to be repressed by FOXP2 under normal conditions, can be induced upon knockdown of FOXP2. Further supporting this hypothesis, it was found that the 2 kb immediate up-stream DNA region of lnc-KCNJ9-2:2 contains several potential FOXP2 binding motifs. These findings indicate that FOXP2 may repress transcription of lnc-KCNJ9-2:2 by binding within its promoter region. However, upon MSC-induced knockdown of FOXP2, the promoter region is not blocked anymore and allows transcriptional activation of lnc-KCNJ9-2:2. Consistent with the idea that LOX induced EMT and microRNA-induced knockdown of FOXP2 are integrated into two different pathways upon MSC-induction, it was shown that LOX overexpression in MDA-MB-231 cells did not result in lnc-KCNJ9-2:2 expression.

While lnc-KCNJ9-2:2 could be associated as a target gene of the MSC-induced microRNA network and subsequent FOXP2 knockdown, lnc-AC016722.1.1-1:3 seemed to be regulated differently. This lncRNA was not significantly induced upon LOX or miR-A2 and miR-214 overexpression, suggesting that there might be additional triggers needed that lead to its transcriptional activation.

5.3 Biological Function of Lnc-AC016722.1.1-1:3 and Lnc-KCNJ9-2:2

Next generation sequencing techniques allowed the discovery of thousands of lncRNAs. However, up to now less than 1 % of all have been actually functionally characterized (TSAI et al., 2011). A common method to identify lncRNAs that may play a relevant role in disease progression is to compare their expression between normal and tumor samples (HUARTE and RINN, 2010).

To test the functional relevance of lnc-AC016722.1.1-1:3 and lnc-KCNJ9-2:2, their expression was determined within a cohort of 73 breast cancer samples and compared to 5 normal breast tissue samples. The cancer samples comprised all common clinical subtypes including luminal A, luminal B, HER2-positive and basal-like/triple-negative breast cancer. In summary, lnc-KCNJ9-2:2 expression was found to be on average 3.7 fold overexpressed in clinical breast cancer samples. The result was statistically significant and provided a first hint of the clinical relevancy of this novel lncRNA. Furthermore, it was demonstrated that 50 % of all basal like cancer samples have an at least 2 times higher expression of lnc-KCNJ9-2:2 than normal control samples. While this indicated that this particular lncRNA might be enriched in certain basal-like/triple negative breast cancer samples, overexpression in other cancer subtypes was found at lower levels.

The triple-negative breast cancer subtype itself comprises a highly diverse collection of cancers (TURNER and REIS-FILHO, 2013). Detailed molecular characterization of this subtype will be of clinical significance in future. It might be possible that high expression of lnc-KCNJ9-2:2 is a trait that is found only in a certain subgroup of basal-like/triple-negative breast cancer and thus might serve as a potential biomarker. Currently, the use of lncRNAs as potential biomarker is under exhaustive investigation (ZHANG et al., 2012; ARITA et al., 2013). Moreover, expression of particular lncRNAs, including HOTAIR and MALAT-1 were recently shown to be a marker for poor patients prognosis (SCHMIDT et al., 2011; NAKAGAWA et al., 2013). It will be of interest to evaluate the prognostic value of lnc-KCNJ9-2:2 expression in breast cancer patients in the future. Importantly, by comparing large-scale genomic dataset it could be shown that lnc-KCNJ9-2:2 is amplified in a variety

of malignant diseases apart from breast carcinoma, including lung adenocarcinoma and hepatocellular carcinoma. This result may indicate its functional relevancy and suggests that it may be important in the progression of various malignant diseases.

While the acquired data suggests that the MSC-induced lnc-KCNJ9-2:2 might be of clinical significance, lnc-AC016722.1.1-1:3 was not detectable in any clinical breast cancer sample. Therefore, further efforts of this thesis concentrated on determining the function of lnc-KCNJ9-2:2.

To evaluate the molecular function of lnc-KCNJ9-2:2, an overexpression construct was developed. The construct mimicked the MSC-induced expression of this particular lncRNA in BCC. The lentiviral PLJM1-GFP vector served as a backbone and was selected due to its CMV promoter that allows mammalian expression (SANCAK et al., 2008). The GFP insert was swapped with lnc-KCNJ9-2:2 sequence and the resulting new construct was used for overexpression in MCF10A breast epithelial and MDA-MB-231 cancer cells. Verification of the overexpression cell lines proved the functionality of the new PLJM1-lnc-KCNJ9-2:2 construct and resulted in an approximately 2000 fold increase of lncRNA expression levels in MDA-MB-231 cells as compared to PLJM1-GFP control cells.

Therefore, the expression levels of lnc-KCNJ9-2:2 overexpressing 231 cells were comparable to levels found in MSC-activated cancer cells. However, the same construct led to an only approximately 20 fold overexpression of lnc-KCNJ9-2:2 in MCF10A cells. This can be explained by the surprisingly high endogenous levels of lnc-KCNJ9-2:2 in this particular cell line. In the future it will be interesting to further evaluate the function of lnc-KCNJ9-2:2 by knocking down this lncRNA in normal MCF10A cells.

To test the potential influence of lnc-KCNJ9-2:2 on EMT, 231 PLJM1-lnc-KCNJ9-2:2 cells were tested for enhanced expression of several EMT markers. No significant difference could be determined between the overexpression cell line and regular MDA-MB-231 control cells. This result indicates that lnc-KCNJ9-2:2 may not have an influence on EMT. However, it has to be mentioned that regular MDA-MB-231 cancer cells already have a very aggressive phenotype and are highly metastatic and

express mesenchymal markers (CHAO et al., 2010). Therefore, it might be possible that lnc-KCNJ9-2:2 is not able to further enhance EMT in this particular cell line because it is already pushed to high levels.
Upon testing of lnc-KCNJ9-2:2 overexpressing MCF10A cells it was found that one of the EMT master-regulators, called SNAIL, is 4 fold (significantly) overexpressed when compared to control MCF10A cells. While this would actually indicate that lnc-KCNJ9-2:2 might be a driver of EMT, it was controversial that E-cadherin was 1.5 fold overexpressed in the same cell line. Loss of the cell adhesion protein E-cadherin is considered one major hallmark of EMT that results in tumor progression (PEINADO et al., 2007). Even more importantly, SNAIL is a known transcriptional repressor of E-cadherin (BATLLE et al., 2000). Therefore, it was very surprising to find enhanced levels of SNAIL and E-cadherin at the same time in lnc-KCNJ9-2:2 overexpressing MCF10A cells. Taken together, it remains to be determined if the lncRNA KCNJ9-2:2 is a *bona-fide* driver of EMT.

To test the potential influence of lnc-KCNJ9-2:2 on the formation of cancer stem cells, the 231 lnc-KCNJ9-2:2 overexpressing cells were tested for ALDH positive cells and compared to control 231 cells. High ALDH-1 activity is found in embryonic cells and CSC seem to express increased levels as well (MOREB, 2008). Interestingly, it could be demonstrated that lnc-KCNJ9-2:2 overexpressing 231 cells have an approximately 2 fold larger CSC population than MDA-MB-231 control cells. Although this finding was not statistically significant, it still indicates that the MSC-induced lncRNA KCNJ9-2:2 increases the CSC population in MDA-MB-231 cells.

This finding is in accordance with its demonstrated regulation by FOXP2. Knockdown of FOXP2 was shown to lead to an increased CSC population and enhanced tumor-initiating potential of MDA-MB-231 cancer cells (Cuiffo BG, unpublished data). Moreover, this study demonstrated that knockdown of FOXP2 results in significantly increased levels of the novel lncRNA KCNJ9-2:2 in MDA-MB-231 cancer cells. Therefore, it is not completely surprising that this lncRNA may have a potential role in CSC formation. As a next step, tumor initiation studies with cancer cells that overexpress lnc-KCNJ9-2:2 need to be done to better evaluate the role of this lncRNA on CSC formation.

5.4 MSC Induce Up-regulation of HOTAIR in Breast Cancer Cells

The lncRNA HOTAIR is overexpressed in many primary breast tumors and to an even higher extent in distant metastases (GUPTA et al., 2010). Moreover, it was shown that HOTAIR overexpression is a powerful predictor of future metastasis and an indicator for patient's prognosis in a variety of malignant cancer diseases (GUPTA et al., 2010; GENG et al., 2011; KIM et al., 2013b; LI et al., 2013b; LIU et al., 2013). Currently, it is postulated that HOTAIR promotes re-localization of the protein group polycomb repressive complex 2 (PRC2), and due to methylation, epigenetically silences metastasis suppressor genes (GUPTA et al., 2010). Despite its biological function, the molecular pathology that leads to increased metastasis remains largely unknown.

In this study it was demonstrated that MSC-activated cancer cells, which are highly metastatic compared to regular cancer cells, display increased levels of HOTAIR. This result matches well with the observation that HOTAIR itself induces cell invasion and may lead to breast cancer metastasis (GUPTA et al., 2010). Additionally, it was shown that induction of HOTAIR in MSC-activated cancer cells is independent of LOX and MSC-induced microRNAs. In contrast to lnc-KCNJ9-2:2, HOTAIR seems to be independent of the transcriptional repressor FOXP2, suggesting an alternative way of MSC-dependent induction in cancer cells.

