TECHNISCHE UNIVERSITÄT MÜNCHEN

Fakultät für Medizin

Investigation of the epithelial-mesenchymal transition in human mammary epithelial cells to identify factors that impact breast cancer progression

Laura Eichelberger

Vollständiger Abdruck der von der Fakultät für Medizin der Technischen Universität München zur Erlangung des akademischen Grades einer

Doktorin der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitz: Prof. Dr. Radu Roland Rad

Prüfer*innen der Dissertation:

1. Prof. Dr. Heiko Lickert

2. Prof. Dr. Wolfgang Wurst

Die Dissertation wurde am 06.07.2021 bei der Technischen Universität München eingereicht und durch die Fakultät für Medizin am 07.12.2021 angenommen.

Parts of this thesis have been published in a pre-print at bioRxiv (Eichelberger et al., 2020):

Maintenance of epithelial traits and resistance to mesenchymal reprogramming promote proliferation in metastatic breast cancer

Laura Eichelberger, Massimo Saini, Helena Domínguez Moreno, Corinna Klein, Johanna M. Bartsch, Mattia Falcone, Manuel Reitberger, Elisa Espinet, Vanessa Vogel, Elisabeth Graf, Thomas Schwarzmayr, Tim-Matthias Strom, Mareike Lehmann, Melanie Königshoff, Nicole Pfarr, Roberto Würth, Elisa Donato, Simon Haas, Saskia Spaich, Marc Sütterlin, Andreas Schneeweiss, Wilko Weichert, Gunnar Schotta, Andreas Trumpp, Martin R. Sprick, and Christina H. Scheel

Table of contents

Table of contents

Summary

The outgrowth of metastatic cancer cells at distant organs causes the majority of breast cancer-related deaths. For cancer cells to escape from the primary tumour, disseminate within the body, and reach distant sites, the epithelial-mesenchymal transition (EMT) is an essential mechanism. During EMT, cancer cells lose epithelial traits and gain mesenchymal and migratory characteristics. However, distant breast cancer metastases typically display an epithelial identity, and breast cancer cells that can revert EMT have an enhanced tumorigenic potential. Therefore, the aim of this study was to unravel why some breast cancer cells that had undergone an EMT can revert to an epithelial state, whereas others remain mesenchymal. To do so, I used immortalised human mammary epithelial cells (HMLE-Twist1-ER), in which EMT is induced by activating the EMT-transcription factor (TF) Twist1. Upon transient activation of Twist1 in these cells, I observed the existence of two subsets: a subset of EMT-competent cells and a subset of EMT-resistant cells.

EMT-competent cells underwent an irreversible Twist1-mediated EMT, indicated by the decreased expression of epithelial markers like E-cadherin (CDH1) and EPCAM, the increased expression of mesenchymal markers like Vimentin, and the gain of a mesenchymal morphology. This irreversible trans-differentiation to a mesenchymal state inhibited the colonising capacity of the cells in a three-dimensional environment.

In EMT-resistant cells, the expression of mesenchymal markers also increased during transient Twist1-activation. However, the expression of epithelial markers was maintained, suggesting that these cells transiently gained an epithelial-mesenchymal hybrid state. As these cells could revert to a fully epithelial state after transient Twist1-activation, they could colonise as epithelial organoids in a three-dimensional environment.

To unravel what defines resistance to Twist1-mediated EMT or competence to undergo such an EMT, I set out to analyse changes in chromatin accessibility in both subsets of cells, as I hypothesised that EMT-resistance and EMT-competence were regulated on the chromatin level. To investigate this hypothesis, an assay for transposase-accessible chromatin (ATAC-sequencing) was performed in both subsets of cells. This analysis revealed that Twist1-activation resulted in different changes in chromatin accessibility in cells that resisted EMT and cells that underwent trans-differentiation to a mesenchymal state. Interestingly, simultaneous analysis of the transcriptome by RNA-sequencing revealed that changes in chromatin accessibility were reflected on the transcript level. The TF binding motif analysis of chromatin regions that displayed differential accessibility in both subsets of cells during Twist1-activation identified ZEB1 and GRHL2 as important regulators of EMT. Further analyses of transcript and protein levels revealed high levels of the EMT-inducing TF ZEB1 and the consequent repression of genes associated with an epithelial state, including *GRHL2*, *OVOL2*, and *MIR200* in cells that underwent Twist1-mediated EMT. In contrast, in cells that resisted trans-

differentiation upon Twist1-activation, ZEB1 levels were lower and, as a result, not sufficient to repress the expression of epithelial genes. By performing a CRISPR/Cas9-mediated knockout of *ZEB1* in cells that were previously competent to undergo Twist1-mediated EMT, I further underpinned the importance of ZEB1 for EMT, as the loss of *ZEB1* resulted in inhibition of Twist1-mediated EMT.

I hypothesised that high levels of ZEB1 could trigger EMT in the subset of cells that previously resisted Twist1-mediated EMT. Remarkably, by introducing an inducible *ZEB1* overexpression in both EMT-resistant and EMT-competent cells, I could trigger EMT in both subsets of cells. After transient induction of *ZEB1* overexpression, EMT-competent cells remained in a mesenchymal state, whereas EMT-resistant cells could revert to an epithelial state. Of note, upon withdrawal of induction, in cells that reverted to an epithelial state, *ZEB1* levels were decreased below a certain threshold resulting in the re-expression of epithelial genes and the reversion from the mesenchymal to an epithelial state.

In conclusion, the present study shows that induction of EMT-TFs induces an irreversible EMT in a subset of breast cancer cells, which inhibits epithelial colonisation in a three-dimensional environment. On the contrary, the complete loss of the epithelial identity is prevented in the other subset of breast cancer cells, which maintains the ability to colonise in a three-dimensional environment.

Zusammenfassung

Das Auswachsen metastasierter Tumorzellen in Organen bedingt die Mehrheit aller Brustkrebs-assoziierten Todesfälle. Um sich vom Primärtumor zu lösen, sich im Körper zu verbreiten und entfernte Gewebe zu erreichen, ist die epithelial-mesenchymale Transition (EMT) ein wesentlicher Mechanismus für Tumorzellen. Während der EMT verlieren Krebszellen epitheliale Eigenschaften und gewinnen stattdessen mesenchymale und migratorische Charakteristika. Brustkrebs-assoziierte Metastasen weisen jedoch typischerweise epitheliale Merkmale auf, und Brustkrebszellen, die zu einem epithelialen Status zurückkehren können, haben ein erhöhtes tumorigenes Potenzial. Daher war das Ziel dieser Arbeit, herauszufinden warum manche Brustkrebszellen nach einer EMT zu einem epithelialen Status zurückkehren können, wohingegen andere mesenchymal bleiben.

Hierfür habe ich Versuche mit humanen Brustepithelzellen (HMLE-Twist1-ER) durchgeführt, in denen EMT durch die Aktivierung des EMT-Transkriptionsfaktors (TF) Twist1 induziert werden kann. Durch die transiente Aktivierung von Twist1 in diesen Zellen zeigte sich, dass zwei Untergruppen existierten: eine Untergruppe von EMT-kompetenten Zellen und eine Untergruppe von EMT-resistenten Zellen.

EMT-kompetente Zellen durchliefen eine irreversible Twist1-induzierte EMT, was durch die verminderte Expression von epithelialen Markern wie E-Cadherin (CDH1) und EPCAM, die erhöhte Expression von mesenchymalen Markern wie Vimentin und das Erlangen einer mesenchymalen Morphologie verdeutlicht wurde. Diese irreversible Transdifferenzierung in einen mesenchymalen Status inhibierte das Auswachsen der Zellen in einer dreidimensionalen Umgebung.

Auch in EMT-resistenten Zellen nahm die Expression von mesenchymalen Markern während der transienten Twist1-Aktiviertung zu. Die Expression von epithelialen Markern blieb jedoch erhalten, was darauf hindeutete, dass diese Zellen vorübergehend einen epithelial-mesenchymalen Hybridzustand erlangten. Da diese Zellen nach transienter Twist1-Aktivierung einen epithelialen Status zurückerlangten, konnten sie als epitheliale Organoide in einer dreidimensionalen Umgebung auswachsen.

Um herauszufinden, was die Resistenz gegen eine Twist1-vemittelte EMT beziehungsweise die Kompetenz eine solche EMT zu durchlaufen definiert, war mein Ziel, die Veränderungen in der Zugänglichkeit des Chromatins in beiden Untergruppen zu analysieren. Grund hierfür war die Hypothese, dass EMT-Resistenz und EMT-Kompetenz auf Chromatin-Ebene reguliert wurden. Daher wurde in beiden Untergruppen ein Assay für Transposase-zugängliches Chromatin (ATAC-Sequenzierung) durchgeführt. Dies zeigte, dass die Twist1-Aktivierung zu unterschiedlichen Veränderungen der Zugänglichkeit von Chromatin in EMT-kompetenten und EMT-resistenten Zellen führte. Die gleichzeitige Analyse des Transkriptoms mithilfe von RNA-Sequenzierung zeigte, dass sich die Veränderungen der Zugänglichkeit des Chromatins auf

der Transkriptionsebene widerspiegelte. Durch die Analyse der TF-Bindungsmotive in Chromatinregionen, die während der Twist1-Aktivierung eine unterschiedliche Zugänglichkeit in beiden Untergruppen zeigten, konnten ZEB1 und GRHL2 als wichtige Regulatoren der EMT identifiziert werden. Untersuchungen der Transkript- und Proteinlevels zeigten hohe Levels des EMT-induzierenden TF ZEB1 in Zellen, die eine Twist1-vermittelten EMT durchlaufen konnten. Die hohen Levels an ZEB1 führten zu der Repression von Genen, die mit einem epithelialen Status assoziiert sind, einschließlich *GRHL2*, *OVOL2* und *MIR200*. Im Gegensatz dazu waren die ZEB1-Levels in Zellen, die resistent gegenüber einer Twist1-vermittelten Transdifferenzierung waren, niedriger und daher nicht ausreichend, um die Expression von epithelialen Genen zu unterdrücken. Durch eine CRISPR/Cas9-vermittelte Deletion von *ZEB1* in Zellen, die zuvor eine Twist1-vermittelte EMT durchlaufen konnten, konnte ich die Bedeutung von ZEB1 für die EMT weiter aufzeigen, da der Verlust von *ZEB1* zur Inhibition der Twist1-vermittelten EMT führte.

Ich stellte die Hypothese auf, dass hohe ZEB1-Levels eine EMT in Zellen induzieren könnten, die zuvor resistent gegenüber einer Twist1-vermittelten EMT waren. Mithilfe einer induzierbaren ZEB1-Überexpression konnte ich eine EMT in beiden Untergruppen an Zellen induzieren. Nach der transienten Induktion der ZEB1-Überexpression verweilten EMT-kompetente Zellen in einem mesenchymalen Zustand, während EMT-resistente Zellen in einen epithelialen Zustand zurückkehren konnten. Nachdem die Induktion gestoppt wurde, sanken die ZEB1-Levels in Zellen, die in einen epithelialen Zustand zurückkehrten, unter einen bestimmten Schwellenwert, was zur Re-Expression von epithelialen Genen und zur Transition vom mesenchymalen in einen epithelialen Status führte.

Zusammenfassend zeigt die vorliegende Arbeit, dass die Induktion von EMT-TFs eine irreversible EMT in einer Untergruppe von Brustkrebszellen induziert, was wiederum dazu führt, dass das Auswachsen in einer dreidimensionalen Umgebung gehemmt wird. Im Gegensatz dazu wird der vollständige Verlust der epithelialen Identität in der anderen Untergruppe von Brustkrebszellen verhindert, wodurch die Fähigkeit in einer dreidimensionalen Umgebung auszuwachsen erhalten bleibt.

List of Tables

List of Figures

List of Abbreviations

–	Negative
%	Percent
°C	Degree Celsius
+	Positive
2D	Two-dimensional
3'UTR	Three-prime untranslated region
3D	Three-dimensional
7-AAD	7-Aminoactinomycin D
A	Ampere
ACTB	Actin beta
AP-1	Activator protein-1
APS	Ammonium peroxydisulfate
ATAC	Assay for transposase-accessible chromatin
ATM	Ataxia telangiectasia mutated
bHLH	Basic helix-loop-helix
Blast	Blasticidin
bp	Base pair(s)
BPE	Bovine pituitary extract
BRCA1	Breast cancer associated gene 1
BRCA2	Breast cancer associated gene 2
BRIP1	BRCA1-interacting protein 1
BSA	Albumin fraction V, Bovine serum albumin
C_2H_2	Cysteine 2-Histidine 2
Cas9	CRISPR associated protein 9
CDH1	Cadherin-1, E-cadherin
cDNA	Complementary DNA
CHEK2	Checkpoint kinase 2
ChIP	Chromatin immunoprecipitation
CID	CtBP interaction domain
cm	Centimetres
CMV	Cytomegalovirus
CO_2	Carbon dioxide
CRISPR	Clustered regularly interspaced short palindromic repeats
Ct	Cycle threshold
CtBP1	C-terminal binding protein 1

CtBP2	C-terminal binding protein 2
CTC	Circulating tumour cell
Ctrl	Control
d	Days
DAPI	4′,6-diamidino-2-phenylindole
DC	Detergent-compatible
DCIS	Ductal carcinoma in situ
DMEM	Dulbecco's modified eagle medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
dNTP	Deoxyribonucleotide
Dox	Doxycycline
E-GUS	E-SCC containing an inducible GUS overexpression construct
E-SCC	EMT-resistant SCC
E-ZEB1	E-SCC containing an inducible ZEB1 overexpression construct
E12	Alternative splicing product of TCF3
E47	Alternative splicing product of TCF3
ECL	Enhanced chemiluminescence
ECM	Extracellular matrix
EDTA	Ethylenediaminetetraacetic acid
EGF	Epidermal growth factor
EMT	Epithelial-mesenchymal transition
EPCAM	Epithelial cell adhesion molecule
ER	Oestrogen receptor
ESRP1	Epithelial splicing regulatory protein 1
FACS	Fluorescent-activated cell sorting
FCS	Fetal calf serum
FITC	Fluorescein isothiocyanate
FN1	Fibronectin 1
FOXC2	Forkhead box C2
g	Gram
GATA3	GATA binding protein 3
gDNA	Genomic DNA
GEO	Gene expression omnibus
GFI3	Growth factor independent 1 transcriptional repressor

GFP	Green fluorescent protein
GRHL1	Grainyhead like transcription factor 1
GRHL2	Grainyhead like transcription factor 2
GRHL3	Grainyhead like transcription factor 3
gRNA	Guide RNA
GUS	beta-glucuronidase
h	Hours
H_2O	Water
HAS2	Hyaluronan Synthase 2
HCl	Hydrochloric acid
HDAC1	Histone deacetylase 1
HDAC2	Histone deacetylase 2
HEPES	N-2-hydroxyethylpiperazine-N-2-ethane sulfonic acid
HER2	Human epidermal growth factor receptor 2
HLH	Helix-loop-helix
HMLE	Immortalised human mammary epithelial cells
HR	Hormone receptor
HRP	Horseradish peroxidase
hTERT	Human telomerase reverse transcriptase
hU6	Human U6 promoter
ID1-4	Inhibitor of DNA binding 1-4
IDC	Invasive ductal carcinoma
IF	Immunofluorescence
IgG	Immunoglobulin G
ILC	Invasive lobular carcinoma
IRES2	Internal ribosomal entry site type II
kb	Kilobase(s)
KO	Knockout
l	Litres
LB	Lysogeny broth
lg	$\log_{10}$
LKB1	Liver kinase B1
M	Molar
m	Milli
M-GUS	M-SCC containing an inducible GUS overexpression construct
M-SCC	EMT-competent SCC

M-ZEB1	M-SCC containing an inducible ZEB1 overexpression construct
m/v	mass/volume
mBC	Metastatic breast cancer
mCRC	Metastatic colorectal cancer
MECGM	Mammary epithelial cell growth medium
MET	Mesenchymal-epithelial transition
Milli-Q	Ultrapure water
min	Minute(s)
miR-200	Micro RNA 200 family
MIR141	Micro RNA 141
MIR200A	Micro RNA 200a
MIR200B	Micro RNA 200b
MIR200C	Micro RNA 200c
miRNA	Micro RNA
mPC	Metastatic pancreatic cancer
mRNA	Messenger RNA
mtDNA	Mitochondrial DNA
mTwist1	Mouse Twist1
n	Number or nano
n.s.	Not significant
N20	20-nucleotide
NaCl	Sodium chloride
NaOH	Sodium hydroxide
neg	Negative
NM	NCBI reference sequence
nm	Nanometres
NP40	Nonidet P40
nt	Nucleotide(s)
O_2	Oxygen
OD	Optical density
OE	Overexpression
Oligo	Oligonucleotide
ORF	Open reading frame
OVOL2	Ovo like zinc finger 2
p_{adj}	Adjusted p-value
PALB2	Partner and localizer of BRCA2

PB	PiggyBac
PBS	Phosphate buffered saline
PC	Principal component
PCR	Polymerase chain reaction
PFA	Paraformaldehyde
PIK3CA	Phosphatidylinositol-4,5-Bisphosphate 3-Kinase Catalytic Subunit Alpha
pos	Positive
PR	Progesterone receptor
PRRX1	Paired Related Homeobox 1
PTEN	Phosphatase and tensin homolog
PVDF	Polyvinylidene fluoride
qPCR	Real-time semi-quantitative PCR
RAD51C	RAD51 paralog C
RAD51D	RAD51 paralog D
RB	Retinoblastoma
rcf	Relative centrifugal force
RIPA	Radioimmunoprecipitation assay
RN	Normalized reporter value
RNA	Ribonucleic acid
RNase	Ribonuclease
RPL32	Ribosomal protein 32
rpm	revolutions per minute
RT	Reverse transcription
RTase	Reverse transcriptase
SBD	Smad binding domain
SCC	Single-cell clone
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium dodecyl sulphate-Polyacrylamide gel electrophoresis
sec	Second (s)
SEM	standard error of the mean
SEM	Standard error of the mean
shRNA	short hairpin RNA
SIN3A	SIN3 transcription regulator family member A
SIP1	Alias for ZEB2
SLUG	Alias for SNAI2
SMUC	Alias for SNAI3

SNAG	SNAIL/GFI3
SNAI1	Snail Family Transcriptional Repressor 1
SNAI2	Snail Family Transcriptional Repressor 2
SNAI3	Snail Family Transcriptional Repressor 3
SNAIL	Alias for SNAI1
STAgR	String assembly gRNA
STK11	Serine/threonine kinase 11
SV40	Simian Virus 40
TAE	Tris-acetate-EDTA
TAM	(Z)-4-Hydroxytamoxifen, Tamoxifen
TBS	Tris-buffered saline
TBS/T	Tris-buffered saline with Tween-20
TC	Tissue culture
TCF3	Transcription factor 3
TCF4	Transcription Factor 4
TE	Tris-EDTA
TEMED	Tetramethylenediamine
TF	Transcription factor
TGFBRI	TGFβ receptor 1
TGFβ	Transforming growth factor beta
Tn5	Tn5 transposase
TNBC	Triple-negative breast cancer
TNS	Trypsin Neutralizing Solution
TP53	Tumour protein 53
TWIST1	Twist family bHLH transcription factor 1
V	Voltage
v/v	Volume/volume
VIM	Vimentin
WB	Western blot/Immunoblot
WNT5A	Wnt family member 5A
ZEB1	Zinc finger E-box-binding homeobox 1
ZEB2	Zinc finger E-box-binding homeobox 2
Δ	Delta
δEF1	Alias for ZEB1
λ	Lambda, wavelength
μ	Micro

1   Introduction

1.1   Development of the normal human breast

The bi-layered structure of the mammary gland is formed by two main epithelial cell types, the basal and the luminal cells. The basal cells compromise the outer layer of the glands, which is surrounded by the basement membrane, and the luminal cells are positioned on top of the basal cells facing the lumen (Macias & Hinck, 2012) (Figure 1-1a).

Within the three major stages of female breast development – embryonic, pubertal, and reproductive – changes in shape, size, and function of the mammary gland occur (Macias & Hinck, 2012; Russo & Russo, 2004) (Figure 1-1b).
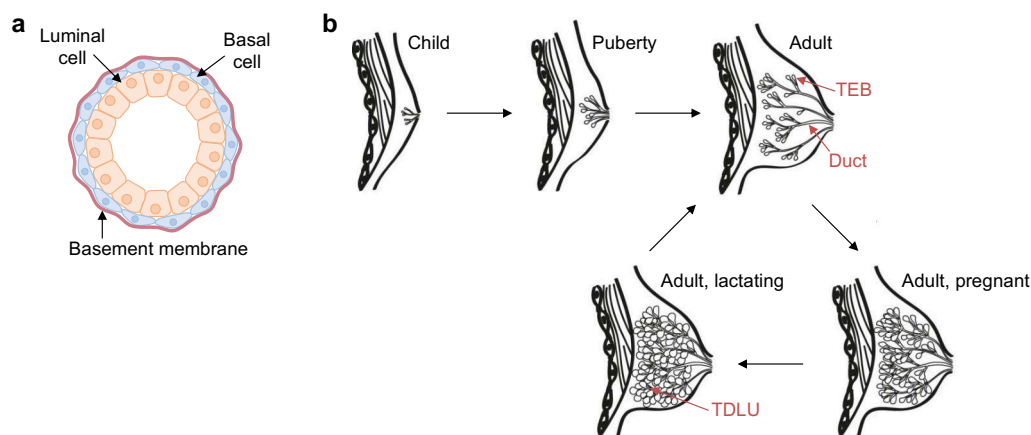
Figure 1-1: Structure and developmental stages of the female mammary gland
a) Ducts of the human mammary gland are composed of two cell types surrounded by a basement membrane. Basal cells compromise the outer cell layer, and luminal cells are positioned on top of the basal cells facing the lumen. **b)** Upon the onset of puberty in females, the primitive structure of epithelial ducts starts branching, forms terminal end buds (TEBs) and undergoes intensive budding. During pregnancy, the functional units of the mammary gland, the terminal ductal lobular units (TDLUs), further grow and fully differentiate during lactation. After weaning, the glands revert to a non-pregnant state by involution. Courtesy of Jelena Krendl (neé Linnemann), modified.

During embryogenesis, the mammary gland arises from one epithelial bud that develops into a primitive structure of multiple ducts (Gusterson & Stein, 2012; Russo & Russo, 2004) (Figure 1-1b). The mammary gland then remains quiescent until hormonal changes during female puberty initiate further development. During this time, ducts elongate and form club-shaped terminal end buds, giving rise to new branches, small ducts, and alveolar terminal end buds (TEBs, Figure 1-1b). As the breast develops further, the functional units of the mammary gland are formed. These consist of several acini and are known as the terminal ductal lobular units (TDLUs) (Parmar & Cunha, 2004; Russo & Russo, 2004) (Figure 1-1b). During pregnancy, the mammary gland undergoes further intense budding, and the secretory activity of the luminal cells is activated (Figure 1-1b). However, the contractility of the basal cells, which facilitates

the transport of the secreted milk to the nipple, is triggered as soon as the child is born. Only when the child is weaned, the changes in the mammary gland are reversed through a process known as involution (Russo & Russo, 2004) (Figure 1-1b).