Recent evidence suggested that HOTAIR might be transcriptionally activated by the extracellular matrix protein Collagen I in lung cancer cells (ZHUANG et al., 2013). Microarray screening of MSC-activated MDA-MB-231 cancer cells revealed that these cells express high amounts of Collagen when compared to regular MDA-MB-231 cells. Thus, Collagen I expression levels in MSC-activated MDA-MB-231 cell were confirmed by qRT-PCR and it was shown that they express >40,000 fold more Collagen I when compared to control. However, when cultured on Collagen I thin-layer coated dishes, regular MDA-MB-231 cells did not have significantly increased amounts of HOTAIR. This suggests that additional triggers are needed to induce the expression.

It was demonstrated that HOTAIR is not significantly induced upon LOX overexpression in MDA-MB-231 cells. Besides its pro-malignant functions, LOX is normally needed for cross-linking of extracellular matrix proteins, e.g. collagen. Therefore, it might be interesting in the future to determine HOTAIR levels in cancer cells that overexpress both, LOX and Collagen I.

In another study, it was shown that estrogen might lead to HOTAIR induction in ER-positive breast cancer cell lines (BHAN et al., 2013). As MSC lead to many transcriptional changes upon induction of cancer cells, it was tested whether MSC-activated cancer cells express increased levels of ER. The triple negative MDA-MB-231 cancer cell line does not express any ER, thus it was surprising that it could be shown that these cells start to re-express ER-1 and ER-2 upon MSC-activation.

Although this finding does not explain the HOTAIR induction that can be found in MSC-activated cancer cells, it might point out future therapy approaches for treatment of MSC-driven breast cancer malignancy. Triple-negative breast cancer comprises a diverse group of often very aggressive-cancer types with poor prognosis and few treatment options (XU et al., 2013a). Therefore, the finding that MSC-activated triple negative cancer cells may re-express estrogen receptors might provide a future therapy target.

5.5 HOTAIR may Induce EMT of Breast Epithelial Cells

To evaluate the molecular function of HOTAIR, an overexpression construct was developed. The MCF10A breast epithelial cell line and the MDA-MB-231 cancer cell line were infected with the new designed PLJM1-HOTAIR construct and overexpression was verified by qRT-PCR. Interestingly, no significant difference in EMT marker expression was determined between HOTAIR overexpressing MDA-MB-231 cells and MDA-MB-231 control cells. This might be due to the fact that the MDA-MB-231 cancer cells already express high levels of mesenchymal markers (CHAO et al., 2010). Therefore, HOTAIR may not be able to further induce their expression in these cancer cells. However, when EMT marker expression was

evaluated in HOTAIR overexpressing MCF10A cells, it was determined that these cells express increased levels of TWIST and SNAIL when compared to control.
These two master transcription factors are strongly associated with EMT and overexpression might induce their transition to mesenchymal cells (PEINADO et al., 2007). Supporting this data, a recent study could demonstrate that knockdown of HOTAIR led to reversion of the EMT phenotype in gastric cancer cells (XU et al., 2013b). This finding further indicates that HOTAIR might play an important role in EMT. Although HOTAIR might be a potential driver of EMT in breast epithelial cells, no evidence was found that this particular lncRNA influences ALDH activity in cancer cells. Therefore, it can be assumed that MSC-induced HOTAIR may contribute to cancer progression by EMT but does not influence expansion of CSC.

5.6 The Updated Model of MSC-driven Breast Cancer Malignancy

LncRNAs play important roles in gene regulation of cells. Similar to protein coding transcripts lncRNAs can when deregulated contribute to cancer progression. This study adds a new layer of complexity to the currently established model of MSC-driven breast cancer malignancy. MSC are thought to travel to the site of primary tumor lesion, activate cancer cells and thereby contribute to their progression. Due to the MSC-activation, protein-coding genes, miRNAs and, as shown in this study, lncRNAs become deregulated in breast cancer cells. Besides the established LOX-TWIST axis that promotes EMT and the microRNA network that leads to CSC, data from this study suggests that MSC-induced lncRNAs might contribute to the development of both of these phenotypes.

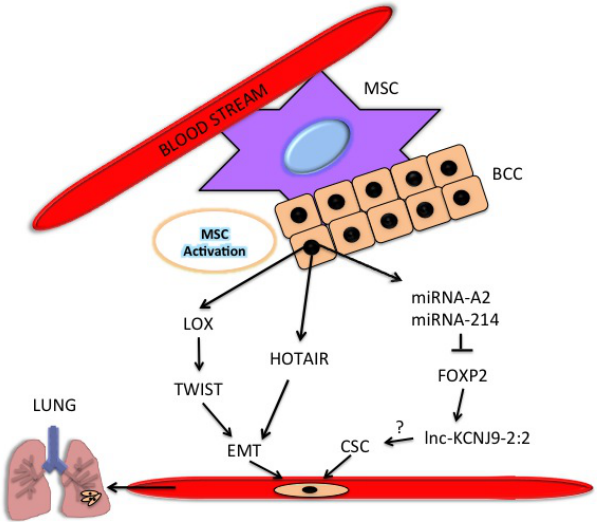

Figure 57 The Updated Model of MSC-Driven Breast Cancer Malignancy
LOX overexpression is triggered by MSC, leading to TWIST up-regulation and EMT. MSC-induced microRNAs lead to knockdown of FOXP2, resulting in an increased stem cell population within the cancer cells. Also, lnc-KCNJ9-2:2 is up-regulated upon FOXP2 knockdown and might contribute to CSC expansion. Additionally, MSC-induced HOTAIR might directly enhance EMT. All these triggers together, ultimately lead to the highly malignant phenotype of MSC-induced cancer cells.

6 Zusammenfassung

Brustkrebs in Frauen ist die am häufigsten auftretende Krebserkrankung der Welt und die wahrscheinlichste Ursache für eine Frau an Krebs zu sterben. Es konnte gezeigt werden, dass mesenchymale Stammzellen (MSC) zu primären Tumoren wandern und die Metastasierung von Brustkrebs fördern. Es wird vermutet, dass die Kommunikation zwischen diesen multipotenten Zellen und den Tumorzellen die Genexpression der Tumorzellen verändert und dadurch deren Bösartigkeit fördert. Neuste Forschungsergebnisse belegen, dass sogenannte „long non-coding RNAs“ (lncRNAs) eine wichtige Rolle in der Krebsentwicklung spielen.

Die Arbeit ist unterteilt in zwei zusammenhängende Projekte, welche beide die Rolle von lncRNAs in Brustkrebszellen nach deren Interaktion mit mesenchymalen Stammzellen beschreiben.

Das erste Projekt beschreibt die Identität und die Rolle von neuen lncRNAs in MSC aktivierten, metastasierenden Brustkrebszellen, welche durch Microarray-screening entdeckt wurden. Die neuen lncRNAs wurden in ein lentivirales Konstrukt geklont um deren Funktion durch Überexpression in Brustkrebszellen zu testen. Dadurch wurden zwei neue lncRNAs Transkripte identifiziert. Eines von diesen spielt möglicherweise eine Rolle in der Ausbreitung von Brustkrebs Stammzellen.

Das zweite Projekt beschreibt den Mechanismus der zu Expression von einer speziellen lncRNA, genannt HOTAIR, in Brustkrebszellen führt. HOTAIR Überexpression ist assoziiert mit Metastasierung und einer schlechten Prognose in vielen malignen Krankheiten. Es konnte gezeigt werden, dass MSC zu einer Hochregulation von HOTAIR in bestimmten Brustkrebszellen führen und das die Überexpression von HOTAIR zur erhöhten Expression von mesenchymalen Markern in Brustkrebszellen führt.

7 Summary

Breast cancer is the most prevalent cancer disease in the world and the most likely cause for a woman to die from cancer. It was shown that mesenchymal stem cells (MSC) migrate to the primary tumor and promote breast cancer cell metastasis. The crosstalk between these mulitpotent progenitor cells and tumor cells alters the gene expression of the tumor cells and enhances their malignant phenotype. Recent evidence suggests that long non-coding RNAs (lncRNAs) play a critical role in cancer progression.

This thesis is divided into two inter-related projects, both describing the involvement of long non-coding RNA induction in breast cancer cells following their interactions with mesenchymal stem cells.

The first project describes the identity and role of novel and previously un-described long non-coding RNAs in MSC-primed and highly metastatic breast cancer cells. Deregulated lncRNAs in MSC-activated cells that were previously found by microarray screening were investigated in terms of their regulation and their mode-of-action. To describe their potential function in breast cancer progression, the novel sequences were cloned into a lentiviral shuttle construct and activities and phenotypes of breast cancer cells overexpressing such constructs were probed using a variety of *in vitro* assays. As a result, novel lncRNA transcripts were identified that are highly induced in metastatic breast cancer cells upon MSC-activation. One of them might be associated with expansion of the cancer stem cell population.

The second project aimed to describe the mechanisms that lead to the expression of the lncRNA HOTAIR in breast cancer cells and to investigate if it plays a role in mesenchymal-stem-cell-catalyzed breast cancer progression. HOTAIR is known to promote cancer metastasis and is associated with poor prognosis in a variety of malignant diseases. It could be shown that MSC up-regulate HOTAIR expression in certain breast cancer cells and may lead to increased expression levels of mesenchymal markers.

8 Literature

(1996): Breast cancer and hormonal contraceptives: further results. Collaborative Group on Hormonal Factors in Breast Cancer. Contraception **54**, 1S-106S.