1.2 Breast Cancer

1.2.1 Current breast cancer statistics

Throughout the lifetime of a woman, the mammary gland undergoes several cycles of thorough remodelling involving extensive proliferation and subsequent cell death (see chapter 1.1). Hence, it is not unusual in a tissue of such high regenerative potential, that cancer is likely to occur. Indeed, breast cancer is one of the most common cancer types in women. Approximately one out of eight women will be diagnosed with breast cancer during a lifetime. For 2020, nearly 70,000 new cases of female breast cancer were estimated in Germany, accounting for roughly 30% of all new cancer cases in women. Of these, approximately 18,400 women were suspected of dying from the disease (Robert-Koch-Institut, 2019).

The stage of cancer at the time of diagnosis drastically influences the 5-year survival rate of patients. If breast cancer is detected early, when the cancer is still localised, the 5-year survival rate is roughly 99%. However, a decline of the 5-year survival rate to less than 30% is observed in patients diagnosed when cancer has already metastasised (Howlader et al., 2020).

1.2.2 Breast cancer classification

21 years ago, researchers started intensive gene-expression analyses of breast cancers using approaches like microarrays and hierarchical clustering (Perou et al., 2000; Sørlie et al., 2001, 2003). By doing so, they identified different subtypes of breast cancer that largely corresponded to already known immunohistochemical subtypes: an oestrogen receptor (ER)-positive luminal-like subtype, an ER-, progesterone receptor (PR)- and human epidermal growth factor receptor (HER2)-negative basal-like subtype, a HER2-enriched subtype, and a potential normal breast-like subtype (Perou et al., 2000). Within the group of luminal-like breast cancers, two subgroups were found that differed mainly in proliferative rates and are now known as Luminal A and Luminal B subtypes (Sørlie et al., 2001, 2003). Some years later, an additional subtype within the subgroup of triple-negative breast cancer (TNBC) was identified by gene expression analyses. This subtype is known as claudin-low breast cancer as cells have a low or absent expression of luminal markers and tight junction genes like *CDH1* (E-cadherin) and claudins (*CLDN3*, *CLDN4*, *CLDN7*) (Herschkowitz et al., 2007; Prat et al., 2010). Nowadays, breast cancers are classified using the parameters tumour size (T), affected lymph nodes (N), distant metastases (M), and grade. The determination of the grade is usually based

on a modification of the grading proposed by Bloom and Richardson, according to Elston and Ellis (Leitlinienprogramm Onkologie, 2020). It includes three parameters used to determine the grade of breast cancers: tubular formation, nuclear pleomorphism and mitotic rate. The amount of tubular formation determines the degree of cell differentiation, which is defined by how well tumour cells form mammary glands. Nuclear pleomorphism focuses on the size and shape of cells and their nuclei, as these parameters typically correlate with tissue atypia. Last but not least, the mitotic rate is calculated by the number of proliferative tumour cells (Leitlinienprogramm Onkologie, 2020). If the extent of the individual parameters is known, the grade of the tumour can be determined, which directly correlates with the tumour's aggressiveness. The histopathological analysis of breast cancers also includes the expression of the oncogene HER2, hormone receptors (HRs) for oestrogen (ER) and progesterone (PR), and the Ki-67 proliferation index. By doing so, four main breast cancer subtypes are classified as shown in Table 1-1 (Hortobagyi et al., 2017).

Table 1-1: Parameters for the classification of different breast cancer subtypes

Breast cancer subtype	Parameters
Luminal A	High HR and low proliferation rate (low Ki-67), HER2-negative
Luminal B	Lower HR and high proliferation rate (high Ki-67), HER2-negative or -positive
HER2-enriched	HR-negative or -positive and HER2-positive
Basal-like	HR-negative and HER2-negative

To understand the frequency and type of mutated genes in breast cancers, several analyses were carried out. The cancer genome atlas network, for example, investigated around 500 tumour samples from breast cancer patients (Cancer Genome Atlas Network, 2012). This analysis revealed that heterogeneity occurs in all four main breast cancer subtypes. In the luminal breast cancer subtypes, somatic mutations were detected in genes like *PI3KCA* (Phosphatidylinositol-4,5-Bisphosphate 3-Kinase Catalytic Subunit Alpha; 45% in Luminal A, 29% in Luminal B), *TP53* (Tumour Protein P53; 12%, 29%), and *GATA3* (GATA Binding Protein 3; 14%, 15%). In contrast to that, mutations in the *TP53* gene occurred at a much higher frequency in the basal-like subtype (80%), whereas mutations in *PI3KCA* and *GATA3* were found in only 9% and 2% of these cases, respectively. Mutations in the *TP53* gene were also found in 72% of HER2-enriched tumour samples, whereas 39% of the HER2-enriched samples displayed a mutated *PI3KCA*, and only 2% harboured mutations in *GATA3* (Cancer Genome Atlas Network, 2012).

1.2.3 Hereditary breast cancer

Approximately 10% of all female breast cancer cases occur due to germline mutations (Thomssen & Wand, 2012). The penetrance of a mutation within a specific gene typically varies and depends on the affected gene. Therefore, three groups of genes are defined based on their penetrance (Figure 1-2). The group of genes with high penetrance when mutated includes genes such as breast cancer-associated genes *BRCA1* and *BRCA2*, serine/threonine kinase 11/liver kinase B1 (*STK11/LKB1*), phosphatase and tensin homolog (*PTEN*), E-cadherin (*CDH1*), and *TP53* (Figure 1-2). In the group of moderate penetrance, genes are included that play a role in DNA-repair mechanisms, like checkpoint kinase 2 (*CHEK2*), ataxia telangiectasia mutated (*ATM*), BRCA1-interacting protein 1 (*BRIP1*), and partner and localizer of BRCA2 (*PALB2*). The group of genes associated with low penetrance when mutated includes two members of the RAD51 recombinase family (*RAD51C*, *RAD51D*) and mitochondrial DNA (*mtDNA*) (Mahdavi et al., 2019).

High penetrance genes   **Low penetrance genes**

BRCA1 & BRCA2
STK11/LKB1
PTEN
CDH1
TP53

CHEK2
ATM
BRIP1
PALB2

RAD51C
RAD51D
mtDNA

Moderate penetrance genes
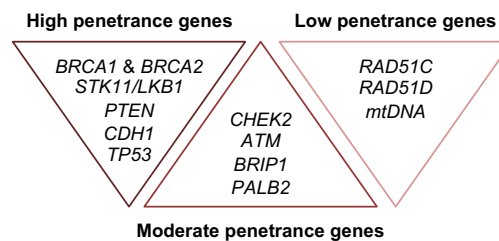
Figure 1-2: Hereditary breast cancer genes
List of genes associated with high, moderate, or low penetrance. Mahdavi et al., 2019, modified.

In proportional terms, mutations in *BRCA1* or *BRCA2* amount with approximately 25 to 55% to the largest share within female breast cancer cases with germline mutations (Thomssen & Wand, 2012). Mutations in *BRCA1* and or *BRCA2* can strongly predict the probability of developing breast cancer, as carriers have an increased risk of 40 to 80% for the development of breast cancer within their lifetime (Fackenthal & Olopade, 2007). However, besides predisposing to breast, mutations in *BRCA* genes also increase prostate, ovarian, and male breast cancer risk. Until now, more than 1600 different mutations were detected in *BRCA1* and more than 1800 different mutations in *BRCA2*, which mainly result in truncation or loss of the proteins (Gorodetska, Kozeretska, & Dubrovska, 2019). Both BRCA1 and BRCA2 proteins are defined as tumour suppressors because of their role in DNA repair and concomitant cell cycle control (Roy, Chun, & Powell, 2011).

1.2.4 In situ and invasive breast cancer

Breast cancers are categorized depending on whether the cancer is still localised (in situ) or has already breached the basement membrane and invaded into the surrounding tissue (invasive). Within breast carcinomas, the in situ carcinomas are divided into two categories: ductal carcinoma in situ (DCIS) and lobular carcinoma in situ (LCIS) (Alkabban & Ferguson, 2020). When LCIS develops, cancer cells proliferate within the acini of the TDLUs of the mammary glands and fill more than half of the space of all acini (Wen & Brogi, 2018). DCIS, which accounts for approximately 10% of all new breast cancer cases in Germany (Deutsche Krebsgesellschaft, 2020), is characterised by malignant epithelial cells that proliferate within the ductal units of the mammary gland without breaching the basement membrane (Bane, 2013). Nevertheless, DCIS can further progress to invasive breast cancer (IBC). However, not all cases of DCIS progress to an invasive form of breast cancer. Unfortunately, many patients are often overtreated as it is not clear yet how to determine whether DCIS remains localised or progresses to IBC. One of the reasons for this is that DCIS often exhibits intratumoral heterogeneity (Bergholtz et al., 2020; Sinha & Piwnica-Worms, 2018).

To identify whether the different molecular subtypes of breast cancer (see chapter 1.2.2) are associated with different progression rates, a recent study compared DCIS samples with IBC samples obtained from 57 and 313 patients, respectively (Bergholtz et al., 2020). Interestingly, the most frequently detected breast cancer subtype in both DCIS and IBC was the Luminal A subtype (43.9% of all DCIS, 47.3% of all IBC). However, differences in the frequency of Luminal B, HER2-enriched, and basal-like subtypes were detected when comparing DCIS and IBC (Bergholtz et al., 2020). Of all DCIS samples, only 8.8% were categorised as Luminal B breast cancer subtype, whereas this breast cancer subtype was identified in 26.5% of all IBC samples. Interestingly, an opposite trend was observed for the HER2-enriched and the basal-like breast cancer subtypes (Bergholtz et al., 2020).

1.2.5 Breast cancer subtype outcome and systemic treatment

Treatment of breast cancers strongly depends on the breast cancer subtype. Although most patients undergo surgery to remove the primary tumour, systemic therapies are applied before (neoadjuvant) and/or after surgery (adjuvant) to shrink the tumour before surgery or target residual cancer cells within the body, respectively. Patients with Luminal A cancers, for example, can benefit from endocrine therapies targeting hormone receptor pathways, including drugs like Tamoxifen and aromatase inhibitors (American Cancer Society, 2020; Hortobagyi et al., 2017). Tamoxifen acts as an oestrogen antagonist in mammary epithelial cells, thereby inhibiting breast cancer cells' proliferation (Sporn & Lippman, 2003).

Aromatase inhibitors inhibit the aromatase enzyme activity, which leads to a repressed production of oestrogen (American Cancer Society, 2020).

As Luminal A tumours also have a lower proliferation rate, patients diagnosed with this breast cancer subtype have a particularly good prognosis. For Luminal B tumours, which have a worse prognosis due to higher proliferation rates, it is known that, despite the expression of HRs, this subtype is more successfully treated with chemotherapy (Hortobagyi et al., 2017). Chemotherapeutics like anthracyclines and taxanes are mostly used in combination and target proliferating cells by interrupting cell division processes. This treatment option is also used for patients diagnosed with basal-like breast cancer. Due to the lack of expression of HRs and HER2 (triple-negative breast cancer, TNBC), this breast cancer subtype is the most difficult to treat, and chemotherapy is the primary option. Nevertheless, the combination of chemotherapeutic treatments and immune-checkpoint inhibitors, like PD-L1, increases response rate, progression-free survival, and overall survival of patients (Schardt, 2020).

Besides TNBC, HER2-enriched tumours are also considered highly aggressive tumours, despite the development of the anti-HER2 therapy, which improved patient outcome to some extent (Hortobagyi et al., 2017). Trastuzumab, for example, is an antibody that binds the extracellular domain of HER2 and thereby inhibits downstream cell proliferation signalling (Nahta & Esteva, 2006).

1.3   The epithelial-mesenchymal transition

The epithelial-mesenchymal transition (EMT) is a process that was first observed during embryonic development (Trelstad, Hay, & Revel, 1967). However, it soon became apparent that EMT is also important during wound healing, fibrosis, and cancer progression (Thiery, Acloque, Huang, & Nieto, 2009). During EMT, epithelial cobblestone-like cells with an apico-basal polarity trans-differentiate to cells with a spindle-like morphology and a front-to-back polarity. These changes are accompanied by the dissolution of cell-cell adhesions and the acquirement of migratory traits (Kalluri & Weinberg, 2009; Yang & Weinberg, 2008).

1.3.1   EMT in development

EMT can be observed at different stages during the development of a vertebrate embryo. Either already early during development, when the mesoderm and the neural crest formations occur, or during later stages in development, for example, when the cardiac valve or the secondary palate form (Yang & Weinberg, 2008). During EMT in the early phase of vertebrate development, cells undergo EMT and migrate to other parts within the embryo, where they further differentiate to generate new tissues. For the formation of the mesoderm during

gastrulation, the epithelial cells that sit in the primitive streak of the epiblast undergo EMT, breach the basement membrane, lose their cell-cell contacts, and migrate between the endo- and the ectoderm (Yang & Weinberg, 2008). After completion of the gastrulation process, when the neural plate and epidermal ectoderm are defined, the formation of the neural crest takes place. During this process, the neural plate first invaginates, and neural crest cells undergo EMT. These cells then invade through the basal lamina into the extracellular matrix (ECM) and distribute within the embryo. Later during development, these cells contribute to diverse structures, including the head skeleton, peripheral nervous system, and different types of neurons (Bronner, 2012; Yang & Weinberg, 2008).

1.3.2 EMT in breast cancer progression – the metastatic cascade

Metastases cause approximately 90% of all cancer deaths (Chaffer & Weinberg, 2011; Fidler, 2002). Indeed, the 5-year survival rate of patients diagnosed with metastasised breast cancer is less than 30% (Howlader et al., 2020). Interestingly, the majority of primary breast tumours, as well as metastases, show epithelial differentiation, and cells within these tumours form strong cell-cell contacts, indicated by the expression of E-cadherin (CDH1), one of the most studied epithelial cell-cell junction proteins (Kowalski, Rubin, & Kleer, 2003).
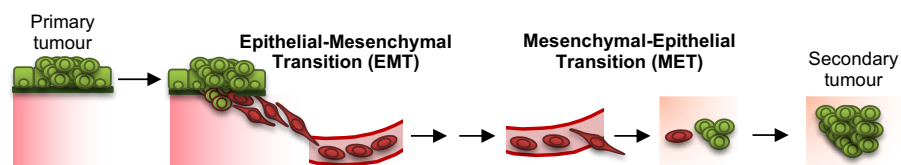
Figure 1-3: The metastatic cascade during tumour progression
Epithelial cells from the primary tumour undergo an epithelial-mesenchymal transition (EMT), gain migratory traits, breach the basement membrane, and invade into the surrounding tissue. Cells then intravasate into blood and lymph vessels and distribute within the body. After extravasation at distant sites, cells need to undergo a mesenchymal-epithelial transition (MET) to regain epithelial traits and form micro-metastases that develop into macro-metastases. From Scheel & Weinberg, 2012, modified.

However, to form metastases, cells of the primary tumour need to undergo different trans-differentiation steps, which altogether are known as the metastatic cascade. Briefly, cancer cells leave the primary tumour, distribute within the body, and finally settle and colonise at distant sites (Scheel & Weinberg, 2012) (Figure 1-3). In detail, the metastatic cascade starts with polarised epithelial cells that undergo an epithelial-mesenchymal transition (EMT) (Geiger & Peeper, 2009). During EMT in cancer progression, a characteristic and crucial step is the decrease of epithelial proteins like E-cadherin (CDH1), the epithelial cell adhesion molecule

(EPCAM), and Claudins (CLDN), as well as the upregulation of mesenchymal markers like Vimentin (VIM) and Fibronectin (FN1) (Dongre & Weinberg, 2019; Geiger & Peeper, 2009; Lamouille, Xu, & Derynck, 2014; Thiery et al., 2009). The acquirement of migratory traits during EMT enables tumour cells to break through the basement membrane, invade into the tumour-surrounding tissue, and intravasate into lymph or blood vessels. It allows cancer cells to circulate and distribute within the body. For the formation of metastases, the circulating tumour cells (CTCs) extravasate at distant sites. However, the site of extravasation and the formation of metastases strongly depend on the primary tumour type. This phenomenon is best summarised by the "seed and soil hypothesis", which was first described in the late 19th century by Paget (Paget, 1889). Upon post-mortem investigation of more than 700 breast cancer patients, he described that the growth of tumour cells (seed) at distant sites appears to depend on the environment of the specific organ (soil) (Paget, 1889). Nowadays, it is known that breast cancer cells preferably metastasise to bone, liver, brain, lung, and distant lymph nodes, dependent on the breast cancer subtype. As breast cancer metastases typically grow as epithelial clusters (Kowalski et al., 2003), it is assumed that tumour cells undergo a mesenchymal-epithelial transition (MET) at the site of metastasis in order to revert to an epithelial morphology. Therefore, EMT followed by MET is also often described as epithelial-mesenchymal plasticity (EMP).

1.3.3 Circulating tumour cells

The detection of cancer cells that escaped from the primary tumour and circulate throughout the body within the bloodstream (circulating tumour cells, CTCs) can be used for the prognosis of metastatic breast cancer (Cristofanilli et al., 2005). An approved method for detecting CTCs is the CellSearch system, which uses cytokeratins and the epithelial cell adhesion molecule (EPCAM) to detect epithelial cells within the blood of patients with invasive cancer (Allard et al., 2004). The reason for choosing EPCAM as a marker to enrich for CTCs is the absence of epithelial cells within the bloodstream of healthy humans (Riethdorf, O'Flaherty, Hille, & Pantel, 2018). Studies isolating CTCs from patients with metastatic breast cancer (mBC), metastatic pancreatic cancer (mPC), or metastatic colorectal cancer (mCRC) revealed that the abundance of ≥ 3 CTCs (for mCRC) or ≥ 5 CTCs (for mBC and mPC) per 7.5 ml of blood correlates with a decreased overall as well as progression-free survival (Cohen et al., 2008; Cristofanilli et al., 2004; de Bono et al., 2008). Moreover, xenograft transplantation assays with isolated, EPCAM-expressing CTCs from blood samples of patients with metastatic breast cancer have identified a metastasis initiating capacity of these cells *in vivo* (Baccelli et al., 2013). However, the EPCAM-based method to enrich for CTCs has already been criticised in

other studies, as CTCs that had undergone an EMT and lost EPCAM expression are not detected when using this method (Gorges et al., 2012; Yu et al., 2013).

Besides single CTCs, also clusters of CTCs can be detected within the circulation of patients with metastatic cancer (Cho et al., 2012). Moreover, Aceto and colleagues have shown that CTC-clusters formed by several cells connected via plakoglobin, a component of desmosomes and adherens junctions, have an increased metastasis initiating capacity compared to single CTCs (Aceto et al., 2014).

1.3.4 Transcription factors orchestrating EMT

The trans-differentiation from an epithelial to a mesenchymal cell state is orchestrated by different EMT-inducing transcription factors like SNAIL1 and 2 (Snail Family Transcriptional Repressor 1 and 2), ZEB1 and 2 (Zinc finger E-box-binding homeobox 1 and 2), TWIST1 (Twist Family BHLH Transcription Factor 1), FOXC2 (Forkhead box C2), TCF4 (Transcription Factor 4), PRRX1 (Paired Related Homeobox 1) and others. All these EMT-transcription factors regulate the expression of, for example, E-cadherin (*CDH1*) and Vimentin (*VIM*), two genes whose repression and activation, respectively, are crucial for EMT. However, the most investigated and often termed master EMT-transcription factors belong to the SNAIL, TWIST, and ZEB families (De Craene & Berx, 2013).

1.3.4.1 SNAIL family

The SNAIL family consists of the three zinc-finger transcription factors SNAIL (*SNAI1*), SLUG (*SNAI2*), and SMUC (*SNAI3*). They all harbour four to six C_2H_2-type zinc fingers within the highly conserved C-terminal domain and mediate specific DNA binding to the E-box sequence 5'-CAGGTG-3'. The N-terminus of SNAIL family members contains the SNAG (SNAIL/GFI3) domain, which is necessary for protein interactions with co-repressors like histone deacetylases (HDACs) and proteins of the polycomb complex (de Herreros, Peiró, Nassour, & Savagner, 2010; Wu & Zhou, 2010). Within the SNAIL family, the importance of SNAIL and SLUG for EMT has been studied in most detail. First of all, SLUG has been identified by antisense oligonucleotide treatment in the developing chick embryo as an essential transcription factor mediating EMT during gastrulation (Nieto, Sargent, Wilkinson, & Cooke, 1994). SLUG represses the transcription of *CDH1* by binding to the promoter region (Bolós et al., 2003). Moreover, the upregulation of *SLUG* expression is important for Twist1-mediated EMT (Casas et al., 2011). Like SLUG, the transcription factor SNAIL represses the expression of *CDH1*. By recruiting HDAC1, HDAC2, and the co-repressor SIN3A to the promoter region,

it mediates chromatin modifications that result in repressed gene expression (Peinado, Ballestar, Esteller, & Cano, 2004).

1.3.4.2 TWIST family

The TWIST family belongs to the family of basic helix-loop-helix (bHLH) transcription factors that consists of many proteins that regulate the transcription of genes implicated in the development and progression of different types of cancer. Based on dimerization capability, DNA-binding specificity, and tissue distribution, bHLH proteins are categorized into seven classes (Massari & Murre, 2000). Importantly, all bHLH proteins contain two characteristic features: a basic domain that mediates binding to DNA and an HLH domain necessary for the formation of homo- or heterodimers with other bHLH factors (Murre et al., 1994). Of the seven classes of bHLH proteins, proteins from classes I, II, and V are implicated in EMT. In detail, E12 and E47 (class I proteins) form heterodimers with TWIST1 and TWIST2 (class II proteins). They can also bind to ID1-4 (class V proteins), although the latter misses the basic domain (Xu, Lamouille, & Derynck, 2009). In 2004, Twist1 was for the first time identified as an important regulator of EMT, E-cadherin (*CDH1*) repression, and metastasis formation (Yang et al., 2004). However, TWIST1 does not repress *CDH1* expression directly, but possibly by upregulating the expression of other EMT-transcription factors (Casas et al., 2011). In 2015, an analysis revealed that TWIST1 preferentially binds to a double E-box motif (5'-CANNTG-3') spaced by exactly five nucleotides and thereby regulates transcription of genes and induces EMT (Chang et al., 2015).

1.3.4.3 ZEB family

The family of zinc finger E-box binding homeobox (ZEB) proteins consists of ZEB1 (also known as δEF1) and ZEB2 (also known as SIP1). Both have a central homeodomain, a CtBP (C-terminal binding protein) interaction domain (CID), a Smad binding domain (SBD), and two zinc-finger clusters located on each end of the protein. The latter enable binding of the ZEB transcription factors to the E-box sequence 5'-CACCTG-3' (Vandewalle, Van Roy, & Berx, 2009). ZEB1 represses the expression of *CDH1* and *EPCAM* directly by binding to the E-boxes within the promoter region (Eger et al., 2005; Vannier, Mock, Brabletz, & Driever, 2013). For repression, ZEB1 interacts with BRG1, a component of the SWI/SNF chromatin-remodelling complex (Sánchez-Tilló et al., 2010) or recruits C-terminal binding proteins (CtBP1, CtBP2) that bind to the CtBP binding domain and lead to the repression of target genes (Furusawa, Moribe, Kondoh, & Higashi, 1999). However, ZEB1 activity is also controlled by various double-negative feedback loops, such as another important factor during development, the

transcription factor ovo like zinc finger 2 (OVOL2) (Hong et al., 2015; Watanabe et al., 2014). Moreover, ZEB1 and grainyhead-like transcription factor 2 (GRHL2) repress each other by binding to each other's promoter region (Cieply et al., 2012; Werner et al., 2013). Additionally, micro RNAs of the miR-200 family repress both ZEB1 and ZEB2 by binding to the three-prime untranslated region (3'UTR) of the corresponding mRNAs (Gregory et al., 2008; Park, Gaur, Lengyel, & Peter, 2008). However, ZEB1 can also repress the expression of miR-200 family members by binding to the E-boxes that are located within the promoter regions of the *MIR200* genes (Bracken et al., 2008; Burk et al., 2008).

1.3.5 Controversies around EMT and its implication in cancer progression

The importance of EMT during metastasis is under constant discussion. In the last years, several studies attempted to unravel the impact of EMT on tumour progression in different kinds of cancers. The reason for EMT being considered as a critical process during metastatic cancer progression is the observation of mesenchymal tumour cells escaping from the primary tumour and invading into the surrounding tissue (Williams, Gao, Redfern, & Thompson, 2019). Also, it has been shown that tumour cells with a mesenchymal phenotype have enhanced survival and tumour progression capacities compared to tumour cells with an epithelial phenotype (Williams et al., 2019).

However, there is only little evidence that mesenchymal tumour cells transition back to an epithelial state via MET at the site of metastasis. A study in 2016 that investigated tumour cells from a breast cancer mouse model with low or high E-cadherin levels has shown that both fractions of cells were able to grow liver metastases, suggesting that cells with low E-cadherin levels indeed transition back to an epithelial state (Beerling et al., 2016).

Nevertheless, other studies have shown that a transient rather than the permanent activity of EMT-transcription factors like Prrx1, Twist1, and Zeb1 was necessary for the metastatic outgrowth (Ocaña et al., 2012; Schmidt et al., 2015; Tsai, Donaher, Murphy, Chau, & Yang, 2012). In addition to these findings, metastases of different cancer types typically display epithelial features (Kowalski et al., 2003; Revenco et al., 2019).

Nonetheless, several other studies suggested an essential role of EMT for cancer metastasis. In a mouse model of squamous cell carcinoma, the abundance of tumour cells within the circulation increased, although the cells displayed a decrease in proliferation upon induction of the EMT-transcription factor Twist1 (Tsai et al., 2012). Moreover, other researchers suggested that the EMT-transcription factor Zeb1 is a crucial factor for metastasis because its deletion in a mouse model of pancreatic cancer decreased the invasion of tumour cells and their metastasising capacity to the lungs (Krebs et al., 2017). However, it did not affect primary tumour growth (Krebs et al., 2017). Another study in a mouse model of squamous skin

carcinoma revealed that at least some degree of EMT within the primary tumour was necessary for metastasis to the lungs, although EMT was not required for metastasis when cancer cells were transplanted subcutaneously (Revenco et al., 2019).

The assumption that EMT is required for metastasis was further questioned with the finding that the inhibition of EMT by overexpressing micro RNAs from the miR-200 family in a breast cancer mouse model did not inhibit metastasis to the lungs (Fischer et al., 2015). Furthermore, invasion of tumour cells and disease progression in a mouse model of pancreatic cancer were not affected when the EMT-transcription factors *Snai1* or *Twist1* were deleted (Zheng et al., 2015). Moreover, in models of invasive breast cancers, loss of E-cadherin enhanced the invasion of cancer cells but at the same time led to a decrease in metastatic outgrowth at distant sites (Padmanaban et al., 2019).

Further challenges in understanding the role of EMT in tumour progression have arisen due to the observation of a higher metastatic potential of skin and breast cancer cells with a hybrid epithelial-mesenchymal phenotype, defined by co-expression of epithelial and mesenchymal markers, compared to fully mesenchymal cells (Pastushenko et al., 2018).

On the one hand, all these findings suggest that at least certain aspects of EMT are required for metastatic progression. On the other hand, how epithelial, mesenchymal, or hybrid phenotypes develop at the molecular level and how they contribute exactly to different steps of the metastatic cascade remains largely unanswered.

1.4 Aim of the study

Twist1 is a basic helix-loop-helix transcription factor that is expressed in breast cancer (Martin, Goyal, Watkins, & Jiang, 2005) and that plays an important role in EMT and metastasis formation (Yang et al., 2004). Previously, a study using immortalised human mammary epithelial cells (HMLE) that express the EMT-transcription factor Twist1 fused to a modified oestrogen receptor (ER) domain (HMLE-Twist1-ER) has shown that upon activation of Twist1 by Tamoxifen treatment (+TAM), all cells underwent EMT, indicated by the trans-differentiation from an epithelial to a mesenchymal phenotype (Figure 1-4a, b) (Schmidt et al., 2015). However, after withdrawal of Twist1-activation by withdrawal of Tamoxifen, only a subset of cells underwent MET and reverted to an epithelial phenotype, whereas the other subset remained mesenchymal (Figure 1-4a, b) (Schmidt et al., 2015). Importantly, cells that had reverted showed an increased capacity for clonal growth in anchorage independence (Schmidt et al., 2015), indicating that a transient EMT-stimulus followed by a reversal of a mesenchymal phenotype back to an epithelial state is essential for cancer progression.
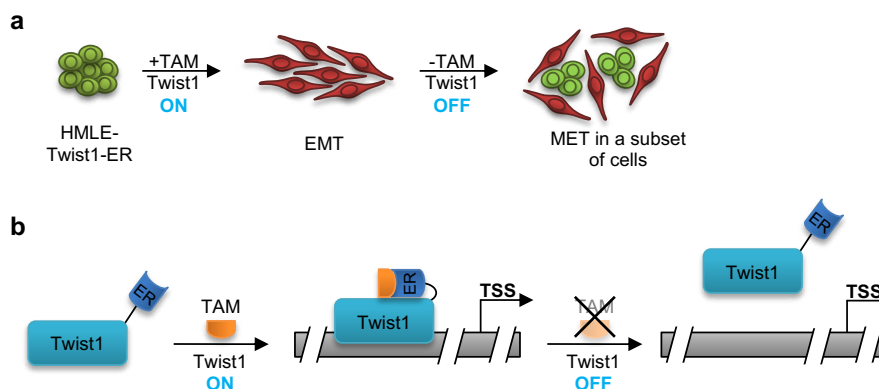
Figure 1-4: Overview of previous findings

a) The activation of Twist1 by Tamoxifen (+TAM) treatment in HMLE-Twist1-ER bulk cells induced trans-differentiation from epithelial (green) to mesenchymal (red) cells. Tamoxifen withdrawal (-TAM) and concomitant removal of Twist1-activation resulted in a mesenchymal-epithelial transition (MET) of a subset of cells (Schmidt et al., 2015). **b)** In cells that express Twist1 fused to a modified oestrogen receptor (ER) domain, Twist1 can be activated by Tamoxifen (TAM) treatment. Upon binding of Tamoxifen to the ER domain, Twist1-ER undergoes a conformational change and can act as a transcription factor. TAM withdrawal leads to loss of Twist1-activation and loss of its transcriptional activity. TSS: transcriptional start site.

In addition to these findings, it has been shown that EPCAM-positive (EPCAM[pos]) circulating tumour cells (CTCs) isolated from the blood of patients with luminal breast cancer can initiate metastatic tumours in mice (Baccelli et al., 2013). Moreover, EPCAM[pos] cells isolated from a breast cancer mouse model were found to metastasize to the bone more efficiently than EPCAM-negative (EPCAM[neg]) cells (Hiraga, Ito, & Nakamura, 2016). Additionally, experiments were performed with human metastatic breast cancer cells isolated from pleural effusions of

stage IV breast cancer patients and sorted for different EPCAM levels. These experiments have shown an increased tumour progression and metastasis rate to the lungs in EPCAM[high] cells compared to EPCAM[low] cells when injected into the mammary fat pad of immunodeficient mice (data obtained by M. Saini, pre-printed at Eichelberger et al., 2020).

Taken together, the findings of these studies suggest that cells that can regain an epithelial state at the site of metastasis are more efficient in colonising at distant sites.

In the present study, I aimed to investigate the differences between cells that can revert to an epithelial state and grow at distant sites and cells that remain in a mesenchymal state and have a reduced tumour progression capacity. Moreover, the goal of this study was to unravel which factors or mechanisms determine these different cell states.

As approximately 90% of all cancer deaths result from the development of distant metastases (Chaffer & Weinberg, 2011; Fidler, 2002), it is of utmost importance to better understand what kind of cells are capable of generating metastases at distant sites and which mechanisms promote the outgrowth of metastasised cells. Although treatment options have improved during the past years, the metastatic disease generally remains incurable. New findings could allow the development of treatments that target metastasised breast cancer cells at distant sites and increase patient survival.

2   Materials

2.1   Cell lines

Cell line	Supplier
HMLE-Twist1-ER cells	Robert A. Weinberg (Whitehead Institute, USA), chapter 3.1.1.1
HMLE-Twist1-ER CD24high	Johanna Bartsch (Helmholtz Centre Munich, Germany), chapter 3.1.1.1
HMLE-Twist1-ER single-cell clones	Johanna Bartsch (Helmholtz Centre Munich, Germany), chapter 3.1.1.2
HMLE-Twist1-ER M-SCC Twist1 motif knockout single-cell clones	Chapter 3.1.1.3
HMLE-Twist1-ER M-SCC *ZEB1* knockout single-cell clones	Chapter 3.1.1.4
HMLE-Twist1-ER E-SCC *ZEB1* or *GUS* overexpression single-cell clones	Chapter 3.1.1.5
HMLE-Twist1-ER M-SCC *ZEB1* or *GUS* overexpression single-cell clones	Chapter 3.1.1.5

2.2   Cell culture media and solutions

Product	Supplier (Catalogue No.)
Dulbecco's Modified Eagle Medium (DMEM)	Gibco®, Life Technologies™; Thermo Fisher Scientific (41965-039)
Fetal calf serum (FCS)	PAN-Biotech (30-2602)
Mammary Epithelial Cell Growth Medium	PromoCell (C-21010)
Mammary Epithelial Cell Growth Medium Supplement Mix	PromoCell (C-39115)
Penicillin/Streptomycin (10,000 U/ml)	Gibco®, Life Technologies™; Thermo Fisher Scientific (15140122)
Trypsin Neutralizing Solution (TNS)	PromoCell (C-41120)
Trypsin-EDTA 0.05%, 0.25%	Gibco®, Life Technologies™; Thermo Fisher Scientific (25300, 25200)

2.3   Chemicals and Reagents

Reagent	Supplier (Catalogue No.)
(Z)-4-Hydroxytamoxifen	Sigma (H7904)
1,7-Dichloro-octamethyltetrasiloxane	Santa Cruz (sc-229834)
2-Mercaptoethanol	Sigma (63689)
2-Propanol	Roth (9866)
Acetic acid (glacial)	Th.Geyer (2234)
Agarose, universal, peqGOLD	VWR (35-1020)
Agencourt AMPure XP	Beckman Coulter (A63880)
Albumin Fraction V (BSA)	Roth (8076)
Aluminium potassium sulphate	Sigma (A-7167)
Ammonium peroxydisulphate (APS)	Roth (9592)
Ampicillin sodium salt	Roth (K029)
Ampicillin sodium salt	Serva Electrophoresis (13399)
Aqua-Poly/Mount	Polysciences (18606)
Blasticidin S HCl (10 mg/ml)	Gibco®, Life Technologies™; Thermo Fisher Scientific (A1113903)
Bromophenol Blue	Sigma (B0126)
Buffer P1 (500 ml)	Qiagen (19051)
Buffer P2 (500 ml)	Qiagen (19052)
Buffer P3 (500 ml)	Qiagen (19053)
Carmine	Sigma (C-1022)
Collagen type I rat tail	Corning (354236)
DAPI	Sigma (D9542)
Deoxynucleotide (dNTP) Solution Mix	NEB (N0447)
Dimethyl sulfoxide (DMSO)	Sigma (D8418)
DNAzol™ Reagent	Invitrogen™, Thermo Fisher Scientific (10503027)
Doxycycline Hydrochloride	Sigma (D3072)
Ethanol	Merck (100983)
Ethylenediaminetetraacetic acid (EDTA)	Roth (X986)
Gel Loading Dye, Orange (6X)	New England Biolabs (B7022S)
Glycerol	Roth (3783)
Glycine	Roth (3790)
Hydrochloric acid 32% (HCl)	Merck (100319)
Kanamycin sulphate	Sigma (60615)

Materials

Reagent	Supplier (Catalogue No.)
LB Broth (Lennox)	Roth (X964)
LB Broth with agar (Lennox)	Sigma (L2897)
Methanol	Roth (4627)
N-2-hydroxyethyl piperazine-N-2-ethane sulfonic acid, 1M (HEPES)	Gibco®, Life Technologies™; Thermo Fisher Scientific (15630)
n-heptane	Applichem (1948)
non-fat dried milk powder	Roth (T145)
Nonidet® P40 (NP40)	Applichem (A1694)
Normal goat serum	GeneTex (GTX73206)
PageRuler™ Plus protein ladder	Thermo Fisher Scientific (26619)
Paraformaldehyde 16% (w/v) (PFA)	VWR International (43368.9M)
Phosphatase Inhibitor Cocktail 2	Sigma (P5726)
Phosphatase Inhibitor Cocktail 3	Sigma (P0044)
Phosphate Buffered Saline (PBS) pH 7.4	Gibco®, Life Technologies™; Thermo Fisher Scientific (10010-015)
Protease Inhibitor Cocktail	Sigma (P8340)
Puromycin Dihydrochloride	Gibco®, Life Technologies™; Thermo Fisher Scientific (A1113803)
Rotiphorese® Gel 30	Roth (3029)
S.O.C. Medium	Invitrogen™, Thermo Fisher Scientific (15544034)
Sodium chloride (NaCl)	Roth (P029)
Sodium Deoxycholate	Applichem (A1531)
Sodium dodecyl sulfate (SDS)	Roth (4360)
Sodium hydroxide (NaOH)	Merck (30620-M)
SYBR™ Safe DNA Gel Stain	Invitrogen™, Thermo Fisher Scientific (S33102)
TE Buffer	Invitrogen™, Thermo Fisher Scientific (12090015)
TEMED	Roth (2367)
Thymol	Sigma (T0501)
TransIT-X2® dynamic delivery system	Mirus Bio LLC (MIR 6003)
TransIT-X2® Transfection Reagent	Mirus Bio LLC (MIR 6003)
TriDye™ 1 kb Plus DNA Ladder	New England Biolabs (N3270S)
Tris (Trizma base)	Sigma (T1503)

Reagent	Supplier (Catalogue No.)
Triton X-100	Sigma (T8787)
TWEEN® 20	Sigma (P1379)
UltraPure™ DNase/RNase-Free Distilled Water	Thermo Fisher Scientific (10977035)
Water (sterile-filtered)	Sigma (W3500)

2.4 Kits

Kit	Supplier (Catalogue No.)
Amersham™ ECL™ Prime Western Blotting Reagent Kit	GE Healthcare (RPN2232)
DC™ Protein Assay Kit	Bio-Rad Laboratories (5000112)
Dead Cell Removal Kit	Miltenyi Biotec (130-090-101)
DNA Clean & Concentrator™-5 Kit	Zymo Research (D4014)
Hemacolor® Rapid Staining Kit	Merck (111661)
miRNeasy Mini Kit	Qiagen (217004)
miScript II RT Kit	Qiagen (218161)
miScript SYBR® Green PCR Kit	Qiagen (218073)
Mix2Seq Kit	Eurofins (3094-000MSK)
Monarch® DNA Gel Extraction Kit	New England Biolabs (T1020)
NE-PER™ Nuclear and Cytoplasmic Extraction Reagents Kit	Thermo Fisher Scientific (78833)
OneScript® Plus cDNA Synthesis Kit	Applied Biological Materials (G236)
Plasmid Midi Kit	Qiagen (12143)
RNase-Free DNase Set	Qiagen (79254)
RNeasy Mini Kit	Qiagen (74106)

2.5 Bacterial strains

Bacterial Strain	Supplier (Catalogue No.)
MAX Efficiency™ Stbl2™ Competent Cells	Invitrogen™, Thermo Fisher Scientific (10268019)
XL10-Gold Ultracompetent Cells	Agilent Technologies (200314)

2.6 Plasmids

Plasmid	Source
PB-TAC-ERP2	Addgene #80478
pcDNA3.1+/C-(K)DYK with *ZEB1* (NM_001128128.2) cDNA ORF clone OHu27170C	GenScript®
pCMV-hyPBase	Kindly provided by Roland Rad
pENTR1A-no ccdb	Addgene #17398
pSpCas9(BB)-2A-GFP (PX458)	Kindly provided by Christopher Breunig and Stefan Stricker (Addgene #48138)
STAgR_gRNAScaffold_hU6	Kindly provided by Christopher Breunig and Stefan Stricker (Addgene #102840)
StagR-neo	Kindly provided by Christopher Breunig and Stefan Stricker (Addgene #102992)

2.7 Enzymes and enzyme mixes

Enzyme/Enzyme Mix	Supplier (Catalogue No.)
All DNA digestion enzymes	New England Biolabs
Clonase™ Gateway™ LR Clonase II Enzym-Mix	Invitrogen™, Thermo Fisher Scientific (11791020)
Gibson Assembly® Master Mix	New England Biolabs (E2611)
Phusion® High-Fidelity DNA Polymerase	New England Biolabs (M0530)
Power SYBR™ Green PCR Master Mix	Applied Biosystems (4367659)
Q5® High-Fidelity DNA Polymerase	New England Biolabs (M0202S)
T4 DNA ligase	New England Biolabs (M0267S)
Taq DNA polymerase	New England Biolabs (M0491S)

2.8 Guide RNAs (gRNAs) used for CRISPR/Cas9

Label	Target region	gRNA sequence (5' – 3')
Twist1 motif gRNA1	Twist1 motif (*ZEB1*)	ATGTTGATCGCCAGAGAAAG
Twist1 motif gRNA2	Twist1 motif (*ZEB1*)	GCAGCTGCACAGACGCTGGG
Twist1 motif gRNA3	Twist1 motif (*ZEB1*)	ATCAACAGATGACTTAAGGG
ZEB1 gRNA1	*ZEB1* Exon 5	GCCTCTATCACAATATGGAC

Label	Target region	gRNA sequence (5' – 3')
ZEB1 gRNA2	*ZEB1* Exon 5	ACAACTCAGCCCTCAATGGA
ZEB1 gRNA3	*ZEB1* Exon 7	AGTTCTGTCACAAGCATGCA
ZEB1 gRNA4	*ZEB1* Exon 7	TTGCCGTATCTGTGGTCGTG

2.9 Primer pairs for PCRs on gDNA of CRISPR/Cas9 cells

Target region	Orientation	Sequence (5' – 3')	Annealing temperature
Twist1 motif	forward	AGCAAGGAGTGGAGCATAG	63°C
	reverse	ATACAAGGACAGAGCCAAAG	
ZEB1 Exon 5	forward	GCATAGGGACTCAGTGGAAACT	67°C
	reverse	AGGAGGCAACTCCCTTTACTAC	
ZEB1 Exon7	forward	GGTCGGTGAAATGGGATAAGAAAAA	67°C
	reverse	ACCACCAGTGAAAACCCCATT	

2.10 Primer pairs for qPCRs on mRNAs

Gene	Orientation	Sequence (5' – 3')
CDH1	forward	TGCCCAGAAAATGAAAAAGG
	reverse	GTGTATGTGGCAATGCGTTC
EPCAM	forward	ATAACCTGCTCTGAGCGAGTG
	reverse	TGCAGTCCGCAAACTTTTACTA
FN1	forward	CAGTGGGAGACCTCGAGAAG
	reverse	TCCCTCGGAACATCAGAAAC
GRHL2	forward	GAAAGTCCAGTTTCACCAGAGG
	reverse	GGCACTAAGGCCACTAGTCTTTT
mCherry	forward	GAACGGCCACGAGTTCGAGA
	reverse	CTTGGAGCCGTACATGAACTGAGG
mTwist1	forward	GGACAAGCTGAGCAAGATTCA
	reverse	CGGAGAAGGCGTAGCTGAG
OVOL2	forward	ACAGGCATTCGTCCCTACAAA
	reverse	CGCTGCTTATAGGCATACTGC
RPL32	forward	CAGGGTTCGTAGAAGATTCAAGGG
	reverse	CTTGGAGGAAACATTGTGAGCGATC

Gene	Orientation	Sequence (5' – 3')
VIM	forward	GAGAACTTTGCCGTTGAAGC
	reverse	GCTTCCTGTAGGTGGCAATC
WNT5A	forward	ATGGCTGGAAGTGCAATGTCT
	reverse	ATACCTAGCGACCACCAAGAA
ZEB1	forward	GCACAAGAAGAGCCACAAGTAG
	reverse	GCAAGACAAGTTCAAGGGTTC

2.11 Primer Assays for qPCR on miRNAs

Gene	MiScript Primer Assay	Supplier
MIR141	HS-miR-141_1	Qiagen
MIR200A	HS-miR-200a_1	Qiagen
MIR200B	HS-miR-200b_3	Qiagen
MIR200C	HS-miR200c_1	Qiagen
RNU6-2	HS-RNU6-2_11	Qiagen

2.12 Flow cytometry dyes and antibodies

Dye	Supplier (Catalogue No.)	Laser (nm)	Filter (nm)	Dilution
7-AAD	BD (559925)	488	695/40	1:25
EPCAM-FITC (VU-1D9)	Biozol (GTX79849)	488	530/30 I	1:500
SYTOX™ Blue Dead Cell Stain	Invitrogen™, Thermo Fisher Scientific (S34857)	405	450/40	1:125

2.13 Primary antibodies

Epitope (Clone)	Host	Supplier (Catalogue No.)	Application (Dilution)
alpha-Tubulin (B-5-1-2)	Mouse	Sigma (T5168)	WB (1:5000)
beta-actin (AC-15)	Mouse	Abcam (ab6276)	WB (1:6000)
E-cadherin (EP700Y)	Rabbit	Biozol (GTX61329)	WB (1:25000)
E-cadherin-Alexa Fluor-488 (24E10)	Rabbit	New England (3199)	IF 3D (1:50)
Histone H3	Rabbit	Abcam (ab1791)	WB (1:5000)