ACLOQUE, H., ADAMS, M. S., FISHWICK, K., BRONNER-FRASER, M. & NIETO, M. A. (2009): Epithelial-mesenchymal transitions: the importance of changing cell state in development and disease. J Clin Invest **119**, 1438-49.

AL-HAJJ, M., WICHA, M. S., BENITO-HERNANDEZ, A., MORRISON, S. J. & CLARKE, M. F. (2003): Prospective identification of tumorigenic breast cancer cells. Proc Natl Acad Sci U S A **100**, 3983-8.

ALT, E., YAN, Y., GEHMERT, S., SONG, Y. H., ALTMAN, A., GEHMERT, S., VYKOUKAL, D. & BAI, X. (2011): Fibroblasts share mesenchymal phenotypes with stem cells, but lack their differentiation and colony-forming potential. Biol Cell **103**, 197-208.

ARITA, T., ICHIKAWA, D., KONISHI, H., KOMATSU, S., SHIOZAKI, A., SHODA, K., KAWAGUCHI, T., HIRAJIMA, S., NAGATA, H., KUBOTA, T., FUJIWARA, H., OKAMOTO, K. & OTSUJI, E. (2013): Circulating long non-coding RNAs in plasma of patients with gastric cancer. Anticancer Res **33**, 3185-93.

BAKER, M. (2011): Long noncoding RNAs: the search for function. Nat Meth **8**, 379-383.

BARCELLOS-HOFF, M. H., LYDEN, D. & WANG, T. C. (2013): The evolution of the cancer niche during multistage carcinogenesis. Nat Rev Cancer **13**, 511-8.

BATLLE, E., SANCHO, E., FRANCI, C., DOMINGUEZ, D., MONFAR, M., BAULIDA, J. & GARCIA DE HERREROS, A. (2000): The transcription factor snail is a repressor of E-cadherin gene expression in epithelial tumour cells. Nat Cell Biol **2**, 84-9.

BERNARDO, M. E., LOCATELLI, F. & FIBBE, W. E. (2009): Mesenchymal stromal cells. Ann N Y Acad Sci **1176**, 101-17.

BERRY, D. A., CRONIN, K. A., PLEVRITIS, S. K., FRYBACK, D. G., CLARKE, L., ZELEN, M., MANDELBLATT, J. S., YAKOVLEV, A. Y., HABBEMA, J. D., FEUER, E. J., CANCER, I. & SURVEILLANCE MODELING NETWORK, C. (2005): Effect of screening and adjuvant therapy on mortality from breast cancer. N Engl J Med **353**, 1784-92.

BHAN, A., HUSSAIN, I., ANSARI, K. I., KASIRI, S., BASHYAL, A. & MANDAL, S. S. (2013): Antisense transcript long noncoding RNA (lncRNA) HOTAIR is transcriptionally induced by estradiol. J Mol Biol **425**, 3707-22.

BOYLE, P., LEVIN, B., INTERNATIONAL AGENCY FOR RESEARCH ON CANCER. & WORLD HEALTH ORGANIZATION. (2008): World cancer report 2008*,* Lyon
Geneva, International Agency for Research on Cancer ;
Distributed by WHO Press.

CABILI, M. N., TRAPNELL, C., GOFF, L., KOZIOL, M., TAZON-VEGA, B., REGEV, A. & RINN, J. L. (2011): Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev **25**, 1915-27.

CALIN, G. A., LIU, C. G., FERRACIN, M., HYSLOP, T., SPIZZO, R., SEVIGNANI, C., FABBRI, M., CIMMINO, A., LEE, E. J., WOJCIK, S. E., SHIMIZU, M., TILI, E., ROSSI, S., TACCIOLI, C., PICHIORRI, F., LIU, X., ZUPO, S., HERLEA, V., GRAMANTIERI, L., LANZA, G., ALDER, H., RASSENTI, L., VOLINIA, S., SCHMITTGEN, T. D., KIPPS, T. J., NEGRINI, M. & CROCE, C. M. (2007):

Ultraconserved regions encoding ncRNAs are altered in human leukemias and carcinomas. Cancer Cell **12**, 215-29.

CERAMI, E., GAO, J., DOGRUSOZ, U., GROSS, B. E., SUMER, S. O., AKSOY, B. A., JACOBSEN, A., BYRNE, C. J., HEUER, M. L., LARSSON, E., ANTIPIN, Y., REVA, B., GOLDBERG, A. P., SANDER, C. & SCHULTZ, N. (2012): The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov **2**, 401-4.

CHAO, Y. L., SHEPARD, C. R. & WELLS, A. (2010): Breast carcinoma cells re-express E-cadherin during mesenchymal to epithelial reverting transition. Mol Cancer **9**, 179.

COLLABORATIVE GROUP ON HORMONAL FACTORS IN BREAST, C. (2002): Breast cancer and breastfeeding: collaborative reanalysis of individual data from 47 epidemiological studies in 30 countries, including 50302 women with breast cancer and 96973 women without the disease. Lancet **360**, 187-95.

CRISCITIELLO, C., AZIM, H. A., JR., SCHOUTEN, P. C., LINN, S. C. & SOTIRIOU, C. (2012): Understanding the biology of triple-negative breast cancer. Ann Oncol **23 Suppl 6**, vi13-8.

CUIFFO, B. G. & KARNOUB, A. E. (2012): Mesenchymal stem cells in tumor development: emerging roles and concepts. Cell Adh Migr **6**, 220-30.

DE CRAENE, B. & BERX, G. (2013): Regulatory networks defining EMT during cancer initiation and progression. Nat Rev Cancer **13**, 97-110.

DUMONT, N., LIU, B., DEFILIPPIS, R. A., CHANG, H., RABBAN, J. T., KARNEZIS, A. N., TJOE, J. A., MARX, J., PARVIN, B. & TLSTY, T. D. (2013): Breast fibroblasts modulate early dissemination, tumorigenesis, and metastasis through alteration of extracellular matrix characteristics. Neoplasia **15**, 249-62.

EGEBLAD, M., NAKASONE, E. S. & WERB, Z. (2010): Tumors as organs: complex tissues that interface with the entire organism. Dev Cell **18**, 884-901.

EL-HAIBI, C. P., BELL, G. W., ZHANG, J., COLLMANN, A. Y., WOOD, D., SCHERBER, C. M., CSIZMADIA, E., MARIANI, O., ZHU, C., CAMPAGNE, A., TONER, M., BHATIA, S. N., IRIMIA, D., VINCENT-SALOMON, A. & KARNOUB, A. E. (2012): Critical role for lysyl oxidase in mesenchymal stem cell-driven breast cancer malignancy. Proc Natl Acad Sci U S A **109**, 17460-5.

EL-HAIBI, C. P. & KARNOUB, A. E. (2010): Mesenchymal stem cells in the pathogenesis and therapy of breast cancer. J Mammary Gland Biol Neoplasia **15**, 399-409.

ELSTON, C. W. & ELLIS, I. O. (1991): Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. Histopathology **19**, 403-10.

FERLAY, J., SHIN, H. R., BRAY, F., FORMAN, D., MATHERS, C. & PARKIN, D. M. (2010): Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. Int J Cancer **127**, 2893-917.

FIDLER, I. J. (2003): The pathogenesis of cancer metastasis: the 'seed and soil' hypothesis revisited. Nat Rev Cancer **3**, 453-8.

FRANK, T. S., MANLEY, S. A., OLOPADE, O. I., CUMMINGS, S., GARBER, J. E., BERNHARDT, B., ANTMAN, K., RUSSO, D., WOOD, M. E., MULLINEAU, L., ISAACS, C., PESHKIN, B., BUYS, S., VENNE, V., ROWLEY, P. T., LOADER, S., OFFIT, K., ROBSON, M., HAMPEL, H., BRENER, D., WINER, E. P., CLARK, S., WEBER, B., STRONG, L. C., THOMAS, A. & ET AL. (1998): Sequence analysis of BRCA1 and BRCA2: correlation of mutations with family history and ovarian cancer risk. J Clin Oncol **16**, 2417-25.

GAO, J., AKSOY, B. A., DOGRUSOZ, U., DRESDNER, G., GROSS, B., SUMER, S. O., SUN, Y., JACOBSEN, A., SINHA, R., LARSSON, E., CERAMI, E., SANDER, C. & SCHULTZ, N. (2013a): Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal **6**, pl1.

GAO, M. Q., KIM, B. G., KANG, S., CHOI, Y. P., YOON, J. H. & CHO, N. H. (2013b): Human breast cancer-associated fibroblasts enhance cancer cell proliferation through increased TGF-alpha cleavage by ADAM17. Cancer Lett **336**, 240-6.

GENG, Y. J., XIE, S. L., LI, Q., MA, J. & WANG, G. Y. (2011): Large intervening non-coding RNA HOTAIR is associated with hepatocellular carcinoma progression. J Int Med Res **39**, 2119-28.

GRIVENNIKOV, S. I., GRETEN, F. R. & KARIN, M. (2010): Immunity, inflammation, and cancer. Cell **140**, 883-99.

GUPTA, G. P. & MASSAGUE, J. (2006): Cancer metastasis: building a framework. Cell **127**, 679-95.

GUPTA, R. A., SHAH, N., WANG, K. C., KIM, J., HORLINGS, H. M., WONG, D. J., TSAI, M. C., HUNG, T., ARGANI, P., RINN, J. L., WANG, Y., BRZOSKA, P., KONG, B., LI, R., WEST, R. B., VAN DE VIJVER, M. J., SUKUMAR, S. & CHANG, H. Y. (2010): Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. Nature **464**, 1071-6.

GUTSCHNER, T. & DIEDERICHS, S. (2012): The hallmarks of cancer: a long non-coding RNA point of view. RNA Biol **9**, 703-19.

GUTTMAN, M., DONAGHEY, J., CAREY, B. W., GARBER, M., GRENIER, J. K., MUNSON, G., YOUNG, G., LUCAS, A. B., ACH, R., BRUHN, L., YANG, X., AMIT, I., MEISSNER, A., REGEV, A., RINN, J. L., ROOT, D. E. & LANDER, E. S. (2011): lincRNAs act in the circuitry controlling pluripotency and differentiation. Nature **477**, 295-300.

HALPERN, J. L., KILBARGER, A. & LYNCH, C. C. (2011): Mesenchymal stem cells promote mammary cancer cell migration in vitro via the CXCR2 receptor. Cancer Lett **308**, 91-9.

HUARTE, M., GUTTMAN, M., FELDSER, D., GARBER, M., KOZIOL, M. J., KENZELMANN-BROZ, D., KHALIL, A. M., ZUK, O., AMIT, I., RABANI, M., ATTARDI, L. D., REGEV, A., LANDER, E. S., JACKS, T. & RINN, J. L. (2010): A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. Cell **142**, 409-19.

HUARTE, M. & RINN, J. L. (2010): Large non-coding RNAs: missing links in cancer? Hum Mol Genet **19**, R152-61.

JEMAL, A., BRAY, F., CENTER, M. M., FERLAY, J., WARD, E. & FORMAN, D. (2011): Global cancer statistics. CA Cancer J Clin **61**, 69-90.

KARNOUB, A. E., DASH, A. B., VO, A. P., SULLIVAN, A., BROOKS, M. W., BELL, G. W., RICHARDSON, A. L., POLYAK, K., TUBO, R. & WEINBERG, R. A. (2007): Mesenchymal stem cells within tumour stroma promote breast cancer metastasis. Nature **449**, 557-63.

KENT, W. J. (2002): BLAT--the BLAST-like alignment tool. Genome Res **12**, 656-64.

KHALIL, A. M., GUTTMAN, M., HUARTE, M., GARBER, M., RAJ, A., RIVEA MORALES, D., THOMAS, K., PRESSER, A., BERNSTEIN, B. E., VAN OUDENAARDEN, A., REGEV, A., LANDER, E. S. & RINN, J. L. (2009): Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. Proc Natl Acad Sci U S A **106**, 11667-72.

KIM, A., BAE, Y. K., GU, M. J., KIM, J. Y., JANG, K. Y., BAE, H. I., LEE, H. J. & HONG, S. M. (2013a): Epithelial-mesenchymal transition phenotype is

associated with patient survival in small intestinal adenocarcinoma. Pathology **45**, 567-73.
KIM, K., JUTOORU, I., CHADALAPAKA, G., JOHNSON, G., FRANK, J., BURGHARDT, R., KIM, S. & SAFE, S. (2013b): HOTAIR is a negative prognostic factor and exhibits pro-oncogenic activity in pancreatic cancer. Oncogene **32**, 1616-25.
LA VECCHIA, C. (2004): Estrogen and combined estrogen-progestogen therapy in the menopause and breast cancer. Breast **13**, 515-8.
LEE, J. M., DEDHAR, S., KALLURI, R. & THOMPSON, E. W. (2006): The epithelial-mesenchymal transition: new insights in signaling, development, and disease. J Cell Biol **172**, 973-81.
LI, J. P., LIU, L. H., LI, J., CHEN, Y., JIANG, X. W., OUYANG, Y. R., LIU, Y. Q., ZHONG, H., LI, H. & XIAO, T. (2013a): Microarray expression profile of long noncoding RNAs in human osteosarcoma. Biochem Biophys Res Commun **433**, 200-6.
LI, X., LEWIS, M. T., HUANG, J., GUTIERREZ, C., OSBORNE, C. K., WU, M. F., HILSENBECK, S. G., PAVLICK, A., ZHANG, X., CHAMNESS, G. C., WONG, H., ROSEN, J. & CHANG, J. C. (2008): Intrinsic resistance of tumorigenic breast cancer cells to chemotherapy. J Natl Cancer Inst **100**, 672-9.
LI, X., WU, Z., MEI, Q., LI, X., GUO, M., FU, X. & HAN, W. (2013b): Long non-coding RNA HOTAIR, a driver of malignancy, predicts negative prognosis and exhibits oncogenic activity in oesophageal squamous cell carcinoma. Br J Cancer **109**, 2266-78.
LIU, X. H., LIU, Z. L., SUN, M., LIU, J., WANG, Z. X. & DE, W. (2013): The long non-coding RNA HOTAIR indicates a poor prognosis and promotes metastasis in non-small cell lung cancer. BMC Cancer **13**, 464.
LU, K. H., LI, W., LIU, X. H., SUN, M., ZHANG, M. L., WU, W. Q., XIE, W. P. & HOU, Y. Y. (2013): Long non-coding RNA MEG3 inhibits NSCLC cells proliferation and induces apoptosis by affecting p53 expression. BMC Cancer **13**, 461.
LUGA, V. & WRANA, J. L. (2013): Tumor-stroma interaction: revealing fibroblast-secreted exosomes as potent regulators of wnt-planar cell polarity signaling in cancer metastasis. Cancer Res **73**, 6843-7.
LUX, M. P., FASCHING, P. A. & BECKMANN, M. W. (2006): Hereditary breast and ovarian cancer: review and future perspectives. J Mol Med (Berl) **84**, 16-28.
LYONS, T. R., SCHEDIN, P. J. & BORGES, V. F. (2009): Pregnancy and breast cancer: when they collide. J Mammary Gland Biol Neoplasia **14**, 87-98.
MA, L., TERUYA-FELDSTEIN, J. & WEINBERG, R. A. (2007): Tumour invasion and metastasis initiated by microRNA-10b in breast cancer. Nature **449**, 682-8.
MA, L., YOUNG, J., PRABHALA, H., PAN, E., MESTDAGH, P., MUTH, D., TERUYA-FELDSTEIN, J., REINHARDT, F., ONDER, T. T., VALASTYAN, S., WESTERMANN, F., SPELEMAN, F., VANDESOMPELE, J. & WEINBERG, R. A. (2010): miR-9, a MYC/MYCN-activated microRNA, regulates E-cadherin and cancer metastasis. Nat Cell Biol **12**, 247-56.
MANDEL, K., YANG, Y., SCHAMBACH, A., GLAGE, S., OTTE, A. & HASS, R. (2013): Mesenchymal stem cells directly interact with breast cancer cells and promote tumor cell growth in vitro and in vivo. Stem Cells Dev **22**, 3114-27.
MATHEWS, D. H., MOSS, W. N. & TURNER, D. H. (2010): Folding and finding RNA secondary structure. Cold Spring Harb Perspect Biol **2**, a003665.
MEDEMA, J. P. (2013): Cancer stem cells: the challenges ahead. Nat Cell Biol **15**, 338-44.

MISHRA, P. J., MISHRA, P. J., HUMENIUK, R., MEDINA, D. J., ALEXE, G., MESIROV, J. P., GANESAN, S., GLOD, J. W. & BANERJEE, D. (2008): Carcinoma-associated fibroblast-like differentiation of human mesenchymal stem cells. Cancer Res **68**, 4331-9.

MOREB, J. S. (2008): Aldehyde dehydrogenase as a marker for stem cells. Curr Stem Cell Res Ther **3**, 237-46.

NAKAGAWA, T., ENDO, H., YOKOYAMA, M., ABE, J., TAMAI, K., TANAKA, N., SATO, I., TAKAHASHI, S., KONDO, T. & SATOH, K. (2013): Large noncoding RNA HOTAIR enhances aggressive biological behavior and is associated with short disease-free survival in human non-small cell lung cancer. Biochem Biophys Res Commun **436**, 319-24.

NAROD, S. A. (2012): Breast cancer in young women. Nat Rev Clin Oncol **9**, 460-70.

O'SHAUGHNESSY, J. (2005): Extending survival with chemotherapy in metastatic breast cancer. Oncologist **10 Suppl 3**, 20-9.

PARKIN, D. M., BRAY, F., FERLAY, J. & PISANI, P. (2005): Global cancer statistics, 2002. CA Cancer J Clin **55**, 74-108.

PEINADO, H., OLMEDA, D. & CANO, A. (2007): Snail, Zeb and bHLH factors in tumour progression: an alliance against the epithelial phenotype? Nat Rev Cancer **7**, 415-28.

PEROU, C. M., SORLIE, T., EISEN, M. B., VAN DE RIJN, M., JEFFREY, S. S., REES, C. A., POLLACK, J. R., ROSS, D. T., JOHNSEN, H., AKSLEN, L. A., FLUGE, O., PERGAMENSCHIKOV, A., WILLIAMS, C., ZHU, S. X., LONNING, P. E., BORRESEN-DALE, A. L., BROWN, P. O. & BOTSTEIN, D. (2000): Molecular portraits of human breast tumours. Nature **406**, 747-52.

POLYAK, K. & WEINBERG, R. A. (2009): Transitions between epithelial and mesenchymal states: acquisition of malignant and stem cell traits. Nat Rev Cancer **9**, 265-73.