Epitope (Clone)	Host	Supplier (Catalogue No.)	Application (Dilution)
Twist1 (Twist2C1a)	Mouse	Santa Cruz (sc-81417)	WB (1:200)
Vimentin (V9)	Mouse	Abnova (MAB3578)	WB (1:1000), IF (1:100)
ZEB1	Rabbit	Sigma (HPA027524)	WB (1:5000), IF (1:500)
ZEB1 (H-102)	Rabbit	Santa Cruz (sc-25388)	WB (1:200)

2.14 Secondary antibodies

Host	Epitope	Conjugation	Supplier (Catalog No.)	Application (Dilution)
Goat	Mouse IgG	Alexa Fluor 546	Life Technologies (A11030)	IF (1:250)
Goat	Mouse IgG	HRP	Jackson ImmunoResearch (115-036-062)	WB (1:12500)
Goat	Rabbit IgG	Alexa Fluor 488	Life Technologies (A11034)	IF (1:250)
Goat	Rabbit IgG	HRP	Jackson ImmunoResearch (111-036-045)	WB (1:12500)

2.15 Consumables

Product	Supplier
15 mL Conical Centrifuge Tubes	Corning
384 PCR Plates	Kisker Biotech
50 mL Conical Centrifuge Tubes	Corning
6-, 12-, 24-, 96-well plates (Falcon)	Becton Dickinson
Cell culture flasks (T75, T175)	Greiner Bio-One
Cell Scraper, UltraCruz® Cell Lifter	Santa Cruz
CellCarrier™-96 Black microplates	PerkinElmer
Cotton swabs	Roth
Falcon Round-Bottom Polypropylene Tubes	Corning
Falcon® Permeable Support for 24 well plates, 8 µm pore (Inserts)	Corning
KOVA™ Glasstic™ Slide 10 with Grids	Kova International
Millex 33mm Durapore PVDF 0.22 µm	Merck
Millex 33mm Durapore PVDF 0.45 µm	Merck

Product	Supplier
Nitril® NextGen® Gloves	Meditrade
Nunc® CryoTubes®	Thermo Fisher Scientific
Nunc™ Cell Culture Dishes (6-cm)	Thermo Fisher Scientific
PARAFILM® M	Bemis
PCR Tubes, 0.5 mL, PCR clean	Eppendorf
Petri dish, polystyrene (for Agar plates)	Greiner Bio-One
Pipette tips (filtered)	Starlab
Pipette tips (unfiltered)	Starlab
PVDF blotting membrane	GE Healthcare
QIAshredder	Qiagen
Reagent reservoirs	Corning
Rotilabo® Blotting papers, thick 1.5 mm, 580x600 mm	Roth
Safe-Lock Tubes 1.5ml PCRclean	Eppendorf
Safe-Lock Tubes 2.0ml	Eppendorf
Scalpels	VWR International
Sterile Serological Pipets (2, 5, 10, 25, 50 ml)	Greiner Bio-One
TC-treated Cell Culture Dish (10-cm)	Corning
Test Tube with Cell Strainer Snap Cap (FACS tube)	Corning
Ultra - Clear - Film for RT - PCR	Kisker Biotech
X960 PCR Tube Strips Eppendorf 8 Tubes (0.2 ml)	Eppendorf

2.16 Instruments

Instrument	Manufacturer
2D Rocker	IKA®
Blue light table, LED transilluminator	Serva Electrophoresis
ChemiDoc™ MP Imaging System	Bio-Rad Laboratories
Eppendorf Mastercycler® nexus gradient	Eppendorf AG
FACSAria IIIu	Becton Dickinson
FLUOVIEW FV1200 inverted Confocal Laser Scanning Microscope	Olympus
HERAcell® 240i incubator	Thermo Fisher Scientific
Heraeus™ Fresco™ 21	Thermo Fisher Scientific
Heraeus™ Megafuge™ 40R centrifuge	Thermo Fisher Scientific

Instrument	Manufacturer
Heraeus™ Pico™ 21	Thermo Fisher Scientific
Horizontal Roller	CAT
iMark Microplate Absorbance Reader	Bio-Rad Laboratories
Leica DFC 450 C stereomicroscope	Leica Microsystems
Leica DM IL LED light microscope	Leica Microsystems
Mini Trans-Blot® Electrophoretic Transfer Cell	Bio-Rad Laboratories
Mini-PROTEAN® Tetra System	Bio-Rad Laboratories
Mr. Frosty™ Freezing Container	Thermo Fisher Scientific
NanoDrop® ND 1000 Spectrophotometer	Thermo Fisher Scientific
Neubauer chamber glas improved	Heinz Herenz
Orbital Multi Bio 3D Shaker	Biosan
PerfectBlue™ gel system	Peqlab
PowerPac™ HC	Bio-Rad Laboratories
QuantStudio™ 12K Flex qPCR System	Applied Biosystems™, Thermo Fisher Scientific
Safe 2020 Class II Biological Safety Cabinet	Thermo Fisher Scientific
SteREO Lumar V.12	Carl Zeiss Microscopy
ThermoMixer® C 1.5 ml and 0.5 ml	Eppendorf
Water Bath	Memmert/Karl Hecht
Zeiss Axio Imager M.2 with Colibri 7	Carl Zeiss Microscopy

2.17 Software

Product	Manufacturer
FACS Diva™ v6.1.3	Becton Dickinson
FlowJo® V10	FlowJo, LLC
FV10-ASW software Ver 4.2.b	Olympus
GIMP 2.10	The GIMP Team
GraphPad Prism 7	GraphPad Software
Image Lab 5.2.1	Bio-Rad Laboratories
ImageJ 1.52	NIH
Leica Application Suite V4	Leica Microsystems
MS Office 2016	Microsoft Corporation
Panorama Stitcher Mini Version 1.10	Alexander Boltnev, Olga Kacher

Product	Manufacturer
QuantStudio™ 12K Flex Real-Time PCR System 1.2.2	Applied Biosystems™, Thermo Fisher Scientific
ZEN Pro 2.3	Carl Zeiss Microscopy

2.18 Data availability

The datasets that support the findings of this work are deposited in the Gene Expression Omnibus (GEO) under the accession codes GSE138329 and GSE156024 and can be accessed upon request unless publicly available.

3 Methods

3.1 Cell culture methods

3.1.1 Generation of cell lines

3.1.1.1 HMLE-Twist1-ER cells

HMLE-Twist1-ER bulk cells were kindly gifted by Robert A. Weinberg (Whitehead Institute, United States). These cells were generated as previously described (Mani et al., 2008; Yang et al., 2004). Briefly, to first generate immortalised human mammary epithelial cells (HMLE), primary human mammary epithelial cells were transduced with virus containing the telomerase catalytic subunit of the human telomerase reverse transcriptase (hTERT) gene to induce the telomerase activity. Additionally, these cells were transduced with viruses containing the simian virus 40 (SV40) large T early region, which binds tumour protein 53 (TP53), retinoblastoma (RB) and protein phosphatase 2 (PP2), and thereby allows constant progression of the cell cycle (Elenbaas et al., 2001). These cells do not express the oestrogen receptor (ER) and are not tumorigenic (Ince et al., 2007). To generate HMLE-Twist1-ER cells, HMLE cells were transduced with a pWZL-Blast-Twist1-ER vector and subsequently propagated in growth medium containing Blasticidin S HCl at a final concentration of 10 µg/ml. HMLE-Twist1-ER cells constitutively overexpress Twist1 fused to a modified oestrogen ligand binding-domain (ER). By adding (Z)-4-Hydroxytamoxifen (TAM) at a final concentration of 20 nM, Twist1 can be activated and act as a transcription factor (Casas et al., 2011). The CD24^{high} population used within this study was derived by FACS-sorting and was kindly provided by Johanna Bartsch (Helmholtz Centre Munich, Germany).

3.1.1.2 HMLE-Twist1-ER single-cell clones

HMLE-Twist1-ER (CD24^{high}) single-cell clones (SCCs) were isolated by Johanna Bartsch (Helmholtz Centre Munich, Germany). Briefly, HMLE-Twist1-ER (CD24^{high}) bulk cells were seeded to a 96-well plate (0.3 cells/well). Wells were checked visually for single cells. Single-cell colonies were further expanded and passaged.

3.1.1.3 HMLE-Twist1-ER M-SCC Twist1 motif knockout single-cell clones

To generate HMLE-Twist1-ER M-SCC Twist1 motif knockout single-cell clones, an HMLE-Twist1-ER M-SCC was genetically modified by CRISPR/Cas9. For this purpose, cells were transfected with the previously generated Cas9-GFP plasmid (pSpCas9(BB)-2A-GFP) and the gRNA containing STAgR-neo plasmid (chapter 3.2.5, Table 3-1).

Table 3-1: Composition of the transfection mix to obtain Twsit1 motif knockout cells

Molar ratio pSpCas9(BB)-2A-GFP:gRNA plasmid = 1:3	
Compound	**Amount**
pSpCas9(BB)-2A-GFP	1 µg
STAgR-neo plasmid containing 2 *ZEB1* gRNAs	1.341 µg
TransIT-X2® Transfection Reagent	7.5 µl
DMEM without supplements	to 200 µl

Single-cell clones were isolated by seeding GFP^{pos} sorted cells into a 96-well plate (0.5 cells/well). Wells were checked visually for single cells. Single-cell colonies were further expanded and passaged.

3.1.1.4 HMLE-Twist1-ER M-SCC *ZEB1* knockout single-cell clones

To generate HMLE-Twist1-ER M-SCC *ZEB1* knockout single-cell clones, an HMLE-Twist1-ER M-SCC was genetically modified by CRISPR/Cas9. For this purpose, cells were transfected with the previously generated Cas9-GFP plasmid (pSpCas9(BB)-2A-GFP) and the gRNA containing STAgR-neo plasmid (chapter 3.2.5, Table 3-2). As a control, cells were transfected with the Cas9-GFP plasmid (pSpCas9(BB)-2A-GFP) and a plasmid harbouring the guide RNA scaffold and termination sequence (Table 3-3).

Table 3-2: Composition of the transfection mix to obtain *ZEB1* knockout cells

Molar ratio pSpCas9(BB)-2A-GFP:gRNA plasmid = 1:3	
Compound	**Amount**
pSpCas9(BB)-2A-GFP	1 µg
STAgR-neo plasmid containing 4 *ZEB1* gRNAs	1.55 µg
TransIT-X2® Transfection Reagent	7.5 µl
DMEM without supplements	to 200 µl

Table 3-3: Composition of the transfection mix to obtain knockout control cells

Molar ratio pSpCas9(BB)-2A-GFP:STAgR-neo plasmid = 1:3	
Compound	**Amount**
pSpCas9(BB)-2A-GFP	1 µg
STAgR-neo plasmid without coding gRNAs	1.19 µg
TransIT-X2® Transfection Reagent	7.5 µl
DMEM without supplements	to 200 µl

After 2.5 days, cells were sorted for GFP expression, and GFP[pos] sorted cells were seeded into a 96-well plate (0.5 cells/well) to obtain single-cell clones. Wells were checked by eye for single cells. Single-cell colonies were further passaged and expanded.

3.1.1.5 HMLE-Twist1-ER E-SCC and M-SCC *ZEB1* overexpression single-cell clones

To generate HMLE-Twist1-ER E-SCC and M-SCC *ZEB1* overexpression single-cell clones, an HMLE-Twist1-ER M-SCC and an E-SCC were genetically modified by PiggyBac. For this purpose, cells were transfected (chapter 3.4.1) with the pCMV-hyPBase plasmid and the previously generated (chapter 3.2.10) and linearized PB-*ZEB1* plasmid (Table 3-4). As a control, cells were transfected with the pCMV-hyPBase plasmid and the linearized PB-*GUS* plasmid (Table 3-5).

Table 3-4: Composition of the transfection mix to obtain *ZEB1* overexpression cells

Mass ratio pCMV-hyPBase:linearized PB-*ZEB1* = 1:8	
Compound	**Amount**
pCMV-hyPBase	1.65 µg
linearized PB-*ZEB1*	13.2 µg
TransIT-X2® Transfection Reagent	45 µl
DMEM without supplements	to 1 ml

Table 3-5: Composition of the transfection mix to obtain *GUS* overexpression cells

Mass ratio pCMV-hyPBase:linearized PB-*GUS* = 1:8	
Compound	**Amount**
pCMV-hyPBase	1.65 µg
linearized PB-*GUS*	13.2 µg
TransIT-X2® Transfection Reagent	45 µl
DMEM without supplements	to 1 ml

Single-cell clones were isolated after selection with medium containing 1 µg/ml Puromycin Dihydrochloride by seeding puromycin resistant cells into a 96-well plate (0.5 cells/well). Wells were checked by eye for single cells. Single-cell colonies were further passaged and expanded.

3.1.2 Cultivation of cells

All cell culture related experiments were carried out in a sterile biosafety cabinet. All cells were propagated in cell culture dishes in a 37°C HERAcell 240i incubator with a humidified atmosphere of 5% CO_2 and ambient O_2 levels. HMLE-Twist1-ER (CD24^{high}) cells and SCCs and HMLE-Twist1-ER M-SCC *ZEB1* or Twist1 motif knockout SCCs were cultured in Mammary Epithelial Cell Growth Medium (MECGM) supplemented with the Mammary Epithelial Cell Growth Medium Supplement Mix (0.004 ml/ml bovine pituitary extract (BPE), 10 ng/ml epidermal growth factor (EGF), 5 µg/ml Insulin and 0.5 µg/ml hydrocortisone), 1% penicillin/streptomycin, and Blasticidin S HCl at a final concentration of 10 µg/ml. For Twist1-activation, cells were treated with (Z)-4-Hydroxytamoxifen (TAM) at a final concentration of 20 nM for the indicated number of days. HMLE-Twist1-ER E-SCC and M-SCC *ZEB1* and *GUS* overexpression SCCs were propagated in medium further supplemented with Puromycin Dihydrochloride at a final concentration of 1 µg/ml. For induction, *ZEB1* and *GUS* overexpression SCCs were treated with 1 µg/ml doxycycline for the indicated number of days. The medium was changed every two to three days.

3.1.3 Passaging of cells

Cells were passaged at a confluency of 70 to 80%. To do so, the medium was aspirated, and the cells were washed with PBS once. To detach cells, 0.15% Trypsin-EDTA was added, followed by an incubation step at 37°C. Trypsinization was stopped by adding three volumes of Trypsin neutralizing solution (TNS). After transferring the cell suspension to a falcon tube, cells were pelleted by centrifugation in a Heraeus Megafuge at 490 rcf at 4°C for 5 minutes. The supernatant was aspirated, cells were resuspended in their respective growth medium, and 1/4 to 1/10 of the cell suspension was transferred to a new cell culture dish containing the appropriate medium.

3.1.4 Cell counting

For cell counting, 10 µl of the cell suspension were transferred to a Neubauer counting chamber. Cells were then counted using a 10x objective of a Leica DM IL LED light microscope. The average number of cells per large square was determined and multiplied by 10^4 to determine the number of cells per ml.

3.1.5 Thawing of cells

Frozen cells were quickly thawed in a 37°C water bath. The cell suspension was then transferred to a falcon containing 3 to 5 ml PBS and centrifuged in a Heraeus Megafuge at 490 rcf at 4°C for 5 minutes. After removing the supernatant, the cell pellet was resuspended in Mammary Epithelial Cell Basal Medium supplemented with the Mammary Epithelial Cell Growth Medium Supplement Mix and 1% penicillin/streptomycin and transferred to a cell culture dish. The next day, cells were washed with PBS once, and the cell line-appropriate growth medium was added.

3.1.6 Freezing of cells

For freezing of cells, the cell suspension was pelleted in a Heraeus Megafuge at 490 rcf at 4°C for 5 minutes. After removing the supernatant, the cell pellet was resuspended in ice-cold freezing medium and transferred to a Nunc® CryoTube®. Tubes were transferred to a pre-cooled freezing container and frozen at -80°C. For long-term storage, tubes containing frozen cells were transferred to liquid nitrogen.

Freezing medium: 50% FCS, 40% Mammary Epithelial Cell Basal Medium supplemented with the Mammary Epithelial Cell Growth Medium Supplement Mix and 1% penicillin/streptomycin, 10% DMSO

3.1.7 Migration Assay

To investigate the migratory capacity of cells, 2.5×10^4 cells were seeded into 24-well culture inserts with 8 µm pores and incubated at 37°C in a HERAcell 240i incubator with a humidified atmosphere of 5% CO_2 and ambient O_2 levels. After 24 hours, the medium was carefully aspirated, and non-migrated cells were carefully removed from the upper side of the insert with a cotton swab. Cells on the bottom of the insert were then fixed and stained with the Hemacolor® Rapid Staining Kit according to the manufacturer's instructions. Briefly, the insert was carefully transferred to the fixation solution for 30 sec. After a washing step in Milli-Q water, the insert was transferred to staining solution 1 for 2 min. Next, the insert was rewashed with Milli-Q water and transferred to staining solution 2 for 2 min. After another washing step, the insert was transferred to Hemacolor® buffer for at least 2 min. Then the insert was taken out, and the residual liquid was carefully removed using a cotton swab. The insert was then transferred into a fresh 24-well plate to dry at room temperature overnight. The next day migrated cells were counted.

3.1.8 Cultivation in floating 3D-collagen gels

To model cell growth behaviour in a three-dimensional environment, cells were cultivated in floating 3D-collagen I gels, as previously described (Linnemann et al., 2015) with some minor modifications. Briefly, 24-well plates were coated with a siloxane coating solution for 30 sec, followed by three wash steps with PBS and one wash step with Milli-Q water. After trypsinization, 300 cells were seeded per gel by mixing the cell suspension with neutralizing solution and collagen I (final collagen I concentration: 1.3 mg/ml). 400 µl of the prepared mixture were transferred per previously coated 24-well. To allow polymerization, plates were transferred into a 37°C HERAcell 240i incubator with a humidified atmosphere of 5% CO_2 and ambient O_2 levels for one hour. 600 µl of the desired medium with respective concentrations of compounds (final volume including collagen gel: 1 ml) were added. 3D-collagen gels were detached by surrounding the gels with a thin pipette tip followed by careful shaking of the 24-well plate until gels were detached entirely. The medium was changed every two to three days, and cells were cultured within 3D-collagen gels for ten days.

Siloxane coating solution: n-Heptane, 25 g/l 1,7-Dichlorooctamethyltetrasiloxane
Neutralizing Solution: 500 mM HEPES in PBS

3.2 Methods for working with DNA

3.2.1 Restriction digest

To obtain specific DNA fragments from plasmids or to validate correct cloning or reproduction of plasmids, restriction enzymes from New England Biolabs were used. A total reaction volume of 20 to 50 µl was prepared according to the manufacturer's protocol of each restriction enzyme. Digests were performed according to the manufacturer's protocol.

3.2.2 Ligation of DNA fragments with T4 DNA ligase

To ligate purified DNA fragments, T4 DNA ligase was used according to the manufacturer's instructions. A molar ratio of 1:3 to 1:10 (backbone:insert) was used. The required mass of the insert was calculated using the NEBioCalculator for DNA ligation (*http://nebiocalculator.neb.com/#!/ligation*). The ligation reaction was set up, gently mixed and incubated at 4°C in a ThermoMixer® for 16 h. To amplify the ligated DNA construct, 25 µl of chemically competent bacteria were transformed with 4 µl of the ligation reaction as described in chapter 3.3.

Formula ligation calculator: required mass insert (g) = desired insert/vector molar ratio x mass of vector (g) x ratio of insert to vector lengths

3.2.3 Agarose gel electrophoresis

To prepare agarose gels, 1x TAE buffer containing 0.8 to 1.2% of agarose powder was microwaved until agarose powder was completely dissolved. After SYBR™ Safe DNA stain was added at a ratio of 1:20000, the solution was poured into an agarose gel chamber followed by immediate application of a gel comb. After curing, agarose gels were transferred into a PerfectBlue™ gel system, 1x TAE buffer was added, and combs were removed carefully. To separate DNA fragments, e.g. PCR products, based on their base-pair length, DNA samples were mixed with 6x orange loading dye and loaded into the agarose gel pockets for separation. To determine the size of the fragments, a TriDye™ 1 kb Plus DNA Ladder was loaded into an agarose gel pocket. DNA fragments were separated at 120 V for 30-60 min. For analysis, agarose gels were imaged on a ChemiDoc™ MP Imaging System using the Image Lab software. For the extraction of specific DNA fragments from the gel, DNA fragments within agarose gels were visualized with a blue light table, and fragments were isolated as described in chapter 3.2.4.

10x TAE: 0.4 M Tris, 0.2 M acetic acid, 10 mM EDTA

3.2.4 Extraction of DNA fragments from agarose gels

To visualize DNA fragments separated by agarose gel electrophoresis (chapter 3.2.3), the agarose gel was placed on a blue light table. The desired fragment was cut out using a scalpel and transferred to a 1.5 ml reaction tube. Isolation of the DNA fragment was performed with the Monarch® DNA Gel Extraction Kit according to the manufacturer's instructions. For elution, 15 µl of pre-heated (50°C) elution buffer were used. DNA concentrations were measured with the NanoDrop® ND 1000 Spectrophotometer. DNA was stored at -20°C.

3.2.5 STAgR cloning of two to four gRNAs

To generate plasmids containing two to four guide RNAs (gRNAs) targeting specific genomic regions, string assembly gRNA cloning (STAgR) was used and performed as described (Breunig et al., 2018). To enable deletion of several base pairs, two gRNAs flanking the genomic region of interest were used. DNA exons and introns of specific genes were obtained from the Ensembl genome browser (https://www.ensembl.org/index.html). gRNAs were designed using Benchling (https://benchling.com/) and the MIT CRISPR tool (http://crispr.mit.edu/).

To amplify the backbone fragment according to the protocol by Breunig and colleagues, the hU6 promoter containing STAgR-neo backbone vector and the following primers were used:

STAgR backbone fragment:

Forward primer: 5'-$N_{20(\text{last gRNA sense})}$GTTTTAGAGCTAGAAATAGCAAGTT-3'

Reverse primer: 5'-$N_{20(\text{first gRNA antisense})}$CGGTGTTTCGTCCTTT-3'

To digest methylated template plasmid, the backbone PCR reaction was digested using DpnI according to the manufacturer's instructions for at least 1h at 37°C.

To amplify STAgR gRNA fragments according to the published protocol, the STAgR_gRNAscaffold_hU6 plasmid and the following primers were used:

1st STAgR gRNA fragment:

Forward primer: 5'-$N_{20(\text{first gRNA sense})}$GTTTTAGAGCTAGAAATAGCAAGTT-3'

Reverse primer: 5'-$N_{20(\text{second gRNA antisense})}$CGGTGTTTCGTCCTTT-3'

2nd STAgR gRNA fragment:

Forward primer: 5'-$N_{20(\text{second gRNA sense})}$GTTTTAGAGCTAGAAATAGCAAGTT-3'

Reverse primer: 5'-$N_{20(\text{third gRNA antisense})}$CGGTGTTTCGTCCTTT-3'

3rd STAgR gRNA fragment:

Forward primer: 5'-$N_{20(\text{third gRNA sense})}$GTTTTAGAGCTAGAAATAGCAAGTT-3'

Reverse primer: 5'-$N_{20(\text{last gRNA antisense})}$CGGTGTTTCGTCCTTT-3'

DpnI-digested backbone fragment and STAgR gRNA fragments were loaded on a 1% agarose gel and purified by gel extraction. As all fragments contained overlapping sequences, Gibson Assembly (chapter 3.2.6) was applied to obtain a single vector. To identify bacterial colonies containing the correct plasmid size, colony PCR was performed (chapter 3.2.7). Colonies harbouring plasmids with the correct size were used for further expansion (chapter 3.3.2). To validate the correctness of the DNA sequence, plasmids were sequenced by Eurofins using the Mix2SeqKit according to the manufacturer's instructions.

Sequencing primers:

Forward primer 1: 5'-GAGTTAGGGGCGGGACTATG-3'

Forward primer 2: 5'-ACTGGATCCGGTACCAAGG-3'

Reverse primer: 5'-TTACGGTTCCTGGCCTTTTG-3'

3.2.6 Gibson assembly of DNA fragments

Gibson assembly was used to assemble DNA fragments with overlapping ends. Gibson assembly was used to assemble a vector backbone and two to four insert fragments (chapter 3.2.5). The Gibson assembly reaction was prepared using the Gibson Assembly Master Mix according to the manufacturer's instructions. 50 to 100 ng of purified vector backbone PCR product and two to four insert fragments (STAgR pieces) were mixed at a ratio of 1:3 (backbone:insert). The Gibson assembly reaction was incubated at 50°C for 60 min. For transformation of 25 µl chemically competent bacteria (chapter 3.3), the Gibson assembly

reaction was diluted 1:4 in nuclease-free water. 2 µl of the diluted Gibson reaction were used for transformation of chemically competent bacteria.

3.2.7 Colony PCR

To screen bacterial colonies obtained from Gibson assembly (chapter 3.2.6), single bacterial colonies were analysed as described (Breunig et al., 2018). Briefly, a single bacterial colony was picked with a pipette tip and smeared into a 0.2 ml PCR reaction tube followed by inoculation of 1 ml of LB medium containing 100 µg/ml Ampicillin. 10 µl of a PCR mix containing Taq polymerase and specific primers was added to each reaction tube and PCR reaction run according to the published protocol (Breunig et al., 2018). PCR products were analysed on a 1% agarose gel. Inoculated colonies (1 ml) harbouring plasmids with the correct length were used for further expansion (chapter 3.3.2).

Primers used for colony PCR:

STAgR_seq_fwd2: ACTGGATCCGGTACCAAGG

STAgR_seq_rev: TTACGGTTCCTGGCCTTTTG

Expected PCR products containing sgRNAs: hU6 265 bp, gRNA scaffold 83 bp, 2 sgRNAs ~740 bp, 4 sgRNAs ~1.5 kb

3.2.8 Isolation of genomic DNA

To isolate genomic DNA (gDNA) from cells, cells were cultivated in 12-well plates. At a confluency of at least 60%, cells were washed with 1 ml PBS followed by lysis with 250 µl DNAzol™ per well. Plates were either sealed with parafilm and stored at 4°C for one to two days, or isolation was performed immediately. Isolation was performed according to the manufacturer's instructions, with minor modifications. Briefly, per 250 µl DNAzol™ reagent, 150 µl of 100% ultra-pure ethanol were added. After 5 min incubation time, precipitated gDNA was transferred into a 1.5 ml reaction tube containing 70% ethanol. After a centrifugation step at 18000 rcf for 1 min, the supernatant was aspirated, and the gDNA pellet was washed once again with 500 µl of 70% ethanol. After removing the supernatant again, the gDNA pellet was dried by either placing the open tube on room temperature or on an 85°C ThermoMixer® until the remaining ethanol was evaporated. Next, gDNA was solubilized in 30-50 µl of 8 mM NaOH and adjusted with 0.1 M HEPES to a final pH of approximately 7.5. DNA concentrations were measured with the NanoDrop® ND 1000 Spectrophotometer at λ=260 nm, and a 260/280 ratio >1.8 was classified as pure. gDNA was stored at 4°C.

3.2.9 Analysis of CRISPR/Cas9 single-cell clones

To investigate DNA modifications of knock out SCCs (chapters 3.1.1.3 and 3.1.1.4), gDNA was isolated (chapter 3.2.8), and PCRs were performed. For this purpose, 200 ng of gDNA were amplified with the Q5® High-Fidelity DNA Polymerase and specific primers flanking the targeted regions according to the manufacturer's protocol on an Eppendorf Mastercycler®. A wildtype gDNA control was used as a reference. PCR products were loaded to a 2% agarose gel. For Twist1 motif knockout single-cell clones, PCR products were cleaned up with the DNA Clean & Concentrator-5 Kit according to the manufacturer's instructions and sequenced by Eurofins using the Mix2SeqKit according to manufacturer's instructions.

3.2.10 Cloning of *ZEB1* into the PiggyBac vector

To overexpress *ZEB1*, the All-in-One PiggyBac system designed for inducible transgene expression (Kim et al., 2016) was used. For this purpose, the pcDNA3.1+/C-(K)DYK vector containing the *ZEB1* (NM00128128) complementary DNA (cDNA; clone OHu27170C) and the pENTR1A-no ccdb vector were enzymatically digested (chapter 3.2.1). The *ZEB1* cDNA fragment and the pENTR1A backbone were cleaned up (chapter 3.2.4) and ligated (chapter 3.2.2) to create the pENTR1A-*ZEB1* vector. To transfer the *ZEB1* cDNA from the pENTR1A-*ZEB1* into the PB-TAC-ERP2 vector, a gateway reaction was set up using the Clonase™ Gateway™ LR Clonase II Enzym-Mix according to the manufacturer's protocol to generate the PB-*ZEB1* vector.

3.2.11 Omni-ATAC-sequencing

ATAC-sequencing was performed in collaboration with Helena Domínguez Moreno and Prof. Gunnar Schotta (Ludwig-Maximilian University, Germany). ATAC-sequencing was performed as described previously (Corces et al., 2017). Briefly, 5×10^4 cells (viability > 90%) were harvested by trypsinisation and pelleted by a centrifugation step at 4°C and 500 rcf for 5 min. If cell viability of cell suspensions used for ATAC-seq was below 90%, dead cells were removed with the Dead Cell Removal Kit according to the manufacturer's instructions before further processing. The following steps were performed by Helena Domínguez Moreno in the laboratory of Gunnar Schotta. Cell pellets were resuspended in ATAC resuspension buffer supplemented with 0.1% NP40, 0.1% TWEEN® 20 and 0.01% digitonin for lysis and incubated on ice for 3 min. Then 1 ml of ATAC resuspension buffer supplemented with 0.1% TWEEN® 20 only was added and centrifuged (500 rcf, 4°C, 10 min) to collect nuclei. Nuclei were subsequently re-suspended in 50 µl of the transposase reaction mix. Reactions were

incubated for 30 min at 37°C in a ThermoMixer® shaking at 900 rpm, and DNA was then purified using the Qiagen PCR clean-up MinElute Kit. Transposed DNA was subsequently amplified in 50 µl reactions with custom primers as described previously (Buenrostro, Giresi, Zaba, Chang, & Greenleaf, 2013). After four cycles of amplification, libraries were monitored with qPCR: 5 µl PCR sample in a 15 µl reaction with the same primers. qPCR output was monitored for the ΔRN; 0.25 ΔRN cycle number was used to estimate the number of additional cycles of the PCR reaction needed for the remaining PCR samples. Amplified libraries were purified with the PCR clean-up MinElute Kit and size-selected for fragments less than 600 bp using the Agencourt AMPure XP beads. Libraries were quality controlled by Qubit and Agilent DNA Bioanalyzer analysis. Deep sequencing was performed on a HiSeq 1500 system according to the standard Illumina protocol for 50 bp single-end reads. The following steps were performed by Helena Domínguez Moreno and/or Gunnar Schotta. ATAC-seq reads were aligned to the human genome hg38 using Bowtie (Langmead, 2010) with options "-q -n 2 --best --chunkmbs 2000 -p 32 -S". ATAC peaks over Input background were identified using Homer (Heinz et al., 2010) findPeaks.pl with option "-style factor". Peaks from all samples were merged using mergePeaks resulting in a unified Peak set. The peak list was filtered for promoter-associated peaks (distance to TSS < 1000bp) with bedtools. Raw ATAC coverage counts were then calculated with annotatePeaks. PCA analysis was performed on ATAC peaks coverage data with the R function prcomp. A list of differential ATAC peaks was determined with the DESeq2 results function for all possible comparisons between the samples and filtered for adjusted p-value <0.05 and $\log_2$ fold change >3. The cluster heatmap was generated based on normalized coverage data of all differential ATAC peaks using the R pheatmap package in k_means mode with 12 clusters. Cluster-specific peaks were further analysed for transcription factor motifs using Homer findMotifsGenome function in *de novo* mode.

ATAC resuspension buffer: 10mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$

2x Tagmentation buffer: 20 mM Tris-HCl pH 7.6, 10 mM $MgCl_2$, 20% dimethyl formamide, H_2O

Transposase reaction mix: 25 µl 2x Tagmentation buffer, 2.5 µl Tn5 Transposase, 5.25 µl H_2O, 16.5 µl PBS, 0.25 µl of 2% digitonin, 0.5 µl of 10% Tween-20

3.3 Methods for working with bacteria

3.3.1 Transformation of chemically competent bacteria

To transform bacteria with plasmid DNA, 25 µl of XL10-Gold ultracompetent cells or MAX Efficiency™ Stbl2™ competent cells were added to a 1.5 ml reaction tube containing the desired amount of plasmid DNA and incubated on ice for 10 min. After a heat shock step at

42°C for 30 sec in a water bath, the reaction tube was immediately transferred to ice for 2 min. Next, S.O.C. Medium was added to a final volume of 200 µl. Bacteria solution was incubated at 37°C and 900 rpm in a thermoblock for one hour. Different volumes of bacteria solution were plated onto agar plates containing the appropriate antibiotic and incubated at 37°C overnight.

3.3.2 Isolation of bacterial plasmid DNA

3.3.2.1 Mini plasmid preparation

To isolate a small amount of plasmid DNA from bacteria, a single bacterial colony was inoculated in 3 ml lysogeny broth (LB) medium containing the appropriate antibiotic and incubated shaking at 37°C overnight. The next day, 2 ml of the bacteria solution were centrifuged (16000 rcf, 3 min) in a 2 ml reaction tube. After removing the supernatant, the bacterial pellet was resuspended in 250 µl buffer P1, and 250 µl of buffer P2 were added, followed by inverting the tube ten times. 300 µl of buffer P3 were added, and the tube was again inverted ten times. After a centrifugation step at 16000 rcf and 4°C for 15 min, the supernatant was transferred to a new 1.5 ml reaction tube. 600 µl of 100% 2-Propanol were added and mixed thoroughly to precipitate DNA. DNA was pelleted by centrifugation at 16000 rcf and 4°C for 20 min. The DNA pellet was washed twice with 500 µl of 70% Ethanol (16000 rcf, 3 min). After removing the supernatant, the 1.5 ml reaction tube was left open to dry the DNA pellet at 85°C or room temperature. The pellet was resuspended in 50 µl elution buffer.

3.3.2.2 Midi plasmid preparation

Larger amounts of plasmid DNA were isolated from bacteria using the Plasmid Midi Kit according to the manufacturer's protocol with minor modifications. Briefly, 1 ml of LB medium containing the appropriate antibiotic were inoculated with 100 µl of a bacteria solution prepared for a Mini plasmid preparation (chapter 3.3.2.1) and incubated shaking at 37°C overnight. The next day, the bacteria solution was spun down and isolated according to the manufacturer's protocol. The DNA pellet was eluted in 150 µl.

3.4 Methods for DNA delivery

3.4.1 Transfection of cell lines

To deliver DNA into cells, 1.5 to 2×10^5 cells were seeded into 6-well plates (CRISPR/Cas9) or 10 cm dishes (PiggyBac) 24 hours before transfection. Cells were transfected using the

TransIT-X2 transfection reagent according to manufacturer's instructions with minor modifications. Briefly, growth medium in 6-well plates or 10 cm dishes was replaced with 1.8 ml or 9 ml, respectively, supplemented medium containing 1% penicillin/streptomycin. The transfection mix was prepared (Table 3-6), incubated at room temperature for 30 min, and carefully added to each well.

Table 3-6: General composition of the transfection mix

Compound	6-well	10 cm dish
DNA	2.5 µg	15 µg
TransIT-X2® Transfection Reagent	7.5 µl	45 µl
DMEM without supplements	to 200 µl	to 1 ml

3.5 Methods for working with RNA

3.5.1 RNA isolation

To isolate total RNA, the miRNeasy Mini Kit or the RNeasy Mini Kit was used according to the manufacturer's instructions. Briefly, cells growing in 6-well plates, 6 cm, or 10 cm dishes were washed once with PBS. 350 to 700 µl of lysis buffer were added, and cells were scraped thoroughly. Lysates were transferred to a 1.5 ml reaction tube and placed to -80°C to disrupt cells further. Next, lysates were thawed on ice, loaded on Qiashredder columns and spun at 14000 rcf for 2 min. Further steps of RNA isolation were performed with the miRNeasy Mini Kit or the RNeasy Mini Kit according to the manufacturer's protocols. The DNase digestion step suggested within the RNA isolation protocols was performed with the RNase-free DNase Set according to the manufacturer's instructions. RNA was eluted in 30 µl nuclease-free water, and concentration was measured with the NanoDrop® ND 1000 Spectrophotometer at λ=260 nm. RNA was stored at -80°C.

3.5.2 Complementary DNA synthesis

3.5.2.1 Reverse transcription of messenger RNA

To obtain complementary DNA (cDNA) of messenger RNA (mRNA), 1 µg of isolated total RNA (chapter 3.5.1) was reverse transcribed with the OneScript Plus cDNA Synthesis Kit using Oligo(dT) primers according to the manufacturer's protocol. A control without RNA (water control) and a control excluding the reverse transcriptase was included. The reverse transcription was performed in an Eppendorf Mastercycler® for 60 min at 50°C and terminated at 85°C for 5 min. cDNA was stored at -20°C.

3.5.2.2 Reverse transcription of micro RNA

To obtain cDNA of mature micro RNA (miRNA), 1 µg of isolated total RNA (chapter 3.5.1) was reverse transcribed with the miScript II RT Kit using the HiSpec buffer according to the manufacturer's instructions. A control without RNA (water control) was included. The reverse transcription was performed in an Eppendorf Mastercycler® for 60 min at 37°C followed by an inactivation step at 95°C for 5 min. cDNA was stored at -20°C.

3.5.3 Real-time semi-quantitative PCR

To measure relative transcript levels of different genes by real-time semi-quantitative PCR (qPCR), cDNA samples (chapters 3.5.2.1 and 3.5.2.2) were analysed using the Power SYBR™ Green PCR Master Mix. Briefly, for each primer pair, a 10 µl reaction was set up for a 384-well (Table 3-7, Table 3-8).

Table 3-7: Reaction mix for mRNA levels analysis via qPCR

Component	Volume per 384-well
Forward Primer [20 µM]	0.25 µl
Reverse primer [20 µM]	0.25 µl
Power SYBR™ Green PCR Master Mix	5.00 µl
Nuclease-free H_2O	2.50 µl
cDNA (1:10 diluted in nuclease-free H_2O)	2.00 µl
Total volume	10.00 µl

Table 3-8: Reaction mix for miRNA levels analysis via qPCR

Component	Volume per 384-well
Universal Primer (miScript SYBR® Green PCR Kit)	1.00 µl
miScript Primer Assay	1.00 µl
Power SYBR™ Green PCR Master Mix	5.00 µl
Nuclease-free H_2O	2.00 µl
cDNA (1:10 diluted in nuclease-free H_2O)	1.00 µl
Total volume	10.00 µl

To normalise expression values of different genes, *RPL32* (mRNA) or *RNU6-2* (miRNA) were used as housekeeper genes. cDNA synthesis controls (chapter 3.5.2.1) were included for *RPL32* or *RNU6-2*. For each primer pair, a control without cDNA was prepared. All samples were analysed as triplicates on a 384-well PCR plate in a QuantStudio 12K Flex qPCR system

using defined cycling steps (Table 3-9, Table 3-10). A post-amplification melting curve was performed to check for reaction specificity.

Table 3-9: qPCR cycling steps for analysis of mRNA levels

Step	Temperature	Duration	Cycles
Initialization	95°C	10 min	1x
Denaturation	95°C	15 sec	
Annealing	60°C	30 sec	40x
Extension	72°C	16 sec	
Melting Curve	60 to 95°C		

Table 3-10: qPCR cycling steps for analysis of miRNA levels

Step	Temperature	Duration	Cycles
Initialization	95°C	10 min	1x
Denaturation	95°C	15 sec	
Annealing	55°C	30 sec	40x
Extension	72°C	30 sec	
Melting Curve	55 to 95°C		

All data were analysed with the ΔCt method, with ΔCt = Ct (Gene of interest) – Ct (*RPL32* or *RNU6-2*). Relative expression levels are displayed as $1000 \times 2^{-\Delta Ct}$ on a $\log_{10}$ (lg) scale.

3.5.4 RNA-sequencing

For RNA-sequencing, total RNA was isolated using the miRNeasy Mini Kit (chapter 3.5.1). RNA-sequencing was performed by Elisabeth Graf, Sandy Lösecke, and Thomas Schwarzmayr of the group of Tim-Matthias Strom (Helmholtz Centre Munich, Germany). The following steps were performed by Elisabeth Graf and Sandy Lösecke in the lab of Tim-Matthias Strom. Library preparation was performed using the Illumina TruSeq® Stranded Total RNA Library Prep Kit with Ribo-Zero. Briefly, RNA integrity number (RIN) was determined with the Agilent 2100 BioAnalyzer from the RNA 6000 Nano Kit. For library preparation, 1 µg of total RNA was depleted for cytoplasmatic rRNAs, fragmented, and reverse transcribed with the Elute, Prime, Fragment Mix. A-tailing, adaptor ligation, and library enrichment were performed as described in the High Throughput protocol of the TruSeq RNA Sample Prep Guide. RNA libraries were assessed for quality and quantity with the Agilent 2100 BioAnalyzer

and the Quant-iT PicoGreen dsDNA Assay Kit and then sequenced as 150 bp paired-end runs on an Illumina HiSeq4000 platform.

The following steps were performed by Thomas Schwarzmayr. The STAR aligner (v 2.4.2a) with modified parameter settings (--twopassMode=Basic) was used for split-read alignment against the human genome assembly hg19 (GRCh37) and UCSC knownGene annotation (Dobin et al., 2013). To quantify the number of reads mapping to annotated genes, HTseq-count (v0.6.0) was used (Anders, Pyl, & Huber, 2015). FPKM (Fragments Per Kilobase of transcript per Million fragments mapped) values were calculated using custom scripts. PCA plots were created with the R package ggplot2 (Wickham, 2016).

3.6 Methods for working with proteins

3.6.1 Flow cytometry

3.6.1.1 Flow cytometric analysis

To assess cell surface protein levels, cells were resuspended and stained in supplemented MECGM containing the specific antibodies (50 µl/1x10^5 cells). After 40 min incubation on ice protected from light, PBS was added, and cells were pelleted at 4°C for 5 min at 490 rcf followed by resuspension in supplemented MECGM (100 µl/1x10^5 cells). Cell suspensions were filtered through a 35 µm-strainer into 5 ml round-bottom tubes. Next, cells were analysed on a FACS Aria IIIu with the FACS Diva™ software. Cell debris and doublets were excluded with the forward and side scatter. Gating strategies can be found in chapter 7 (Figure 7-1). To allow gating for viable cells, 7-AAD or SYTOX™ Blue Dead Cell Stain was added to each sample before analysis. Samples of unstained and single-stained cells were used as references to set negative and positive gates, respectively. Data were recorded with the FACS Diva™ software and further analysed using FlowJo V10.

3.6.1.2 Fluorescence-activated cell sorting

To separate single cells based on their expression levels of cell-surface markers (e.g. EPCAM), cells were stained and prepared as described in chapter 3.6.1.1. For separation based on the expression of fluorochromes (e.g. GFP), cells were directly stained with 7-AAD or SYTOX™ Blue Dead Cell Stain and further processed as described in chapter 3.6.1.1. To set positive and negative gates, single-stained cells and unstained or non-fluorescent cells were used, respectively. Cell debris and doublets were again excluded with the forward and side scatter. 7-AAD or SYTOX™ Blue Dead Cell Stain was used to gate for viable cells. Gating strategies can be found in chapter 7 (Figure 7-1). Fluorescence-activated cell sorting (FACS)

was performed on a FACSAria IIIu flow cytometer using the 70 µm or 100 µm nozzle and set-ups according to the manufacturer's recommendations. Cells were sorted into tubes containing 1 ml supplemented MECGM. Sorted cells were centrifuged at 4°C for 5 min at 490 rcf, followed by resuspension in the appropriate growth medium. A fraction of the cell suspension was used for immediate re-analysis.

3.6.2 Immunoblotting

3.6.2.1 Isolation of total protein

To isolate protein from cells, cells were washed once with PBS. Next, cells were carefully scraped in 1 ml PBS on ice, transferred to a 1.5 ml reaction tube, and pelleted by centrifugation at 4°C and 2500 rcf for 5 min. After aspiration of the supernatant, cells were lysed by resuspending the pellet in 50 to 150 µl RIPA buffer. To enhance the lysis efficiency, reaction tubes were transferred to a -80°C freezer. After freezing for at least overnight, lysates were thawed on ice and centrifuged at 14000 rcf for 10 min to pellet the cell debris. The protein-containing supernatants were transferred to a new 1.5 ml reaction tube, and protein concentrations were either directly determined as described in chapter 3.6.2.3 or protein extracts were stored at -80°C.