QUAIL, D. F. & JOYCE, J. A. (2013): Microenvironmental regulation of tumor progression and metastasis. Nat Med **19**, 1423-37.

REDIG, A. J. & MCALLISTER, S. S. (2013): Breast cancer as a systemic disease: a view of metastasis. J Intern Med **274**, 113-26.

REN, S., LIU, Y., XU, W., SUN, Y., LU, J., WANG, F., WEI, M., SHEN, J., HOU, J., GAO, X., XU, C., HUANG, J., ZHAO, Y. & SUN, Y. (2013): Long Noncoding RNA MALAT-1 is a New Potential Therapeutic Target for Castration Resistant Prostate Cancer. J Urol **190**, 2278-87.

REYA, T., MORRISON, S. J., CLARKE, M. F. & WEISSMAN, I. L. (2001): Stem cells, cancer, and cancer stem cells. Nature **414**, 105-11.

RHODES, L. V., ANTOON, J. W., MUIR, S. E., ELLIOTT, S., BECKMAN, B. S. & BUROW, M. E. (2010): Effects of human mesenchymal stem cells on ER-positive human breast carcinoma cells mediated through ER-SDF-1/CXCR4 crosstalk. Mol Cancer **9**, 295.

RICCI-VITIANI, L., FABRIZI, E., PALIO, E. & DE MARIA, R. (2009): Colon cancer stem cells. J Mol Med (Berl) **87**, 1097-104.

SANCAK, Y., PETERSON, T. R., SHAUL, Y. D., LINDQUIST, R. A., THOREEN, C. C., BAR-PELED, L. & SABATINI, D. M. (2008): The Rag GTPases bind raptor and mediate amino acid signaling to mTORC1. Science **320**, 1496-501.

SCHMIDT, L. H., SPIEKER, T., KOSCHMIEDER, S., SCHAFFERS, S., HUMBERG, J., JUNGEN, D., BULK, E., HASCHER, A., WITTMER, D., MARRA, A., HILLEJAN, L., WIEBE, K., BERDEL, W. E., WIEWRODT, R. & MULLER-TIDOW, C. (2011): The long noncoding MALAT-1 RNA indicates a poor

prognosis in non-small cell lung cancer and induces migration and tumor growth. J Thorac Oncol **6**, 1984-92.
SCHNITT, S. J. (2010): Classification and prognosis of invasive breast cancer: from morphology to molecular taxonomy. Mod Pathol **23 Suppl 2**, S60-4.
SPAETH, E., KLOPP, A., DEMBINSKI, J., ANDREEFF, M. & MARINI, F. (2008): Inflammation and tumor microenvironments: defining the migratory itinerary of mesenchymal stem cells. Gene Ther **15**, 730-8.
STROUD, J. C., WU, Y., BATES, D. L., HAN, A., NOWICK, K., PAABO, S., TONG, H. & CHEN, L. (2006): Structure of the forkhead domain of FOXP2 bound to DNA. Structure **14**, 159-66.
STUMM, L., BURKHARDT, L., STEURER, S., SIMON, R., ADAM, M., BECKER, A., SAUTER, G., MINNER, S., SCHLOMM, T., SIRMA, H. & MICHL, U. (2013): Strong expression of the neuronal transcription factor FOXP2 is linked to an increased risk of early PSA recurrence in ERG fusion-negative cancers. J Clin Pathol **66**, 563-8.
TAM, W. L. & WEINBERG, R. A. (2013): The epigenetics of epithelial-mesenchymal plasticity in cancer. Nat Med **19**, 1438-49.
TAYLOR, M. A., PARVANI, J. G. & SCHIEMANN, W. P. (2010): The pathophysiology of epithelial-mesenchymal transition induced by transforming growth factor-beta in normal and malignant mammary epithelial cells. J Mammary Gland Biol Neoplasia **15**, 169-90.
THIERY, J. P., ACLOQUE, H., HUANG, R. Y. & NIETO, M. A. (2009): Epithelial-mesenchymal transitions in development and disease. Cell **139**, 871-90.
TRAPNELL, C., WILLIAMS, B. A., PERTEA, G., MORTAZAVI, A., KWAN, G., VAN BAREN, M. J., SALZBERG, S. L., WOLD, B. J. & PACHTER, L. (2010): Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol **28**, 511-5.
TSAI, M. C., MANOR, O., WAN, Y., MOSAMMAPARAST, N., WANG, J. K., LAN, F., SHI, Y., SEGAL, E. & CHANG, H. Y. (2010): Long noncoding RNA as modular scaffold of histone modification complexes. Science **329**, 689-93.
TSAI, M. C., SPITALE, R. C. & CHANG, H. Y. (2011): Long intergenic noncoding RNAs: new links in cancer progression. Cancer Res **71**, 3-7.
TURNER, N. C. & REIS-FILHO, J. S. (2013): Tackling the diversity of triple-negative breast cancer. Clin Cancer Res **19**, 6380-8.
ULITSKY, I. & BARTEL, D. P. (2013): lincRNAs: genomics, evolution, and mechanisms. Cell **154**, 26-46.
UNTERGASSER, A., NIJVEEN, H., RAO, X., BISSELING, T., GEURTS, R. & LEUNISSEN, J. A. (2007): Primer3Plus, an enhanced web interface to Primer3. Nucleic Acids Res **35**, W71-4.
VALASTYAN, S., REINHARDT, F., BENAICH, N., CALOGRIAS, D., SZASZ, A. M., WANG, Z. C., BROCK, J. E., RICHARDSON, A. L. & WEINBERG, R. A. (2009): A pleiotropically acting microRNA, miR-31, inhibits breast cancer metastasis. Cell **137**, 1032-46.
VAN DE VIJVER, M. J., HE, Y. D., VAN'T VEER, L. J., DAI, H., HART, A. A., VOSKUIL, D. W., SCHREIBER, G. J., PETERSE, J. L., ROBERTS, C., MARTON, M. J., PARRISH, M., ATSMA, D., WITTEVEEN, A., GLAS, A., DELAHAYE, L., VAN DER VELDE, T., BARTELINK, H., RODENHUIS, S., RUTGERS, E. T., FRIEND, S. H. & BERNARDS, R. (2002): A gene-expression signature as a predictor of survival in breast cancer. N Engl J Med **347**, 1999-2009.

VAN ZIJL, F., MAIR, M., CSISZAR, A., SCHNELLER, D., ZULEHNER, G., HUBER, H., EFERL, R., BEUG, H., DOLZNIG, H. & MIKULITS, W. (2009): Hepatic tumor-stroma crosstalk guides epithelial to mesenchymal transition at the tumor edge. Oncogene **28**, 4022-33.

VERMEULEN, L., DE SOUSA E MELO, F., RICHEL, D. J. & MEDEMA, J. P. (2012): The developing cancer stem-cell model: clinical challenges and opportunities. Lancet Oncol **13**, e83-9.

VICARIO, C. M. (2013): FOXP2 gene and language development: the molecular substrate of the gestural-origin theory of speech? Front Behav Neurosci **7**, 99.

VOLDERS, P. J., HELSENS, K., WANG, X., MENTEN, B., MARTENS, L., GEVAERT, K., VANDESOMPELE, J. & MESTDAGH, P. (2013): LNCipedia: a database for annotated human lncRNA transcript sequences and structures. Nucleic Acids Res **41**, D246-51.

WAHLESTEDT, C. (2013): Targeting long non-coding RNA to therapeutically upregulate gene expression. Nat Rev Drug Discov **12**, 433-46.

WANG, X., SONG, X., GLASS, C. K. & ROSENFELD, M. G. (2011): The long arm of long noncoding RNAs: roles as sensors regulating gene transcriptional programs. Cold Spring Harb Perspect Biol **3**, a003756.

WEIGELT, B. & REIS-FILHO, J. S. (2009): Histological and molecular types of breast cancer: is there a unifying taxonomy? Nat Rev Clin Oncol **6**, 718-30.

XU, H., EIREW, P., MULLALY, S. C. & APARICIO, S. (2013a): The Omics of Triple-Negative Breast Cancers. Clin Chem.

XU, Z. Y., YU, Q. M., DU, Y. A., YANG, L. T., DONG, R. Z., HUANG, L., YU, P. F. & CHENG, X. D. (2013b): Knockdown of long non-coding RNA HOTAIR suppresses tumor invasion and reverses epithelial-mesenchymal transition in gastric cancer. Int J Biol Sci **9**, 587-97.

YU, J. M., JUN, E. S., BAE, Y. C. & JUNG, J. S. (2008): Mesenchymal stem cells derived from human adipose tissues favor tumor cell growth in vivo. Stem Cells Dev **17**, 463-73.

ZEISBERG, E. M., POTENTA, S., XIE, L., ZEISBERG, M. & KALLURI, R. (2007): Discovery of endothelial to mesenchymal transition as a source for carcinoma-associated fibroblasts. Cancer Res **67**, 10123-8.

ZHANG, X., SUN, S., PU, J. K., TSANG, A. C., LEE, D., MAN, V. O., LUI, W. M., WONG, S. T. & LEUNG, G. K. (2012): Long non-coding RNA expression profiles predict clinical phenotypes in glioma. Neurobiol Dis **48**, 1-8.

ZHAO, Y., GUO, Q., CHEN, J., HU, J., WANG, S. & SUN, Y. (2014): Role of long non-coding RNA HULC in cell proliferation, apoptosis and tumor metastasis of gastric cancer: A clinical and in vitro investigation. Oncol Rep **31**, 358-64.