RIPA buffer: 50 mM Tris pH 8.0, 150 mM NaCl, 1% (w/v) NP-40, 0.5% (w/v) sodium deoxycholate, 0.1% (w/v) SDS, 5 mM EDTA pH 8.0, Milli-Q water, 4°C

Lysis buffer (freshly prepared for each use): RIPA buffer + Phosphatase Inhibitor Cocktail 2 and 3 and Protease Inhibitor Cocktail (each 1:100)

3.6.2.2 Isolation of nuclear and cytoplasmic protein fractions

To isolate nuclear and cytoplasmic protein fractions of HMLE-Twist1-ER cells, the NE-PER™ Nuclear and Cytoplasmic Extraction Reagents Kit was used according to the manufacturer's instructions with minor modifications. Briefly, cells growing on 10 cm dishes were harvested with 0.15% Trypsin-EDTA and counted. 2×10^6 cells were pelleted by centrifugation for 5 min at 4°C and 500 rcf, once washed with 1 ml PBS, transferred into a 1.5 ml reaction tube and again pelleted by a centrifugation step for 5 min at 4°C and 500 rcf. After removing the supernatant, the cell pellet was suspended in 100 µl ice-cold CER I supplemented with protease inhibitor cocktail (1:100). To completely suspend the cell pellet, the tube was vortexed on the highest setting for 15 sec followed by an incubation step on ice for 10 min. Next, 5.5 µl of ice-cold CER II were added, and the tube was vortexed on the highest setting for 5 seconds. After incubation on ice for 1 min, the tube was vortexed again for 5 seconds on the highest

setting and then centrifuged for 5 min at 4°C and 21100 rcf. The supernatant containing the cytoplasmic protein extract was transferred to a fresh, pre-chilled 1.5 ml tube and placed on ice. The remaining pellet was suspended in 50 µl NER supplemented with protease inhibitor cocktail (1:100) and vortexed on the highest second for 15 seconds. After an incubation step of 10 min on ice, the tube was vortexed again for 15 seconds on the highest setting and then again incubated on ice for 10 min. The incubation and vortexing steps were repeated for a total of 40 min. Then the tube was centrifuged at 4°C and 21100 rcf for 10 min. The supernatant containing the nuclear protein extract was transferred to a fresh, pre-chilled 1.5 ml tube and placed on ice. Protein concentrations of cytoplasmic and nuclear extracts were either directly determined as described in chapter 3.6.2.3, or protein extracts were stored at -80°C.

3.6.2.3 Determination of protein concentrations

To determine the protein concentration of protein extracts, samples were measured using the detergent-compatible (DC)™ Protein Assay Kit.

For that, 3 to 5 µl of each sample were pipetted into a well of a 96-well plate in duplicates. To prepare a standard curve, protein standards of 0.5 µg, 1 µg, 2 µg, 5 µg, 7.5 µg, 10 µg, and 25 µg (Stock 1 mg/ml BSA/RIPA buffer) and a blank (0 µg/ml, 3µl RIPA buffer) were pipetted as duplicates. All the next steps were performed according to the manufacturer's instructions. After determining the optical density (OD) by measuring the absorbance on an iMark Microplate Absorbance Reader at 750nm, protein concentrations were calculated based on the standard curve obtained from the protein standard.

3.6.2.4 Sodium dodecyl sulphate-Polyacrylamide gel electrophoresis

To separate proteins based on their molecular weight, Sodium dodecyl sulphate-Polyacrylamide gel electrophoresis (SDS-PAGE) was used. For that, buffers for a 10% or 12.5% separating gel were mixed and, directly after adding TEMED and APS, poured into prepared glass plates of the Mini-PROTEAN® Tetra System. To ensure a straight top border, 100% 2-propanol was used to overlay the separating gel until completed polymerization. After removing the 2-propanol, the stacking gel solution was prepared and poured on top of the separating gel, and a 15-well comb was put in. After complete polymerisation, gels were loaded. Therefore, 10 µg of protein were mixed with 5x SDS loading buffer first, incubated at 95°C for 5 min to denature proteins, and then transferred to ice before loading into gel pockets. 8 µl of PageRuler™ Plus protein ladder were loaded into a gel pocket as a marker for protein sizes. Gel electrophoresis was performed at 120 V for 60 to 90 min in 1x Running buffer.

12.5% separating gel: 1.95 ml 30% Acrylamide (Rotiphorese® Gel 30), 2.75 ml Milli-Q water, 1.25 ml 1 M Tris pH 8.8, 50 µl 10% (w/v) SDS, 3.75 µl TEMED, 37.5 µl 10% (w/v) APS

10% separating gel: 1.65 ml 30% Acrylamide (Rotiphorese® Gel 30), 3.05 ml Milli-Q water, 1.25 ml 1 M Tris pH 8.8, 50 µl 10% (w/v) SDS, 3.75 µl TEMED, 37.5 µl 10% (w/v) APS

Stacking gel: 416.5 µl 30% Acrylamide (Rotiphorese® Gel 30), 1.73 ml Milli-Q water, 312.5 µl 1 M Tris pH 6.8, 25 µl 10% (w/v) SDS, 2.5 µl TEMED, 12.5 µl 10% (w/v) APS

5x SDS loading buffer: 30% (v/v) Glycerol, 10% (w/v) SDS, 0.25 M Tris/HCl pH 6.8, 0.05% (w/v) bromophenol blue, 10% (v/v) β-Mercaptoethanol, Milli-Q water

1x Running buffer: 25 mM Tris, 192 mM Glycine, 0.1% (w/v) SDS, Milli-Q water

3.6.2.5 Wet transfer and protein detection

To transfer proteins separated by SDS-PAGE (chapter 3.6.2.4) to a polyvinylidene fluoride (PVDF) blotting membrane, a Mini Trans-Blot® Electrophoretic Transfer Cell was used. Before assembling the wet blot according to the manufacturer's instructions, the PVDF blotting membrane was activated by incubating it in methanol. After the assembly of all parts, blotting was performed at 300 mA for 1.5 to 2 hours in 1x Transfer buffer under cooled conditions. After disassembling, the PVDF blotting membrane was first washed in TBS/T on a 2D Rocker followed by a blocking step with 5% milk or BSA at room temperature for 1 hour. Next, the PVDF blotting membrane was incubated overnight in a 50 ml falcon containing the primary antibody solution, prepared according to the manufacturer's protocol, on a horizontal falcon roller at 4°C. The next day, the membrane was washed three times with TBS/T for 10 min before incubation with the appropriate HRP-conjugated secondary antibody in 5% milk or 5% BSA. After 1 hour of incubation at room temperature, the membrane was washed three times with TBS/T and once with TBS for 10 min. To visualize the protein abundance, 600 µl of ECL™ mix were prepared and applied to the membrane according to the manufacturer's instructions. Chemiluminescence was detected with a ChemiDoc™ MP Imaging System and the Image Lab software.

1x Transfer buffer: 25 mM Tris, 192 mM Glycine, 20 % (v/v) Methanol

10x TBS: 1.5 M NaCl, 0.1 M Tris, MilliQ-water, pH 7.2-7.4 with HCl

1x TBS/T: 10% (v/v) 10x TBS, 0.1% (v/v) TWEEN® 20, MilliQ-water

5% milk: 5% (m/v) non-fat dried milk powder in 1x TBS/T

5% BSA: 5% (m/v) Albumin Fraction V (BSA) in 1x TBS/T

3.7 Immunofluorescence of cells in 2D

Cells were cultured in 96-well CellCarrier™ microplates suited for immunofluorescence. For fixation, medium was aspirated, and cells were washed with PBS. Next, cells were fixed with 4% paraformaldehyde (PFA) at room temperature for 10 to 15 min. After removing the fixation solution, cells were washed with PBS three times and then stored in PBS at 4°C.

For staining, PBS was removed, and cells were permeabilized with a permeabilization solution for a maximum of 2 min followed by three washing steps with PBS and a blocking step with a blocking solution for at least one hour at room temperature. After another three washing steps with PBS, cells were incubated with the primary antibody solution overnight at 4°C. The next day, the primary antibody solution was aspirated, and cells were washed with PBS three times. Cells were then incubated with the secondary antibody solution for two to three hours at room temperature. After removing the secondary antibody solution and another three washing steps with PBS, 4′,6-diamidino-2-phenylindole (DAPI) staining solution was added for 5 min to stain cell nuclei. Next, the DAPI staining solution was removed, and cells were washed twice with PBS and once with Milli-Q water. For mounting, Aqua-Poly/Mount mounting medium was added to each well. Cells were imaged with a Zeiss Axio Imager.M2 using the Zen Pro software, and images were processed with GIMP.

Permeabilization solution: 0.2% Triton X-100 in PBS

Blocking solution: 0.1% BSA in PBS + 10% goat serum

Antibody solution: Respective antibody diluted in 0.1% BSA in PBS

DAPI solution: 167 ng/ml DAPI in PBS

3.8 3D-collagen I gel-based methods

3.8.1 Fixation of floating 3D-collagen I gels

After ten days in culture, cells growing in floating 3D-collagen I gels were fixed. For this purpose, the medium was removed, and gels were washed with 1 ml PBS once. Next, gels were fixed with 4% PFA for 15 min shaking on an orbital shaker. After removal of the fixation solution, gels were washed with PBS three times. Gels were stored in PBS at 4°C.

4% PFA: 1:4 dilution of 16% PFA in PBS

3.8.2 Immunofluorescence of 3D-collagen I gels

All washing and incubation steps were performed on an orbital shaker. First, PBS of fixed gels (chapter 3.8.1) was aspirated, and gels were permeabilised for 10 min in 1 ml permeabilization solution. Next, gels were washed twice with PBS, followed by a blocking step in 1 ml blocking

solution for two to three hours. After another washing step with PBS, gels were incubated with 100 µl of the primary antibody solution shaking at 4°C overnight. The next day, gels were washed three times with PBS and incubated with 100 µl of the secondary antibody solution shaking at room temperature for 2 to 3 hours. After three washing steps, 100 µl of DAPI staining solution were added for 5 min to stain cell nuclei. Gels were twice washed with PBS and once with Milli-Q water before transferring gels to a microscopy slide. Excess water was carefully removed, and gels were mounted with Aqua-Poly/Mount mounting medium and a coverslip. Slides were dried at room temperature for one to two days, sealed with nail polish and stored at 4°C. Images were taken on a FLUOVIEW FV1200 inverted Confocal Laser Scanning Microscope with a 20x objective and the FV10-ASW software. Images were processed with GIMP software.

Permeabilization solution: 0.2% Triton X-100 in PBS

Blocking solution: 0.1% BSA in PBS + 10% goat serum

Antibody solution: Respective antibody diluted in 0.1% BSA in PBS

DAPI solution: 167 ng/ml DAPI in PBS

3.8.3 Carmine staining of 3D-collagen I gels

To visualize colonies formed by cells growing in 3D-collagen I gels, gels were stained with carmine. To do so, PBS of fixed gels was removed, and 1 ml of a carmine staining solution was added per gel. Gels were incubated overnight at 4°C on an orbital shaker. Gels were then imaged with a Leica DFC450 C Stereomicroscope with a Plan 0.8x LWD objective and the Leica Application Suite V4 software. Single pictures were stitched with Panorama Stitcher to obtain an image of a whole gel. Carmine positive colonies were counted using ImageJ.

Carmine staining solution: 2 g/l carmine, 5 g/l aluminium potassium sulphate, two crystals of thymol

3.9 Data presentation and statistical analysis

Data are presented as mean±SEM of n=x experiments, with x indicating the number of independent experiments performed, unless otherwise stated. Statistical analyses were performed using GraphPad Prism software. In general, a statistical significance was determined using multiple t-tests (without assuming a standard deviation) with the Holm-Sidak correction. An adjusted p-value (p_{adj}) <0.05 was considered significant.
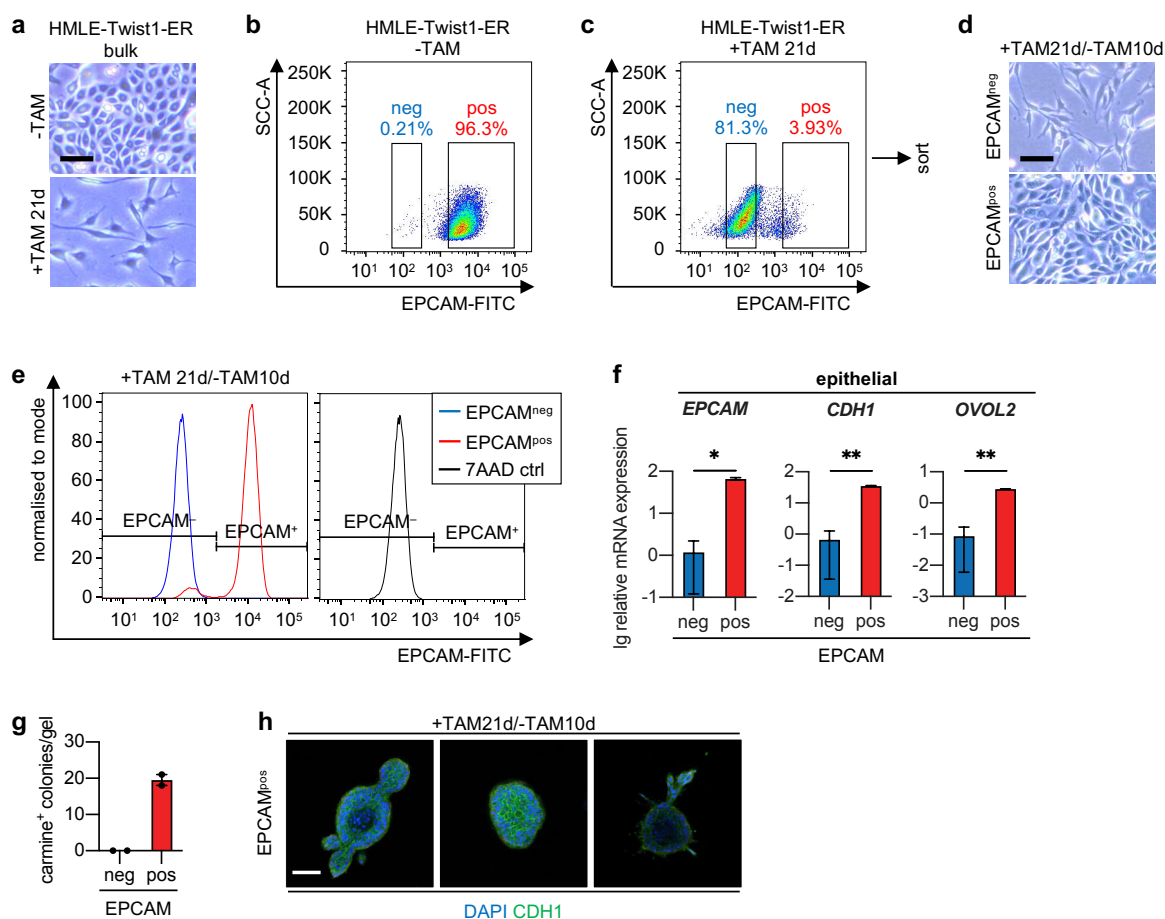
4   Results

4.1   Twist1-activation in HMLE-Twist1-ER bulk cells revealed the existence of two subpopulations identified by different EPCAM levels

The aim of this study was to unravel the heterogeneity observed upon transient activation of the EMT-transcription factor (TF) Twist1 in HMLE-Twist1-ER bulk cells (Schmidt et al., 2015). In detail, Schmidt and colleagues activated Twist1 in HMLE-Twist1-ER bulk cells by treating the cells with Tamoxifen (TAM) and thereby induced EMT in these cells. However, upon treatment followed by removal of Tamoxifen, a subset reverted to an epithelial state, whereas another subset of cells remained mesenchymal (Schmidt et al., 2015).

To reproduce these results and investigate how the heterogeneity evolves, I used HMLE-Twist1-ER bulk cells and investigated the effects of Twist1-activation in these cells. In culture, untreated (-TAM) HMLE-Twist1-ER bulk cells grew as epithelial, cobblestone-like colonies (Figure 4-1a). Moreover, flow cytometric analysis revealed that untreated HMLE-Twist1-ER bulk cells expressed the epithelial cell adhesion molecule (EPCAM, Figure 4-1b), a cell surface protein that is important for the formation of epithelial cell-cell junctions. Moreover, EPCAM is frequently used to detect circulating tumour cells (CTCs, Allard et al., 2004), and its expression has already been shown to correlate with tumour progression and metastasis of breast cancer (Baccelli et al., 2013). Upon treatment with Tamoxifen and concomitant induction of EMT by Twist1-activation for 21 days (+TAM 21d), HMLE-Twist1-ER bulk cells trans-differentiated from an epithelial, cobblestone-like morphology to cells with an elongated, spindle-like morphology, suggesting that the cells have undergone an epithelial-mesenchymal transition (EMT, Figure 4-1a). Additionally, flow cytometric analysis of EPCAM revealed that the majority of cells had lost EPCAM surface protein (neg: EPCAM^{neg}) after three weeks of continuous Twist1-activation (+TAM 21d, Figure 4-1c). Interestingly, a minority of cells still displayed EPCAM protein on their surface (pos: EPCAM^{pos}, Figure 4-1c), despite continuous activation of the EMT-TF Twist1 for 21 days. I excluded failure in Twist1-ER expression in these cells, as all cells were continuously cultured in growth medium containing Blasticidin to select for cells harbouring the Twist1-ER construct.

Together, these results showed that upon activation of the EMT-TF Twist1, the majority of HMLE-Twist1-ER bulk cells underwent EMT, indicated by loss of EPCAM and trans-differentiation from an epithelial to a mesenchymal morphology. Interestingly, a small fraction of HMLE-Twist1-ER cells maintained expression of EPCAM despite long-term activation of Twist1, indicating that these cells resisted Twist1-mediated trans-differentiation. In comparison to the results obtained by Schmidt and colleagues, who observed the same heterogeneity, yet only after transient Twist1-activation (Schmidt et al., 2015), the results obtained in this study revealed that the heterogeneity evolved already during Twist1-activation.

Figure 4-1: Twist1-activation revealed a subset of HMLE-Twist1-ER cells that underwent irreversible EMT and a subset that resisted trans-differentiation

The results in this figure have been published in a pre-print at bioRxiv (Eichelberger et al., 2020).

a) Bright-field images of HMLE-Twist1-ER bulk cells, before (-TAM) or at day 21 of continuous Tamoxifen treatment (+TAM 21d). Scale bar: 100 μm. **b)** Flow cytometric analysis of EPCAM of HMLE-Twist1-ER bulk cells before Tamoxifen treatment (-TAM). neg: EPCAM^{neg}, pos: EPCAM^{pos}. **c)** Flow cytometric analysis of EPCAM of HMLE-Twist1-ER bulk cells treated with Tamoxifen for 21 days (+TAM 21d). Gates for fluorescent activated cell sorting are shown in black. neg: EPCAM^{neg}, pos: EPCAM^{pos}. **d)** Bright-field images of sorted EPCAM^{pos} and EPCAM^{neg} HMLE-Twist1-ER cells, treated with Tamoxifen for 21 days followed by Tamoxifen withdrawal for ten days (+TAM21d/-TAM10d). Scale bar: 100 μm. **e)** Flow cytometric analysis of EPCAM of sorted EPCAM^{pos} and EPCAM^{neg} HMLE-Twist1-ER cells treated as described in d). Gates were set according to unstained control. ctrl: control. **f)** Lg relative *EPCAM*, *CDH1*, and *OVOL2* mRNA expression levels in sorted EPCAM^{pos} (pos) and EPCAM^{neg} (neg) HMLE-Twist1-ER cells treated as described in d). n=2. Mean $\pm$ SEM. Multiple t-testing with Holm-Sidak correction. p_{adj}-values: $*<0.05$, $**<0.005$. **g)** Quantification of carmine positive (carmine$^+$) colonies of sorted EPCAM^{pos} (pos) and EPCAM^{neg} (neg) HMLE-Twist1-ER cells growing in a 3D-collagen assay, treated with Tamoxifen for 21 days followed by Tamoxifen withdrawal for 10 days. n=2. Mean $\pm$ SEM. **h)** Images of immunofluorescent staining of DAPI (blue) and CDH1 (green) of EPCAM^{pos} cells growing in 3D-collagen gels, treated as described in g). Scale bar: 50 μm.

Based on the previous observations, I set out to better understand how the heterogeneity, with respect to the ability of cells to undergo Twist1-mediated EMT, develops. To address this question, I separated both populations that evolved during 21 days of Twist1-activation (+TAM 21d) by fluorescent activated cell sorting (FACS, Figure 4-1c): the population of cells that lost surface EPCAM (neg: EPCAM^{neg}) and the population of cells that maintained surface EPCAM (pos: EPCAM^{pos}). Albeit I have observed heterogeneity already during Twist1-activation, I transferred both sorted subpopulations into Tamoxifen-free conditions for ten days (+TAM 21d/-TAM 10d) because Schmidt and colleagues showed that a transient activation of Twist1 is essential for cancer progression (Schmidt et al., 2015). By doing so, I aimed to investigate whether EPCAM^{neg}-sorted cells could revert to an epithelial state after transient activation of Twist1 (+/-TAM) like it was observed by Schmidt and colleagues (Schmidt et al., 2015).

Propagating both separated populations in Tamoxifen-free growth medium for ten days (+TAM 21d/-TAM 10d) revealed that EPCAM^{pos}-sorted cells continued to grow as cells with an epithelial, cobblestone-like morphology (Figure 4-1d). In line with morphology, flow cytometric analysis of surface EPCAM protein revealed that EPCAM^{pos}-sorted cells displayed surface EPCAM protein (EPCAM^{+}, Figure 4-1e). Intriguingly, no EPCAM^{neg}-sorted cells reverted to cells with an epithelial morphology, but instead, all cells maintained a mesenchymal, spindle-like morphology (Figure 4-1d). In addition to that, flow cytometry showed that EPCAM^{neg}-sorted cells did not re-express EPCAM but instead were still negative for EPCAM protein (EPCAM^{-}, Figure 4-1e), indicating that EPCAM^{neg}-sorted cells did not revert to an epithelial state despite the withdrawal of Twist1-activation.

Real-time semi-quantitative PCR (qPCR) analyses of transcripts of epithelial (*EPCAM, CDH1, OVOL2*) and mesenchymal (*ZEB1*) genes in EPCAM^{neg}-sorted cells and EPCAM^{pos}-sorted cells after transient Twist1-activation (+TAM 21d/-TAM 10d) showed further differences between both subsets of cells (Figure 4-1f). Briefly, qPCR analyses revealed significantly higher transcript levels of epithelial markers and lower expression levels of mesenchymal markers in EPCAM^{pos}-sorted cells than EPCAM^{neg}-sorted cells (p_{adj}<0.05 or <0.005, Figure 4-1f). In detail, transcript levels of epithelial genes like *EPCAM, CDH1*, and *OVOL2* were, respectively, 57-, 53-, and 32-fold higher in EPCAM^{pos}-sorted cells compared to EPCAM^{neg}-sorted cells (Figure 4-1f). Also, transcript levels of *ZEB1*, a gene that encodes an EMT-TF factor that represses epithelial genes, were 13-fold higher in EPCAM^{neg}-sorted cells compared to EPCAM^{pos}-sorted cells (Figure 4-1f).

These results showed that a subset of HMLE-Twist1-ER cells underwent trans-differentiation from an epithelial to a mesenchymal state, indicated by the loss of expression of epithelial genes including *CDH1* and *EPCAM*. Moreover, when Twist1-activation was withdrawn, these cells maintained the repression of epithelial genes and, in turn, remained a mesenchymal state. In contrast, the other subset of HMLE-Twist1-ER cells maintained the expression of

epithelial genes like *EPCAM* and *CDH1* and, concomitantly, an epithelial state, despite the activation of the EMT-TF Twist1.

As it is known that metastases formed by breast cancer cells typically express the epithelial marker CDH1 (Kowalski et al., 2003), my next aim was to investigate colonisation of both sorted populations in a three-dimensional environment. For this purpose, I used a floating 3D-collagen gel assay that has been initially developed to investigate the regeneration of mammary epithelial cells (Linnemann et al., 2015). Nevertheless, this assay has already been used to investigate colonisation as well as invasion of HMLE-Twist1-ER bulk cells (Schmidt et al., 2015). Of note, the latter study revealed that a transient rather than a continuous activation of Twist1 is essential for outgrowth and colonisation of HMLE-Twist1-ER cells in a three-dimensional environment (Schmidt et al., 2015).

Based on this information, I set out to seed EPCAM^{neg} and EPCAM^{pos} cells previously treated with Tamoxifen for 21 days into 3D-collagen gels directly after sorting and withdraw Tamoxifen treatment (+/-TAM). After ten days of propagation in 3D-collagen gels floating in Tamoxifen-free growth medium (+TAM 21d/-TAM 10d), cells were analysed further. Interestingly, carmine staining of 3D-collagen gels showed that only EPCAM^{pos}-sorted cells formed organoids in the three-dimensional environment (Figure 4-1g), whereas EPCAM^{neg}-sorted cells failed to grow in the three-dimensional environment. These organoids expressed the epithelial cell-cell junction protein CDH1, as detected by immunofluorescent staining (Figure 4-1h). Remarkably, in these organoids generated by EPCAM^{pos}-sorted cells, some degree of invasion was visible, although the invading cells expressed the epithelial cell-cell junction protein CDH1 on their surface (Figure 4-1h).

These results showed that only cells that resisted Twist1-mediated EMT and maintained expression of EPCAM were capable of growing in a three-dimensional environment when Twist1-activation was ceased. Moreover, although some degree of invasiveness could be observed in these cells after transient Twist1-activation, all cells displayed expression of the epithelial cell-cell junction protein CDH1. Taken together, these results showed that a subset of HMLE-Twist1-ER bulk cells underwent Twist1-mediated EMT. Though, instead of reverting to an epithelial state after the removal of Twist1-activation, these cells remained in a mesenchymal state. This phenomenon was accompanied by low expression levels of epithelial genes, including *EPCAM* and *CDH1*, and a diminished capacity to colonise in a three-dimensional environment. Importantly, the results in this study indicated that the heterogeneity of HMLE-Twist1-ER bulk cells, in terms of EMT, evolved already during Twist1-activation and not, as previously assumed, after transient activation of the EMT-TF Twist1. In detail, resistance to Twist1-mediated EMT resulted in colonisation, whereas trans-differentiation from an epithelial to a mesenchymal state upon transient Twist1-activation inhibited colonisation in a three-dimensional environment.

4.2 A subset of HMLE-Twist1-ER single-cell clones resisted Twist1-mediated EMT and colonised in a three-dimensional environment

The previous results showed that a subset of HMLE-Twist1-ER cells underwent EMT upon Twist1-activation, whereas another subset of HMLE-Twist1-ER cells resisted Twist1-mediated EMT and thereby maintained colonising capacities. Based on these findings, I wished to understand what characterises and determines whether cells resist or undergo Twist1-mediated trans-differentiation.

To obtain precise results, I set out to perform analyses on a single-cell level. For this purpose, I used several HMLE-Twist1-ER single-cell clones (SCCs) that were isolated from the untreated (-TAM) HMLE-Twist1-ER bulk population.

When propagated in normal and untreated conditions (-TAM), all SCCs formed epithelial, cobblestone-like colonies (Figure 4-2a). However, upon Tamoxifen treatment and concomitant Twist1-activation for up to two weeks (+TAM 7d, +TAM 14d), a subset of SCCs trans-differentiated from an epithelial to a mesenchymal morphology (EMT-competent M-SCCs: M1, M2, M3, Figure 4-2a), suggesting that these cells underwent EMT. On the contrary, the other subset of SCCs continued to grow as epithelial, cobblestone-like colonies despite Twist1-activation (EMT-resistant E-SCCs: E1, E2, E3, Figure 4-2a), indicating that these cells resisted Twist1-mediated trans-differentiation.

Again, these results showed that upon Twist1-activation, a subset of cells was competent to undergo EMT, whereas the other subset remained epithelial. In addition, these results further underpinned the observation that the heterogeneity, with regard to EMT, indeed already evolved during Twist1-activation.

In connection with the previous observations made in the HMLE-Twist1-ER bulk population, I hypothesised that E-SCCs represented EPCAM^{pos}-sorted cells, whereas M-SCCs represented EPCAM^{neg}-sorted cells. To verify this hypothesis, I aimed to investigate the functional traits of E-SCCs and M-SCCs further. For this purpose, I propagated E-SCCs and M-SCCs in floating 3D-collagen gels to investigate their potential to colonise in a three-dimensional environment. In detail, E-SCCs (E1, E2, E3) and M-SCCs (M1, M2, M3) were embedded in collagen gels either untreated (-TAM) or after 14 days of Twist1-activation. Then, cells were propagated in Tamoxifen-free conditions for 10 days (-TAM or +/-TAM).

The propagation of untreated (-TAM) E-SCCs and M-SCCs in 3D-collagen gels floating in Tamoxifen-free conditions revealed a similar frequency of organoid formation in both E-SCCs and M-SCCs (Figure 4-2b). Moreover, immunofluorescent staining showed that organoids formed by untreated E-SCCs and M-SCCs expressed the epithelial cell-cell junction protein CDH1 and some degree of the mesenchymal marker Vimentin (VIM, -TAM, Figure 4-2c).
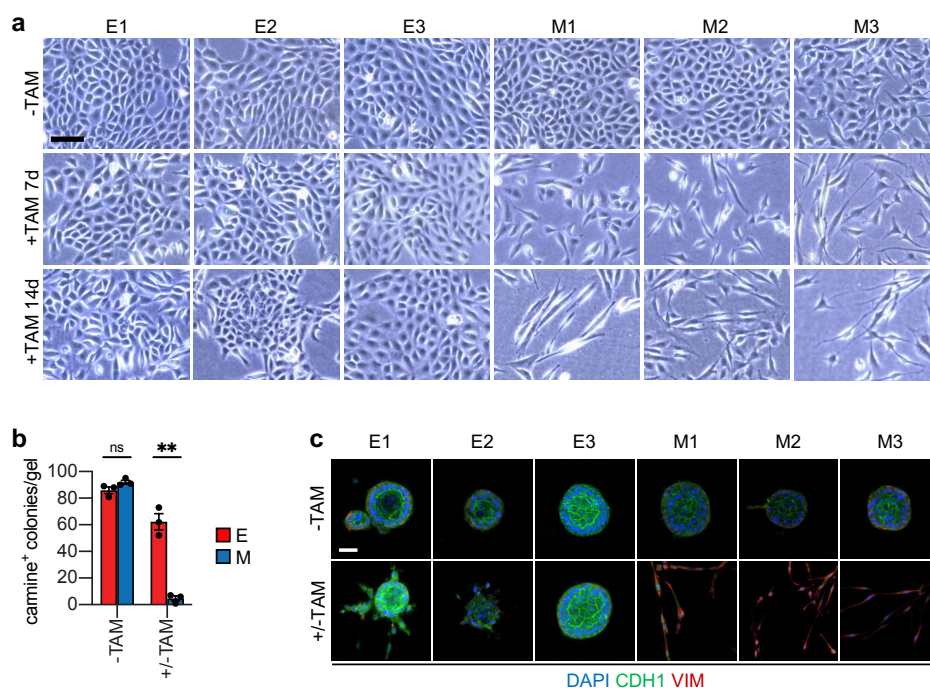
Figure 4-2: A subset of HMLE-Twist1-ER single-cell clones (SCCs) resisted Twist1-activation and maintained the colonising capacity in 3D-collagen gels

The results in this figure have been published in a pre-print at bioRxiv (Eichelberger et al., 2020).

a) Bright-field images of representative EMT-resistant single cells clones (E-SCCs: E1-E3) and representative EMT-competent SCCs (M-SCCs: M1-M3), before (-TAM) or at day 7 (+TAM7d) or day 14 of Tamoxifen treatment (+TAM14d). Scale bar: 100 µm. **b)** Quantification of carmine positive (carmine+) colonies of E-SCCs and M-SCCs growing in a 3D-collagen assay, before (-TAM) or after 14 days of Tamoxifen treatment followed by Tamoxifen withdrawal for 10 days (+/-TAM). n=3. Mean $\pm$ SEM. Multiple t-testing with Holm-Sidak correction. p_{adj}-values: ∗∗<0.005, ns: not significant. **c)** Images of immunofluorescent staining of DAPI (blue), CDH1 (green), and VIM (red) of E-SCCs and M-SCCs growing in a 3D-collagen assay, treated as described in b). Scale bar: 50 µm.

Interestingly, and in line with these findings, Schmidt and colleagues also observed the formation of epithelial organoids when propagating untreated HMLE-Twist1-ER bulk cells in 3D-collagen gels (Schmidt et al., 2015). However, after 14 days of Twist1-activation followed by 10 days of propagation in 3D-collagen gels as well as Tamoxifen withdrawal (+/-TAM), I observed differences in colonising activity between E-SCCs and M-SCCs. In detail, E-SCCs colonised the 3D-collagen gels more efficiently compared to M-SCCs (p_{adj}<0.005, Figure 4-2b). Additionally, organoids formed by E-SCCs after transient Twist1-activation (+/-TAM) expressed the epithelial cell-cell junction protein CDH1 (Figure 4-2c). Again, in these organoids, some cells displayed an invasive phenotype despite the expression of CDH1 (Figure 4-2c). Of note, this was also observed in organoids formed by EPCAM^pos-sorted cells of the HMLE-Twist1-ER bulk (Figure 4-1h). In contrast, the few cells of M-SCCs growing in 3D-collagen gels could be observed as single mesenchymal cells expressing the mesenchymal

marker VIM (+/-TAM, Figure 4-2c). Of note, organoids formed by E-SCCs after transient Twist1-activation showed only some degree of VIM expression (+/- TAM, Figure 4-2c).

Together, these observations supported the previous finding of two subsets of cells within the HMLE-Twist1-ER bulk population: An EMT-resistant population (E-SCCs) and an EMT-competent population (M-SCCs). Moreover, these results further underpinned that resistance to transient Twist1-activation and maintenance of epithelial properties allowed the colonisation of cells in a three-dimensional environment. On the contrary, the competency to trans-differentiate from an epithelial to a mesenchymal state during transient activation of the EMT-TF Twist1 repressed the ability of cells to grow in a three-dimensional environment.

4.3 Twist expression, localisation, and target gene expression were comparable in EMT-competent and EMT-resistant cells

The previous findings showed that a subset of HMLE-Twist1-ER cells underwent EMT upon Twist1-activation (M-SCCs), whereas the other subset of cells resisted Twist1-mediated trans-differentiation (E-SCCs). I previously excluded differential Twist1-activation, as all cells were continuously propagated in growth medium supplemented with Blasticidin to select for cells expressing the Twist1-ER construct. Nevertheless, I aimed to investigate Twist1 expression levels and activity in E-SCCs and M-SCCs in more detail, to examine whether differences could determine whether cells resist or undergo EMT, respectively.

For this purpose, I first assessed *Twist1* transcript levels in E-SCCs and M-SCCs before Tamoxifen treatment (-TAM), after 7 days of Tamoxifen treatment (+TAM 7d), and after 14 days of Tamoxifen treatment (+TAM 14d) by qPCR. Doing so revealed similar *Twist1* transcript levels in E-SCCs and M-SCCs in all conditions (Figure 4-3a).

As Tamoxifen treatment triggers the activation of Twist1, I aimed to investigate whether Twist1 trans-located from the cytoplasm into the nucleus in both E-SCCs and M-SCCs. For this purpose, I isolated nuclear and cytoplasmic protein fractions and assessed Twist1 protein levels by immunoblot (Figure 4-3b). Doing so revealed that upon treatment with Tamoxifen (+TAM 7d and +TAM 14d), Twist1-ER trans-located into the nucleus in both E-SCCs and M-SCCs (Figure 4-3b), indicating that resistance and responsiveness to Twist1-activation were not regulated by differential activation and localisation of Twist1 itself.

Previously, I have observed that EPCAM[neg]-sorted HMLE-Twist1-ER cells that have undergone EMT remained in a mesenchymal state despite the withdrawal of Twist1-activation. Because of that, I set out to determine whether failure in Twist1-deactivation or rather re-localisation from the nucleus to the cytoplasm might control that cells remain in a mesenchymal state. As I hypothesised that M-SCCs and E-SCCs represented EPCAM[neg]-sorted and EPCAM[pos]-sorted cells, respectively, I investigated Twist1 localisation in E-SCCs and M-SCCs

after transient Twist1-activation. In detail, I analysed nuclear and cytoplasmic Twist1 levels in E-SCCs and M-SCCs treated with Tamoxifen for 14 days followed by Tamoxifen withdrawal for 7 days (+TAM 14d/-TAM 7d).

Immunoblot analysis indeed revealed that Twist1 was de-localised from the nucleus in E-SCCs and M-SCCs similarly (Figure 4-3b), suggesting that cells that underwent EMT (M-SCCs) remained mesenchymal although Twist1-activation was withdrawn.
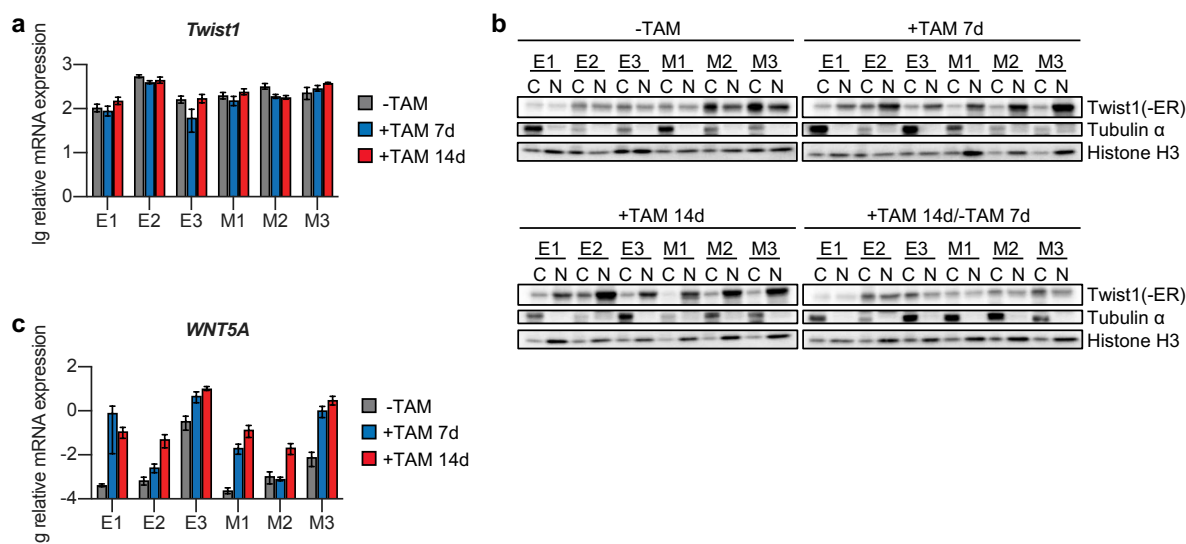
Figure 4-3: Expression and nuclear localisation of Twist1(-ER) and its target gene expression were similar in E-SCCs and M-SCCs

Parts of the results in this figure have been published in a pre-print at bioRxiv (Eichelberger et al., 2020).

a) Lg relative *Twist1* mRNA levels in E-SCCs and M-SCCs, before (-TAM) or at day 7 (+TAM 7d) or day 14 of Tamoxifen treatment (+TAM 14d). n=3. Mean $\pm$ SEM. **b)** Immunoblot of Twist1(-ER), Tubulin α, and Histone H3 in nuclear (N) and cytoplasmic (C) protein fractions of E-SCCs and M-SCCs, before (-TAM), after 7 days (+TAM 7d) or 14 days of Tamoxifen treatment (+TAM 14d), or after 14 days of Tamoxifen treatment followed by withdrawal for 7 days (+TAM 14d/-TAM 7d). **c)** Lg relative *WNT5A* mRNA levels in E-SCCs and M-SCCs treated as described in a). n=3. Mean $\pm$ SEM.

Although this analysis revealed that Twist1 was activated similarly in E-SCCs and M-SCCs upon Tamoxifen treatment (Figure 4-3b), I wished to prove further that Twist1 activity did not differ between E-SCCs and M-SCCs. For this purpose, I analysed the expression of the direct Twist1 target gene *WNT5A* (Wnt Family Member 5A, Shi et al., 2014) by qPCR. Although differences in expression levels of *WNT5A* transcript were observed between all SCCs, no trend towards higher expression levels in M-SCCs than E-SCCs was detected (Figure 4-3c), further indicating that Twist1 activity was similar in all SCCs.

Taken together, these findings showed that Twist1 expression, localisation as well as activity were comparable in both E-SCCs and M-SCCs. In addition, these results suggested that EMT-

resistance and EMT-competence did not arise due to differential Twist1-activation. This finding, in turn, indicated that the heterogeneity in terms of EMT might be regulated by differential activation of factors or pathways downstream of Twist1.

4.4 EMT-resistant cells acquired an epithelial-mesenchymal hybrid state during Twist1-activation

The previous results showed that a subset of HMLE-Twist1-ER bulk cells resisted Twist1-mediated EMT, whereas the other subset of cells underwent trans-differentiation upon activation of Twist1. Moreover, isolation of single-cell clones showed that E-SCCs and M-SCCs represented EMT-resistant and EMT-competent cells, respectively. Importantly, exclusively cells that resisted Twist1-mediated EMT (EPCAM^{pos}-sorted cells and E-SCCs) were competent to form epithelial organoids in a three-dimensional environment after transient Twist1-activation.

Based on these observations, I wished to unravel the determinants of EMT-resistance of E-SCCs and EMT-competence of M-SCCs. Nevertheless, I first set out to characterise both subsets of SCCs in more detail, as the previous investigations focused on determining EMT only by morphology and EPCAM expression. For this purpose, I assessed transcript and protein levels of typical EMT markers in both E-SCCs and M-SCCs at different time points of Twist1-activation by qPCR, immunoblot and flow cytometric analysis.

Analysis of the expression levels of the epithelial gene *EPCAM* by qPCR and flow cytometry revealed that, even before Twist1-activation (-TAM), *EPCAM* transcript, as well as protein expression levels, were higher in E-SCCs compared to M-SCCs ($p_{adj}<0.05$, Figure 4-4a, b). Of note, upon Twist1-activation (+TAM 7d, +TAM 14d), *EPCAM* transcript and protein decreased only marginally in E-SCCs (Figure 4-4a, b). In contrast to that, upon Twist1-activation in M-SCCs, *EPCAM* transcript levels decreased to a significantly lower level compared to E-SCCs (+TAM 7d, +TAM 14d, $p_{adj}<0.005$, Figure 4-4a), which resulted in a complete loss of EPCAM protein (Figure 4-4b).

Besides EPCAM, CDH1 is an important epithelial cell-cell junction protein, and its loss is associated with EMT (Dongre & Weinberg, 2019; Geiger & Peeper, 2009; Thiery et al., 2009). Analysis of CDH1 transcript and protein levels of E-SCCs and M-SCCs at different timepoints of Twist1-activation revealed that expression levels were similar in untreated E-SCCs and M-SCCs (-TAM, Figure 4-4a, c). However, upon Twist1-activation (+TAM 7d, +TAM 14d), *CDH1* transcript levels were decreased exclusively in M-SCCs (Figure 4-4a), resulting in loss of CDH1 protein expression (Figure 4-4c). By contrast, E-SCCs maintained expression of *CDH1* despite activation of Twist1 (+TAM 7d, +TAM 14d, Figure 4-4a, c).

These results showed that upon Twist1-activation, M-SCCs lost the expression of the epithelial markers *CDH1* and *EPCAM*, whereas E-SCCs maintained their expression.
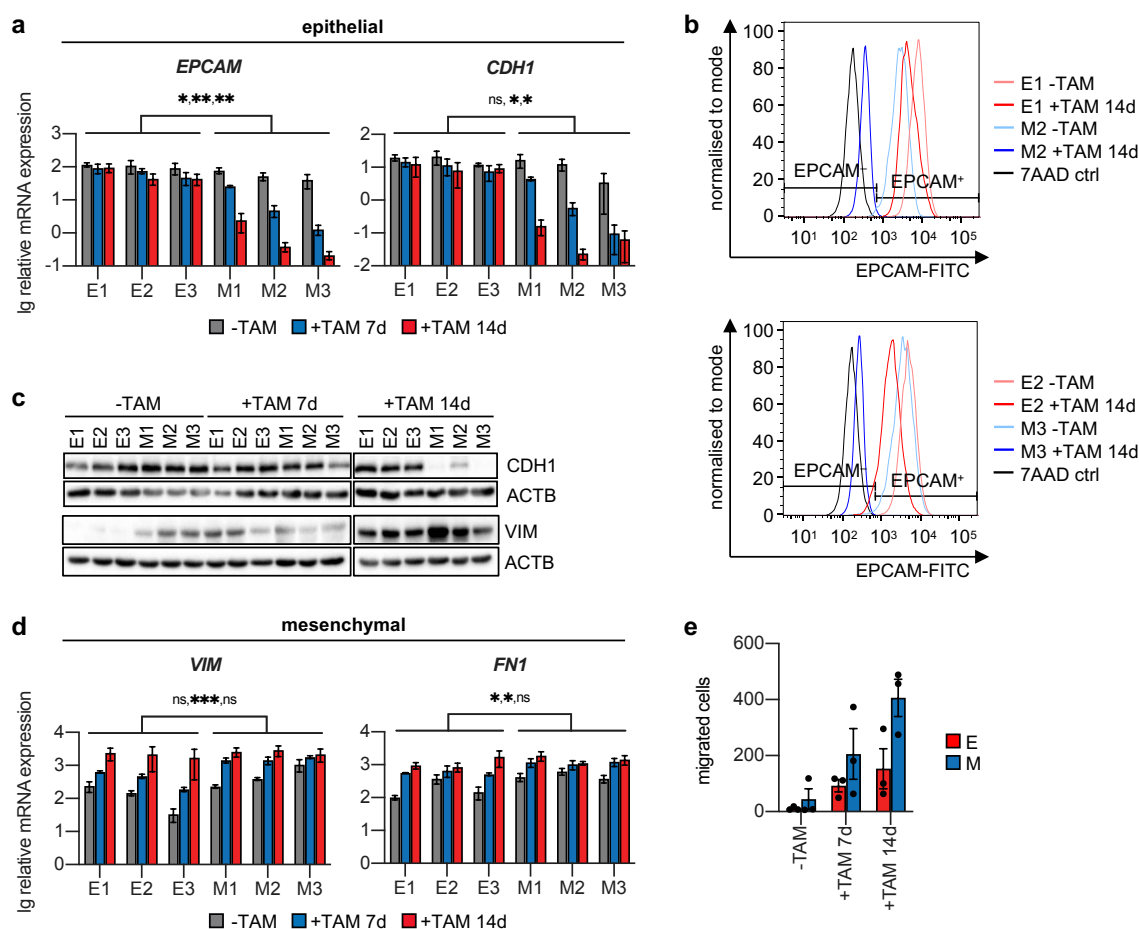
Figure 4-4: EMT-resistant cells maintained epithelial gene expression despite the upregulation of mesenchymal gene expression during Twist1-activation

Parts of the results in this figure have been published in a pre-print at bioRxiv (Eichelberger et al., 2020).

a) Lg relative *EPCAM* and *CDH1* mRNA expression levels in E-SCCs and M-SCCs, before (-TAM) or at day 7 (+TAM 7d) or day 14 of Tamoxifen treatment (+TAM 14d). n=3. Mean $\pm$ SEM. Multiple t-testing with Holm-Sidak correction. P_{adj}-values: $*<0.05$, $**<0.005$, ns: not significant. **b)** Flow cytometric analysis of EPCAM of E-SCCs and M-SCCs, before (-TAM) or at day 14 of Tamoxifen treatment (+TAM 14d). Gates were set according to unstained control. ctrl: control. **c)** Immunoblot of CDH1, VIM, and ACTB (actin beta) of E-SCCs and M-SCCs treated as described in a). **d)** Lg relative *VIM* and *FN1* mRNA expression levels in E-SCCs and M-SCCs treated as described in a). n=3. Mean $\pm$ SEM. Multiple t-testing with Holm-Sidak correction. p_{adj}-values: $*<0.05$, $***<0.0005$, ns: not significant. **e)** Quantification of migrated E-SCCs and M-SCCs treated as described in a). n=3. Mean $\pm$ SEM.