ZHUANG, Y., WANG, X., NGUYEN, H. T., ZHUO, Y., CUI, X., FEWELL, C., FLEMINGTON, E. K. & SHAN, B. (2013): Induction of long intergenic non-coding RNA HOTAIR in lung cancer cells by type I collagen. J Hematol Oncol **6**, 35.

9 Acknowledgements

First, I want to thank Antoine Karnoub, for giving me the extraordinary opportunity to come to Boston and to work at his lab at the Beth Israel Deaconess Medical Center and the Harvard Medical School. This stay was really a great experience for me. I strongly believe that it shaped my way of scientific thinking and made me more independent as a researcher. I am very grateful that I had this opportunity and I enjoyed being a part of the Karnoub lab very much.

In this context, I also want to thank the Karnoub lab members, Benjamin Cuiffo for helpful scientific discussions and especially Clémence Thomas for becoming a very good friend, her constructive comments, warm encouragement and strong support throughout many challenges.

In addition, I want to thank Marina Karaghiosoff, Sabine Macho-Maschler and Helmut Dolznig for their scientific support throughout my studies abroad.

Also, I want to express my gratitude to the Austrian Marshall Plan Foundation and the government of Lower Austria for their financial support of my stay.

Finally, I want to thank my family and friends for their strong support throughout my entire life and especially for the time I was in Boston. I would not be where I am today without them.

10 Appendix

10.1 Microarray Probe Set Analysis

TCONS_00005559_1 / lnc-ST3GAL6-2

Affymetrix probe set	Oligonucleotide sequences	Specific recognition
242005_at	AGCATAATCATTTGTCCGAGGTCAC	yes
	ACAGCTGGAATGTAAGCCTCAGTAC	yes
	CAGTACTTTACTGAATCCTTGGCTT	yes
	ACAGGCATAAAGCTTTTCTCTTCCC	yes
	GTGACTTCGAGTACCAGCTTTCAAA	yes
	CTTTCTCTGCATTGTGGAAGGGTCT	yes
	ATGTATCTTTTATCAATGGCCTTGC	yes
	GAATCCAGATTTAGAGGTCTTCTCC	yes
	AGAGGTCTTCTCCAAACATTTGCAT	yes
	AGAACTTCTACCTATTGTCATTTTA	**no**
	ATTTATTCATTATTTACACCCTTAT	yes
Summary		**10/11 specific**

TCONS_00004205_1 / lnc-AC007401.2.1-1

Affymetrix probe set	Oligonucleotide sequences	Specific recognition
236656_s_at	GGTGCTGTGTTACGGGAGAGAGTGA	yes
	GCTGAATCTTTCTCCCTGGAGTAAG	yes
	TTCTCCCTGGAGTAAGGCCGAAGAC	yes
	GGCCGAAGACTGGATTACTACACGC	yes
	ACTGGATTACTACACGCCTAGACGT	yes
	CACGCCTAGACGTGACACTACACCC	yes
	GACACTACACCCATAGATCTCATGC	yes
	GATCTCATGCATCATTAATGCCATA	yes
	ATGCCATATGACATTGCCATTTTCT	**no**
	TGACATTGCCATTTTCTTTCTCAGT	yes
	TTTCTTTCTCAGTTCACGGACAAAA	yes
236657_at	CCTCCGCACACTGGATGAGAATCCA	**no**
	GCACACTGGATGAGAATCCATCTTC	**no**
	AATCCATCTTCCATTCGAGCTGGGA	yes
	TCCATCTTCCATTCGAGCTGGGAAT	yes
	TTCCATTCGAGCTGGGAATAGACTT	yes
	GATATTATGTAATGGAGTCTCGGGA	yes
	GTAATGGAGTCTCGGGAACCCTGAG	yes
	TGGAGTCTCGGGAACCCTGAGACCT	yes
	TGAGACCTCTCCAGCGAAGCTGAAG	yes
	GACCTCTCCAGCGAAGCTGAAGTGA	yes
	AAACGGTCTTGGTGCTGTGTTACGG	**no**
Summary		**18/22 specific**

TCONS_00026813_1 / lnc-UQCRFS1-9

Affymetrix probe set	Oligonucleotide sequences	Specific recognition
235191_at	GGAGGAGCAATTTTACTGCGACATT	yes
	ATTGATTTCCAAGGGTGTCCATCGC	**no**
	TCACTGCCTTGATCATGTCGGGGTC	yes
	GACCCTGACTCAACGACGGATGAAT	yes
	AGATCTTATGCTTGTCAGTCCTGCT	**no**
	CTCCAAGAAGAGTCCTGTCAGCCAC	yes
	TCAGCCACGGCTCTGACTTGTTGGC	yes
	CCCTCCCGGCATTTATTTAGCACAC	**no**
	GAAGTCTCAGGCTGGAAACACTCTT	**no**
	GGACATTTCTAAGCACATTGAGGCC	yes
	GTGGTGTATGGATTCATCTCACTGA	**no**
Summary	**EXCLUDED**	**6/11 specific**

TCONS_00002647_1 / lnc-AC016722.1.1-1

Affymetrix probe set	Oligonucleotide sequences	Specific recognition
230799_at	GACATCACACCTTATCAGATCCTAC	yes
	GATCCTACGAGTGAGCTACTGACCT	yes
	GACCTCAACCACTGAACTTCAAACT	yes
	CTGTGGCAAGCTACACTCTGTCAGC	yes
	ACCCTGCTAGCCTTTGAACTTGGAA	yes
	TGCCACCCTCTAGCCTTGAGGGACC	yes
	TATATGTCTCTGTGAATCTCCGTTG	yes
	ATCTCCGTTGCATTGGACAGCTTCT	yes
	GCCTGCTTCCATCCTAATTGAGCAA	**no**
	AGCGCATTCCCTGTGAGATTTGTTC	**no**
	CCCCTCCTTAAGCATTGTGAACCTA	**no**
Summary		**8/11 specific**

TCONS_00014854_1 / lnc-MYC-2

Affymetrix probe set	Oligonucleotide sequences	Specific recognition
1558290_a_at	TTTTGCATGTCTGACACCCATGACT	**no**
	ACCCAGAAGCAATTCAGCCCAACAG	**no**
	ATTCAGCCCAACAGGAGGACAGCTT	yes
	GAGGACAGCTTCAACCCATTACGAT	yes
	TCAACCCATTACGATTTCATCTCTG	yes
	ATTACGATTTCATCTCTGCCCCAAC	**no**
	AGCAGCAAGCACCTGTTACCTGTCC	**no**
	GCCTTTGAAAAATCCCTAACCTATG	**no**
	AAATCCCTAACCTATGAGCTTTGAA	**no**
	CCCTAACCTATGAGCTTTGAATAAG	**no**
	AGATGAGTACGAACTTCATCGCCCA	yes
Summary	**EXCLUDED**	**4/11 specific**

TCONS_00013598_1 / lnc-MKLN1-1

Affymetrix probe set	Oligonucleotide sequences	Specific recognition
239814_at	TACACATCCGTCCATTCGTTTAATC	yes
	CTTGTCCGGGCTACTGATGCGAAGA	yes
	GAAGAGAGGCCATGGTTTCCTTTTT	yes
	CTGAATGGCAGCCTTGACGTCACAA	yes
	GACGTCACAACAGCTAGGAACCACT	yes
	GCCCTGCTCTGTCGCAGGGGCAGGG	yes
	ATCCAGTTCTGATGTCTTCTGTGGC	yes
	TTAGGCGTAAAGGCGGTCTCTCTCT	yes
	GTCTCTCTCTAGGCTGATCAAACAG	yes
	TTTGCCCAGTGCTAATTGCCTTCAC	yes
	TTGCCTTCACGGAACCTGGATAAAC	yes
Summary		**11/11 specific**

TCONS_00017736_1 / lnc-MBL2-4

Affymetrix probe set	Oligonucleotide sequences	Specific recognition
1557779_at	GATAACCCCAATCTACGAAGACTAG	yes
	GGAACTTCCTACACTGAGACAACTC	yes
	GCTTCTTCAGTAGCCCTCGAAATAT	yes
	GAAATGTCAACTTCTGGCCACAACG	yes
	GGCCACAACGGCAACCAGTAAAATG	yes
	ATAGTATTTGCAGGTGGAACGCCTA	yes
	GGAACGCCTAGCAGGGCTTGAGTCT	yes
	AACTCTGCTGCTTTTACTCTAATTG	yes
	GTAACATAACAATTCACAGCCTTTT	**no**
	ACAGCCTTTTACTTTGTAGTTATCG	yes
	GTTATCGTGAAGATCTAATCGCAGT	yes
Summary		**10/11 specific**

TCONS_00000659_2 / lnc-KCNJ9-2

Affymetrix probe set	Oligonucleotide sequences	Specific recognition
238901_at	GTAGTCCTAACTATTCAGATTCCTT	**no**
	CCTAACTATTCAGATTCCTTGAGCC	yes
	GATGCTAAGATGACATCCAGACAGG	yes
	GATGACATCCAGACAGGTCAGAAAT	yes
	GATGAGGAAGGATAAACCCTATCTT	yes
	GGATAAACCCTATCTTATAAGACAA	yes
	AAGACAAATTCCTGGCACTCATTTC	**no**
	ACAAATTCCTGGCACTCATTTCTTT	**no**
	TTCCCTAGTCCTGCATTTTTGTTGA	yes
	TAGTCCTGCATTTTTGTTGAAGGCA	yes
	AAGGCAGAAGATTAGAGAGAGACAA	**no**
239370_at	AACAACATGACCGGGAAGATTTCCT	yes
	CTGGCTCTACCTTAAGTCTTTAATA	**no**
	GACTGAAGGTACCAAGGTGTGCTGA	yes
	GAAGCAAAGTTCTCCAAAGTCCAGC	yes
	AGTCCAGCATGGTAGACATCAGTGG	yes
	CAAGGACAGACCCCAAGGCAAGGTG	yes
	GACCCCAAGGCAAGGTGAACCTCAA	yes
	GAACCTCAAGTCTATGCAGTCCAGC	yes
	TCAAGTCTATGCAGTCCAGCTGCCC	yes
	AATATAGACAGAGTAGTCCCTGGCT	yes
	AAGGATGGATTCTCCCATTCCATAC	yes
Summary		**17/22 specific**

TCONS_00019082_1 / lnc-SBF2-2

Affymetrix probe set	Oligonucleotide sequences	Specific recognition
1562606_a_at	GTAGCCCTGCTTGGAATTTAGTGTC	yes
	TTAGTGTCCACCACAGATGTATCCA	yes
	GATGTATCCACTACCCAGGGCCAAA	yes
	AAAGACTCGGCAGCCTCAGAAGACT	**no**
	TCAGGGCCCTCCTCTGTATGTAGGG	yes
	TGGCACATGGCCACTAGCTGGGTGA	yes
	GGGTGGTGTAATGAACAGCATCCTA	yes
	CAGCATCCTAGAAGTCCGGTACAGA	yes
	GGTACAGATTTCACGTGTCACCTTC	yes
	GTCACCTTCTAGTTCCTTATCTGAA	yes
	GGCATATCATTGAGGTCAGGTGTTT	**no**
Summary		**9/11 specific**

10.2 List of Primers

Gene Name	Forward 5' – 3'	Reverse 5' – 3'	Notes
18S	GTAACCCGTTGAACCCCATT	CCATCCAATCGGTAGTAGCG	Designed by Dr. Cuiffo (BIDMC)
A-SMA	ACTGGGACGACATGGAAAAG	GCGTCCAGAGGCATAGAGAG	Designed by Dr. El-Haibi (BIDMC)
CMV-FWD	CGCAAATGGGCGGTAGGCGTG		
COL1A2	GGCAGAGATGGTGAAGATGG	GGCCAAGTCCAACTCCTTTT	
E-Cadherin	AGAACGCATTGCCACATACACTC	CATTCTGATCGGTTACCGTGATC	Designed by Dr. El-Haibi (BIDMC)
ER-1	TGCCAAGGAGACTCGCTACT	GGCAGCTCTTCCTCCTGTTT	
ER-2	TGCTCCTGGCAACTACTTCA	ATCACCCAAACCAAAGCATC	
Fibronectin	CAGTGGGAGACCTCGAGAAG	TCCCTCGGAACATCAGAAAC	Designed by Dr. El-Haibi (BIDMC)
FOXP2	TGCGACAGAGACAATAAGCAA	AATCCACTTGTTTGCTGCTG	Designed by Dr. Cuiffo (BIDMC)
HOTAIR	CAGTGGAATGGAACGGATTT	CTTCCCTCCTCTGGCTCTCT	
HOTAIR-cloning	ATTAGTACCGGTGACTCGCCTGTGCTCTGGAGC	TGCTTAGAATTCTTTTTTTTTTTGAAAATGCATC	
K-1-FWD	CTGATGTAACAGCCTTGGGAAA		
K-1-REV		TTGACGAGACACATTTAATAACATACA	
K-2-FWD	GAGGTTGCAGTGAAAAGCTG		
K-2-REV		TTTTTAAAATGAAACAATGCCAAC	
K-3-FWD	GTCCTGCTGTGGTGGAGAGA		
K-4-FWD	GGAGGAAAGATAATAAAAGGCCAAA		
lnc-AC007401.2.1-1	ACGGGAGAGAGTGACTGGAA	CACCACTTTTGTCCGTGAACT	
lnc-AC016722.1.1-1 isoform 1+2+3+4	GAAGCATCGCCCACAAAC	ATGCGCTGGGTAGACAGG	
lnc-AC016722.1.1-1: isoform 1+2	GATGGACCCAGGATGTGAG	GAAAGGGAACACCTCATGGA	
lnc-AC016722.1.1-1: isoform 3	ATGAGATCAGCCACCCTGTC	GAAAGGGAACACCTCATGGA	
lnc-AC016722.1.1-1: isoform 4	TCTGCTACTCCCGTGCTTG	GAAAGGGAACACCTCATGGA	
lnc-KCNJ9 isoform 1+2+3	CCTTTGGATCTGGAAAGCTG	ACTACAGGTGCTGGGCTCTG	
lnc-KCNJ9 isoform 1	CAGTCTAGCCATGCAGGACA	CAGACCTCACCACCCAAAGT	
lnc-KCNJ9-2:2 (isoform 1+3)	GGCAAGGTGAACCTCAAAAA	TTCCTGCAAGAGGAGAAAGC	
lnc-KCNJ9-2:2-cloning	CGATTCACCGGTATAATAAAAGGCCAAACCTTTGC	TGATTGGAATTCTTGACGAGACACATTTAATAA	
lnc-MBL2-4	TGAGATCATGGAGGAAGTGAA	TTTCGAGGGCTACTGAAGAA	
lnc-MKLN1-1	CATATCACCCAACCTCGCTAA	TGCCATTCAGAGCAGAGAAA	
lnc-MYC-2	TTGGTGCTCTGTGTTCACCT	TGTCCACTAGCAGCAACAGG	
lnc-SBF2-2	TGTGGAAATGCAGAGAGCAC	GGCACCTGGTGTTTTGTCTT	
lnc-ST3GAL6-2	AAGGAGAAGAAGGGGAGGAAT	TGCAAATGTTTGGAGAAGACC	
lnc-UQCRFS1-9	AGCATCCAGGCCACTTACAG	GAATCCATACACCACAAAGCA	
LOX	Qiagen QuantiTect Primer Assay Hs_LOX_1_SG	Qiagen (Valencia, CA)	Cat. No QT00017311
N-Cadherin	ACAGTGGCCACCTACAAAGG	CCGAGATGGGGTTGATAATG	Designed by Dr. El-Haibi (BIDMC)
SLUG	GGGGAGAAGCCTTTTTCTTG	TCCTCATGTTTGTGCAGGAG	Designed by Dr. El-Haibi (BIDMC)
SNAIL	CCTCCCTGTCAGATGAGGAC	CCAGGCTGAGGTATTCCTT	

TWIST1	GGAGTCCGCAGTCTTACGAG	TCTGGAGGACCTGGTAGAGG	
VIMENTIN	GAGAACTTTGCCGTTGAAGC	GCTTCCTGTAGGTGGCAATC	

10.3 List of Reagents and Equipment

Product	**Company**	**Catalog Number**
0.45 micron sterile syringe filter	Corning Inc. (Corning, NY)	CLS431225
1.5 ml safe-lock tubes	Eppendorf (Westbury, NY)	0030 120.086
10X Phosphate-Buffered Saline (PBS), pH 7.4, Liquid	Mediatech, Inc. (Manassas, VA)	MT-46-013-CM
12-well plate	Beckon Dickinson Labware (Franklin Lakes, NJ)	353043
2-Propanol	Sigma-Aldrich (St. Louis, MO)	190764
24-well BD Falcon Companion TC Plate	Beckon Dickinson Labware (Franklin Lakes, NJ)	353504
3/20 ml Syringe Luer-Lok™Tip	Corning Inc. (Corning, NY)	309604
384-Well Skirted PCR Microplates	Axygen Scientifc (Union City, CA)	14-222-317
50X TAE Buffer	Thermo Fisher Scientific (Waltham, MA)	B49
6-well plate	Corning Inc. (Corning, NY)	#3335
6x DNA Loading Dye	Thermo Fisher Scientific (Waltham, MA)	#R0611
7900HT Fast Real-Time PCR System	Applied Biosystems (Foster City, CA)	4329001
Aerosol Barrier Filter Tips (Axygen) 0.5-10 µl	Axygen Scientifc (Union City, CA)	14-222-785
Aerosol Barrier Pipet Tips in Sterile Rack (ExactaCruz), 1-20 µl	Santa Cruz Biotechnology (Santa Cruz, CA)	sc-201723
Aerosol Barrier Pipet Tips in Sterile Rack (ExactaCruz), 1-200 µl	Santa Cruz Biotechnology (Santa Cruz, CA)	sc-201725
Aerosol Barrier Pipet Tips in Sterile Rack (ExactaCruz), 50-1250 µl	Santa Cruz Biotechnology (Santa Cruz, CA)	sc-357416
Affymetrix HT-HG U133 2.0 Plus Array	Affymetrix (Santa Clara, CA)	900751
Agarose, LE, 500g	KSE Scientific (Durham, NC)	BMK-A1705
AgeI - Restriction enzyme	NEB (Ipswich, MA)	R0552S
ALDEFLUOR™ Kit	Stemcell Technologies (Vancouver, Canada)	1700
BD FACSCanto II flow cytometer	BD Biosciences (San Jose, CA)	
BD Falcon 100 x 20 mm sterile tissue culture dish	Corning Inc. (Corning, NY)	08-772-E
BD Falcon 15 mL Conical Centrifuge Tubes	Corning Inc. (Corning, NY)	14-959-49D
BD Falcon 150 x 25 mm sterile tissue culture dish	Corning Inc. (Corning, NY)	353025
BD Falcon 5 mL Round-Bottom Polystyrene Tubes	Corning Inc. (Corning, NY)	14-959-1A
BD Falcon 50 mL Conical Centrifuge Tubes	Corning Inc. (Corning, NY)	14-432-22
BD Falcon 60 x 15 mm sterile tissue culture dish	Corning Inc. (Corning, NY)	08-772B
BD Falcon Cell Culture Inserts for 6-well plates	Corning Inc. (Corning, NY)	353090
BD Falcon serological pipettes 10 ml	Corning Inc. (Corning, NY)	13-675-20
BD Falcon serological pipettes 25 ml	Corning Inc. (Corning, NY)	13-668-2
BD Falcon serological pipettes 5 ml	Corning Inc. (Corning, NY)	13-675-22
Chloroform	Sigma-Aldrich (St. Louis, MO)	288306
CLC Main Workbench 6	CLC bio (Cambridge, MA)	

Collagen I, human, 10 mg	BD Biosciences (San Jose, CA)	08-774-550
Dulbecco's Modification of Eagle's (Mod.) 1X (DMEM)	Mediatech, Inc. (Manassas, VA)	MT-10-013-CV
EcoRI - Restriction enzyme	NEB (Ipswich, MA)	R0101S
Ethidium bromide solution	Sigma-Aldrich (St. Louis, MO)	E1510
Ethyl Alcohol	Pharmco-AAPER (Brookfield, CT)	111000200
Fetal Bovine Serum - Premium Select	Atlanta Biologicals (Flowery Branch, GA)	S11550
Fetal Bovine Serum (HyClone)	Thermo Fisher Scientific (Waltham, MA)	SH3007003
FirstChoice® RLM-RACE Kit	Life technologies (Carlsbad, CA)	AM1700
Forma Orbital Shaker	Thermo Fisher Scientific (Waltham, MA)	
FuGENE® HD Transfection Reagent	Promega (Madison, WI)	PR-E2312
GeneChip HT IVT Labeling Kit	Affymetrix (Santa Clara, CA)	900688
Genechip HT One-Cycle cDNA Synthesis Kit	Affymetrix (Santa Clara, CA)	900687
Ham's F-12 Medium, 1X, w/L-Glu	Mediatech, Inc. (Manassas, VA)	MT-10-080-CV
HEK 293T/17	ATCC (Manassas, VA)	CRL-11268
Hematocytometer Counting Chamber	Thermo Fisher Scientific (Waltham, MA)	267110
Human Adipose-derived Mesenchymal Stem Cells	ScienCell Research Laboratories (Carlsbad, CA)	7510
Human Bone marrow-derived Mesenchymal Stem Cells	Texas AM Health Science Center (Temple, TX)	
L-Glutamine 200 mM, 100X (Gibco®)	Life technologies (Carlsbad, CA)	25030-081
LabGard ES (Energy Saver) NU-425 Class II, Type A2 Biosafety Cabinet	Nuaire (Plymouth, MN)	
LB Broth Miller	Thermo Fisher Scientific (Waltham, MA)	BP1426-2
LZRS-HOTAIR	Addgene (Cambridge MA)	26110
MCF 10A	ATCC (Manassas, VA)	CRL-10317
MDA-MB-231	ATCC (Manassas, VA)	HTB-26
MDA-MB-468	ATCC (Manassas, VA)	HTB-132
MEM α (Gibco®)	Life technologies (Carlsbad, CA)	12561-072
Microcentrifuge Sorval Legend Micro 21	Thermo Fisher Scientific (Waltham, MA)	75002436
miRNeasy Mini Kit	Qiagen (Valencia, CA)	217004
NanoDrop1000 Spectrophotometer	Thermo Fisher Scientific (Waltham, MA)	
Nuclease-Free Water Ambion®	Life technologies (Carlsbad, CA)	AM9937
One Shot® Stbl3™ Chemically Competent E. coli	Life technologies (Carlsbad, CA)	C7373-03
Optigrow LB Agar Miller Powder	Thermo Fisher Scientific (Waltham, MA)	BP1425-10P1
PCR Tubes 0.2 ml	Thermo Fisher Scientific (Waltham, MA)	3412
Penicillin-Streptomycin (Gibco®) 100 mL	Life technologies (Carlsbad, CA)	15140-122
Phusion® High-Fidelity PCR Master Mix with GC Buffer	NEB (Ipswich, MA)	M0532S
Polybrene® 10 mg/ml	Santa Cruz Biotechnology (Santa Cruz, CA)	sc-134220
Polypropylene Adhesive Film	Denville Scientific Inc (South Plainfield, NJ)	B1212-5
PowerPrep® HP Plasmid Maxiprep System	Origene (Rockville, MD)	NP100009

Prism 6	GraphPad Software, Inc. (La Jolla, CA)	
Puromycin dihydrochloride	Sigma-Aldrich (St. Louis, MO)	P9620
QIAprep Spin Miniprep Kit	Qiagen (Valencia, CA)	27104
QIAquick Gel Extraction Kit	Qiagen (Valencia, CA)	28704
QuantiTect Reverse Transcription Kit	Qiagen (Valencia, CA)	205313
QuantiTect SYBR Green PCR Kit	Qiagen (Valencia, CA)	204145
Quantity One 1-D Analysis Software	Bio-Rad Laboratories, Inc (Hercules, CA)	
S.O.C. Medium	Life technologies (Carlsbad, CA)	15544-034
Sequence Detection Software 2.4 for Real-Time PCR System	Applied Biosystems (Foster City, CA)	
Sorvall T-6000 Benchtop Centrifuge	Thermo Fisher Scientific (Waltham, MA)	
Sorvall™ RC 6 Plus Centrifuge	Thermo Fisher Scientific (Waltham, MA)	46910
T-47D	ATCC (Manassas, VA)	HTB-133
T4 DNA Ligase	NEB (Ipswich, MA)	M0202S
Thermo Scientific HyClone Trypan Blue	Thermo Fisher Scientific (Waltham, MA)	SV3008401
Thermo Scientific NAPCO* Series 8000DH CO2 Incubator	Thermo Fisher Scientific (Waltham, MA)	
Thermocycler C1000 Biorad	Bio-Rad Laboratories, Inc (Hercules, CA)	
ThermoGrid™ Rigid Strip 0.2mL PCR Tubes	Denville Scientific Inc (South Plainfield, NJ)	C18064
TrackIt DNA ladder	Life technologies (Carlsbad, CA)	10488-085
Trypsin 0.25%; 100mL	Thermo Fisher Scientific (Waltham, MA)	SH3004201
WI-38	ATCC (Manassas, VA)	CCL-75

10.4 Sequence of lnc-KCNJ9-2:2

ATAATAAAAGGCCAAACCTTTGCTCCAACTTTCTCCTTAGCTTCCCTTTGGATCTGGAAA
GCTGGGGACCCACACGGCAGAGCCATGGTACTGGAGGAGCCATTAACAAAGCTTTCAA
TAAACCTCTCTTTCTTGAAGTTACCTGAGAATGGATCCATTCCCTGCAACTGAAGATTCT
AAGGAACTGGGTTTCTCAGTATACAATGGGAATGGTTGGGAGGAGGTAAAGAGTAGAA
GACAGTATCAAGAATCCAGAGCCCAGCACCTGTAGTCCTAACTATTCAGATTCCTTGAG
CCCAGGAGTTTGAGTCCAGCCTGGACAACATATTGAGACCCCCATCTCTCTAAAAAAAA
AGAGAAAGAAAGAAGGAAAGAAAAAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAA
GAAAGAAAGAAAGAGAAAGAAAGAAGGAAAGAAGGAAAGAAGGAAAGAAAGAAAGAAA
GAGAAAGAAAGAAAAGAAGATTGTAGCTAGGGGGAGAGTAGGTGAAAAGATGAACAAC
ATGACCGGGAAGATTTCCTAATCTCACCACAGCCTGGCTCTACCTTAAGTCTTTAATAAA
AGCTTGACTGAAGGTACCAAGGTGTGCTGAAGTGGAAGCAAAGTTCTCCAAAGTCCAG
CATGGTAGACATCAGTGGTGGTAACCAAGGACAGACCCCAAGGCAAGGTGAACCTCAA
AAATGGAACCTCAAGTCTATGCAGTCCAGCTGCCCTCCCCACCAGAAAGTCCTTGTTCC
AGCCCAACATCAGTGCCTCTGAGTTTGTTTACTAGAAACAAAGGAAGAATTTCCTTGTAA
AAATATAGACAGAGTAGTCCCTGGCTTTCTCCTCTTGCAGGAAGGATGGATTCTCCCAT
TCCATACCATCTTTCCCCCACACTGGCCCCAGAAATACTTAATTCAACTATGTGAAAATA
AAGATTGTTTTTGGTTTGAGGGCATAGGGATCCATTTATCCTTATTCTTTATGAGGCACT
AAATTAGCTTTGTATGTTATTAAATGTGTCTCGTCAA