As EMT is not only defined by loss of epithelial markers but also by the upregulation of mesenchymal markers like Vimentin and Fibronectin (Dongre & Weinberg, 2019; Geiger & Peeper, 2009; Thiery et al., 2009), I investigated the expression of these markers in E-SCCs and M-SCCs by qPCR and immunoblot. This analysis revealed that Vimentin (*VIM*) transcript

levels were similar in untreated E-SCCs and M-SCCs (-TAM, Figure 4-4d). However, immunoblot analysis indicated slightly higher levels of VIM protein in M-SCCs compared to E-SCCs before Twist1-activation (-TAM, Figure 4-4c). In addition, analysis of Fibronectin (*FN1*) transcript revealed significantly higher levels in untreated M-SCCs compared to E-SCCs (-TAM, $p_{adj}<0.05$, Figure 4-4d). When Twist1 was activated continuously for one week (+TAM 7d), M-SCCs revealed significantly higher expression levels of both *VIM* ($p_{adj}<0.0005$) and *FN1* ($p_{adj}<0.05$, Figure 4-4c, d). However, after two weeks of Twist1-activation (+TAM 14d), expression levels of both *FN1* and *VIM* were comparable in E-SCCs and M-SCCs (Figure 4-4c, d), indicating that two weeks of Twist1-activation resulted in high expression levels of mesenchymal markers in both M-SCCs and E-SCCs.

In line with this observation, migration assays revealed that M-SCCs and E-SCCs enhanced their migratory behaviour upon Twist1-activation (+TAM 7d, +TAM 14d, Figure 4-4e). Nevertheless, after 14 days of Twist1-activation (+TAM 14d), the number of migrated cells in the trans-well assay was 3-fold higher for M-SCCs than E-SCCs (Figure 4-4e), indicating that M-SCCs developed a stronger migratory behaviour compared to E-SCCs.

Taken together, these results showed that upon activation of the EMT-transcription factor Twist1, only M-SCCs underwent complete EMT, indicated by the repression of epithelial markers and the upregulation of mesenchymal gene expression as well as migratory traits. By contrast, although Twist1-activation in E-SCCs triggered the upregulation of mesenchymal gene expression as well as their migratory behaviour, these cells concomitantly maintained the expression of epithelial genes. Interestingly, the kind of cell state that E-SCCs developed upon Twist1-activation has been postulated by others as a hybrid epithelial-mesenchymal state (Kröger et al., 2019; Pastushenko et al., 2018; Williams et al., 2019).

4.5 Transient Twist1-activation in EMT-competent cells resulted in stable trans-differentiation to a mesenchymal cell state

The previous results showed that upon Twist1-activation, only M-SCCs fully trans-differentiated to a mesenchymal state, whereas E-SCCs acquired an epithelial-mesenchymal state (Figure 4-4). Besides that, EPCAM[neg]-sorted, mesenchymal HMLE-Twist1-ER cells also displayed a mesenchymal state after a transient activation of Twist1 (Figure 4-1d-f). Moreover, the investigation of Twist1 protein levels in nuclear and cytoplasmic fractions showed that EMT irreversibility of M-SCCs was not regulated by the failure of Twist1 de-localisation after Tamoxifen withdrawal (Figure 4-3b).

Based on these observations, I hypothesised that transient Twist1-activation followed by a longer withdrawal period could allow the reversion of M-SCCs to an epithelial state. At the

same time, I wished to investigate whether E-SCCs maintain the acquired epithelial-mesenchymal hybrid state when Twist1-activation was withdrawn for a longer time.

For these purposes, previously treated (+TAM 14d) E-SCCs and M-SCCs were propagated in Tamoxifen-free conditions for one, two, and three weeks (+TAM 14d/-TAM 7d, +TAM 14d/-TAM 14d, +TAM 14d/-TAM 21d).

Doing so revealed that despite long-term Tamoxifen withdrawal and concomitant removal of Twist1-activation (+TAM 14d/-TAM 7d, +TAM 14d/-TAM 14d, +TAM 14d/-TAM 21d), M-SCCs maintained a mesenchymal, spindle-like morphology (Figure 4-5a). In addition, qPCR analyses of epithelial and mesenchymal markers showed that M-SCCs maintained low expression levels of epithelial genes, including *CDH1* (Figure 4-5b, Figure 4-4a) and high expression levels of mesenchymal genes, like *VIM* and *FN1* (Figure 4-5b, Figure 4-4d).

These results indicated that after transient activation of Twist1 and concomitant EMT, M-SCCs persisted in a fully mesenchymal state despite the withdrawal of Twist1-activation for up to three weeks.
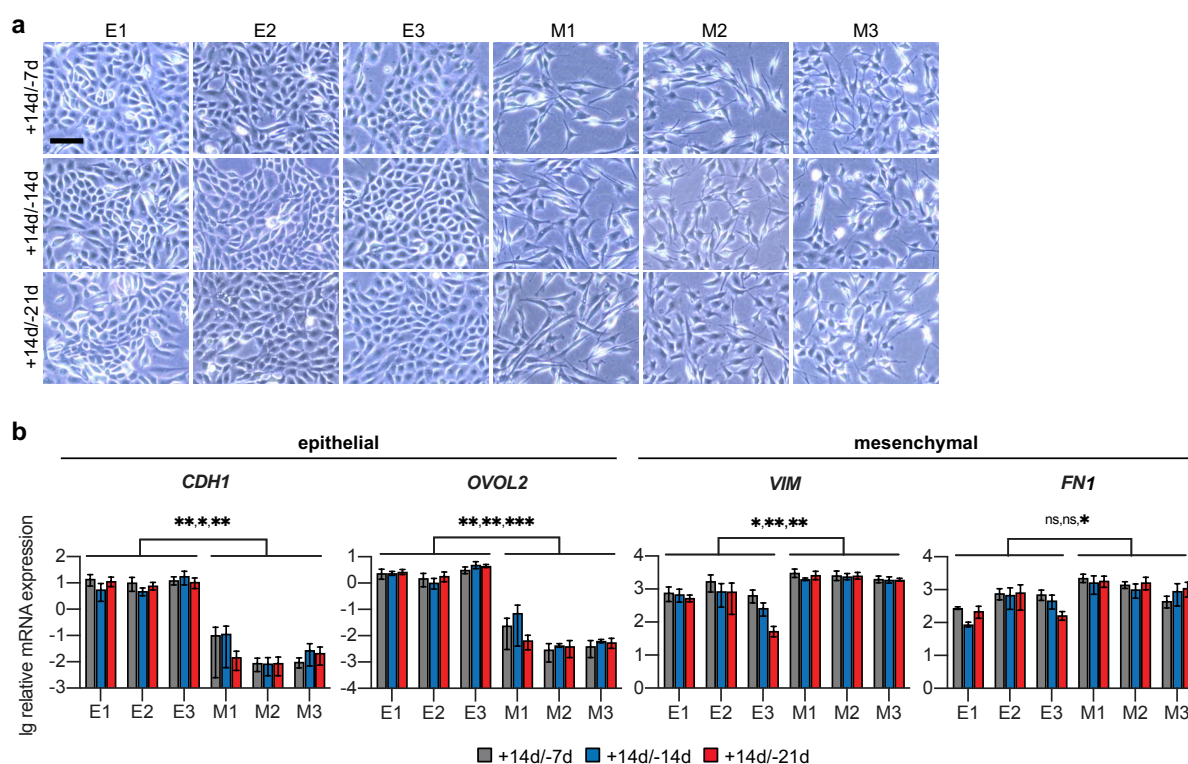
Figure 4-5: EMT-competent cells maintained a mesenchymal state after transient Twist1-activation

The results in this figure have been published in a pre-print at bioRxiv (Eichelberger et al., 2020).

a) Bright-field images of E-SCCs and M-SCCs, treated with Tamoxifen for 14 days followed by Tamoxifen withdrawal for 7 days (+14d/-7d), 14 days (+14d/-14d), or 21 days (+14d/-21d). Scale bar: 100 µm. **b)** Lg relative *CDH1*, *OVOL2*, *VIM*, and *FN1* mRNA expression levels in E-SCCs and M-SCCs treated as described in a). n=3. Mean ± SEM. Multiple t-testing with Holm-Sidak correction. p_{adj}-values: *<0.05, **<0.005, ***<0.0005, ns: not significant.

In contrast to M-SCCs, previously treated E-SCCs did not maintain an epithelial-mesenchymal hybrid cell state when propagated in Tamoxifen-free conditions but instead reverted to a state similar to untreated E-SCCs. In detail, upon Tamoxifen withdrawal and concomitant Twist1 deactivation (+TAM 14d/-TAM 7d, +TAM 14d/-TAM 14d, +TAM 14d/-TAM 21d), expression levels of mesenchymal markers, like *VIM* and *FN1* decreased again (Figure 4-5b, Figure 4-4d). These findings suggested that changes acquired by E-SCCs during Twist1-activation (+TAM) were reverted when Twist1-activation was withdrawn.

Taken together, these findings showed that M-SCCs that have previously undergone Twist1-mediated EMT remained in a mesenchymal state despite the long-term withdrawal of Twist1-activation. This finding suggested that transient activation of the EMT-TF Twist1 resulted in an irreversible, stable trans-differentiation from an epithelial to a mesenchymal state.

Moreover, the obtained results pointed out that E-SCCs could reverse the previously (+TAM) induced upregulation of mesenchymal gene expression. The observation that E-SCCs could revert the temporarily acquired epithelial-mesenchymal state further indicated that E-SCCs harboured some degree of epithelial plasticity.

Overall, the results in this study obtained so far revealed that upon transient activation of the EMT-TF Twist1, M-SCCs stably acquired mesenchymal and strong migratory traits. Nevertheless, this resulted in the loss of colonising capacity in a three-dimensional environment. E-SCCs, on the contrary, acquired an epithelial-mesenchymal state and some degree of migratory ability only temporarily, which allowed the colonisation in a three-dimensional environment.

Based on all these findings, I hypothesized that within a tumour, a subset of cells (M-SCCs) could undergo complete EMT and trigger the breakthrough of the basement membrane. As these cells cannot form epithelial metastases, I hypothesised that tumour cells that resist full trans-differentiation and transiently acquire an epithelial-mesenchymal hybrid state (E-SCCs) form epithelial metastases at distant sites.

4.6 Twist1-activation induced differential changes in chromatin accessibility that determined EMT-resistance and EMT-competence

Based on the previous observations, I hypothesized that M-SCCs represent tumour cells that break through the tumour basement membrane and enable the formation of metastases by cells that E-SCCs represent. Therefore, my next aim was to unravel the mechanistic basis that defines EMT-resistance and EMT-competence of E-SCCs and M-SCCs, respectively.

I hypothesised that EMT-resistance and EMT-competence were regulated on a chromatin level. To investigate this, I set out to analyse the genome-wide chromatin accessibility of E-

SCCs and M-SCCs before (-TAM), during (+TAM 7d and +TAM 14d), and after transient Twist1-activation (+TAM 14d/-TAM 7d).

For this purpose, ATAC-sequencing, a method used to detect accessible and inaccessible chromatin regions (Buenrostro et al., 2013), was performed in collaboration with Helena Domínguez Moreno and Gunnar Schotta. With this analysis, I wished to unravel whether a few or a broad range of differences evolved during transient Twist1-activation in E-SCCs and M-SCCs.

First, the global changes of chromatin accessibility during Twist1-activation were analysed. For that purpose, a principal component analysis (PCA) was performed with the obtained ATAC-sequencing results from E-SCCs and M-SCCs at all different time points (Figure 4-6a). This PCA revealed that untreated (-TAM) E-SCCs and M-SCCs clustered close to each other, indicating that both E-SCCs and M-SCCs harboured a similar chromatin accessibility landscape before Twist1-activation (-TAM, Figure 4-6a). Upon Twist1-activation (+TAM 7d and +TAM 14d), changes in chromatin accessibility occurred in both E-SCCs and M-SCCs, as indicated by shifts on the PCA axes (Figure 4-6a). Interestingly, the most remarkable shift on the PC1 axis was apparent for M-SCCs after seven and 14 days of Twist1-activation (51% variance), indicating that Twist1-activation induced a broad range of changes in M-SCCs. In E-SCCs, on the contrary, changes were observed primarily on the PC2 axis harbouring only a minor variance (33%), indicating that changes induced by Twist1-activation were less pronounced in E-SCCs (+TAM 7d and +TAM 14d, Figure 4-6a). When Twist1-activation was withdrawn (+TAM 14d/-TAM 7d), E-SCCs clustered closely to untreated E-SCCs, indicating that the previously developed changes in chromatin accessibility of E-SCCs were reversed, and the cells displayed a similar chromatin landscape as before Twist1-activation (Figure 4-6a). In contrast, changes of chromatin accessibility in M-SCCs were reversed only marginally upon Tamoxifen withdrawal (+TAM 14d/-TAM 7d), as indicated by a slight shift on the PCA axes (Figure 4-6a).

Together, these results showed that before Twist1-activation, chromatin accessibility was comparable in E-SCCs and M-SCCs. During Twist1-activation, a broad range of changes in chromatin accessibility occurred in E-SCCs and M-SCCs. More precisely, a wide range of changes could be observed during Twist1-activation in M-SCCs, whereas E-SCCs displayed only minor changes. Interestingly, upon withdrawal of Twist1-activation, the previously observed changes were fully reversed exclusively in E-SCCs.

Besides analysing changes in chromatin accessibility, I also aimed to understand the correlations between chromatin accessibility and expression levels better. For this purpose, the transcriptome of E-SCCs and M-SCCs was assessed by RNA-sequencing, which was performed in collaboration with Elisabeth Graf, Sandy Lösecke, and Thomas Schwarzmayr.
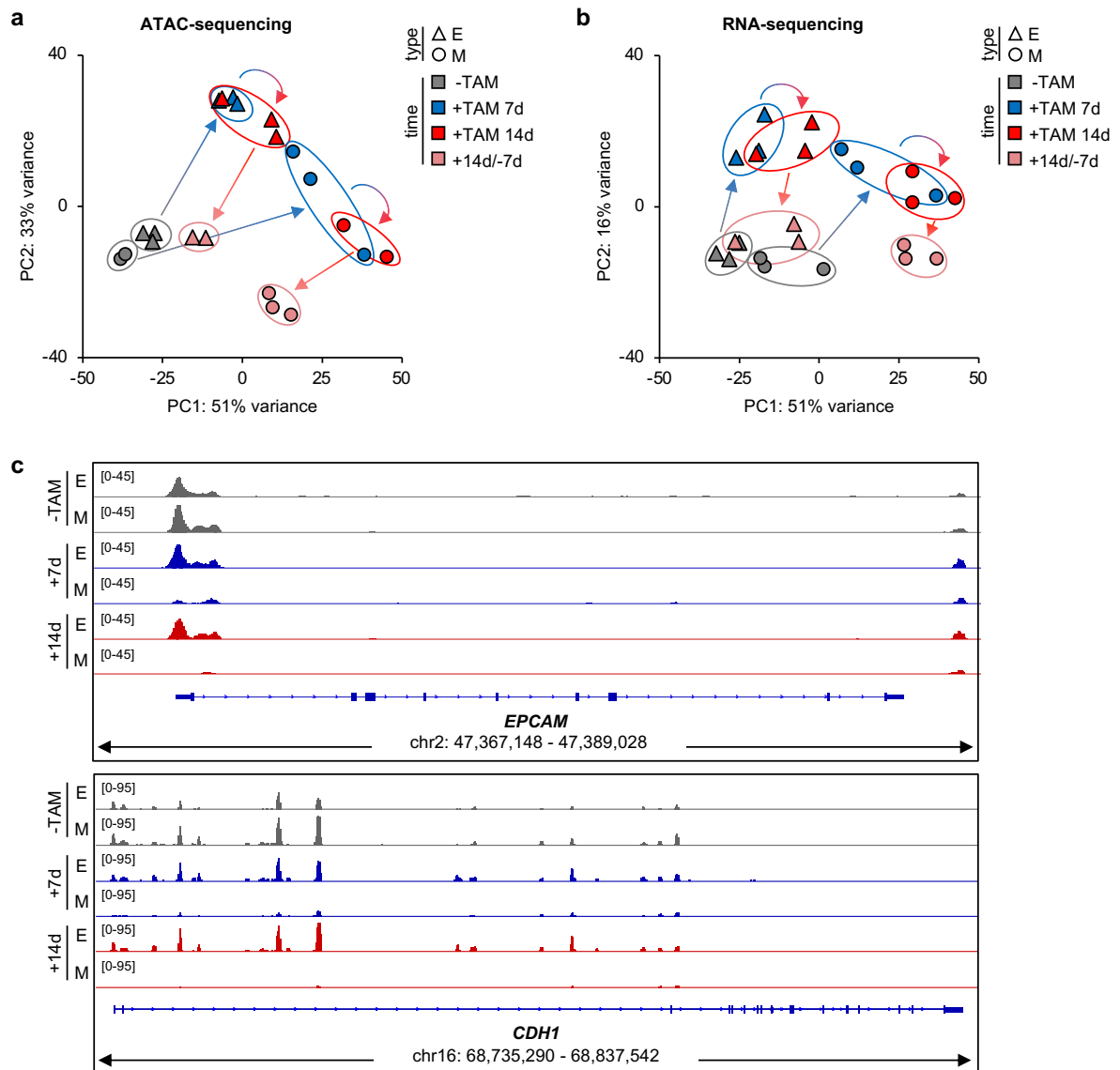
Figure 4-6: Differential changes in chromatin accessibility during Twist1-activation defined EMT-resistance and EMT-competence

The results in this figure have been published in a pre-print at bioRxiv (Eichelberger et al., 2020).

a) Principal component (PC) analysis of ATAC-sequencing results of E-SCCs (△, E) and M-SCCs (○, M), before (-TAM), at day 7 (+TAM 7d) or day 14 of Tamoxifen treatment (+TAM 14d), or after 14 days of Tamoxifen treatment followed by Tamoxifen withdrawal for 7 days (+TAM 14d/-TAM 7d). Each data point represents one SCC at the indicated time point. **b)** Principal component (PC) analysis of RNA-sequencing results of E-SCCs (△, E) and M-SCCs (○, M) treated as described in a). **c)** Genome browser high-resolution screenshot of ATAC-sequencing peaks of *EPCAM* and *CDH1* of one representative E-SCC (E) and one representative M-SCC (M), before (-TAM) or at day 7 (+7d) or day 14 of Tamoxifen treatment (+14d).

To compare ATAC- and RNA-sequencing results, a principal component analysis (PCA) was performed for the RNA-sequencing results as well. Strikingly, the PCA of RNA-sequencing results (Figure 4-6b) revealed similar dynamics as the PCA of the ATAC-sequencing results (Figure 4-6a). This finding suggested that changes in chromatin accessibility were reflected on

the transcript level and vice versa. To verify this observation, I analysed the chromatin accessibility of genes like *CDH1* and *EPCAM* in more detail. Previous analyses of *CDH1* and *EPCAM* transcript and protein levels revealed decreased expression levels upon Twist1-activation exclusively in M-SCCs (Figure 4-4a-c). When focusing on ATAC-sequencing peaks of these genes (Figure 4-6c), peaks of open chromatin were detected in both untreated (-TAM) E-SCCs and M-SCCs. Upon Tamoxifen treatment and concomitant Twist1-activation (+7d, +14d), open chromatin peaks disappeared exclusively in M-SCCs (Figure 4-6c). This observation showed that chromatin accessibility was lost exclusively in M-SCCs. Furthermore, this finding proved that changes in chromatin accessibility resulted in changes in transcript and protein expression levels.

Taken together, these results showed that upon transient Twist1-activation, chromatin was slightly remodelled in E-SCCs, which in turn lead to changes in gene expression. However, when Twist1-activation was removed in these cells, changes were reversed. Of note, this observation was in line with the previous finding of a temporary upregulation of mesenchymal markers in E-SCCs during transient Twist1-activation (Figure 4-4d, Figure 4-5b). In contrast, transient Twist1-activation resulted in major changes in chromatin accessibility and, in turn, gene expression in M-SCCs. However, these changes were only slightly reversed when Twist1-activation was ceased. These results corresponded with the previous observation that Twist1-activation repressed epithelial genes in M-SCCs and resulted in an irreversible trans-differentiation of these cells (Figure 4-4a, b, Figure 4-5b).

In conclusion, these results indicated that a broad range of differential changes in chromatin accessibility occurred in M-SCCs and E-SCCs during transient Twist1-activation. Moreover, these results suggested that the maintenance of chromatin accessibility of epithelial genes could be an important mechanism of E-SCCs to resist Twist1-mediated EMT.

4.7 ATAC-sequencing revealed ZEB and GRHL proteins as potential determinants of irreversible EMT in EMT-competent cells

The previous results indicated that chromatin accessibility was changed globally in E-SCCs and M-SCSs upon Twist1-activation via Tamoxifen treatment. Nevertheless, changes in E-SCCs were not as pronounced as in M-SCCs. In addition, developed changes in chromatin accessibility were fully reversible exclusively in E-SCCs when Twist1-activation was withdrawn.

Therefore, my next aim was to investigate how the differences in chromatin accessibility of E-SCCs and M-SCCs evolved. I hypothesised that TFs might regulate differential chromatin accessibility or, rather, differential expression of genes. To investigate this hypothesis, TF-

motif analyses were performed on the ATAC-sequencing results. First, the ATAC-sequencing results of E-SCCs and M-SCCs were clustered using peak clustering (Figure 4-7a).
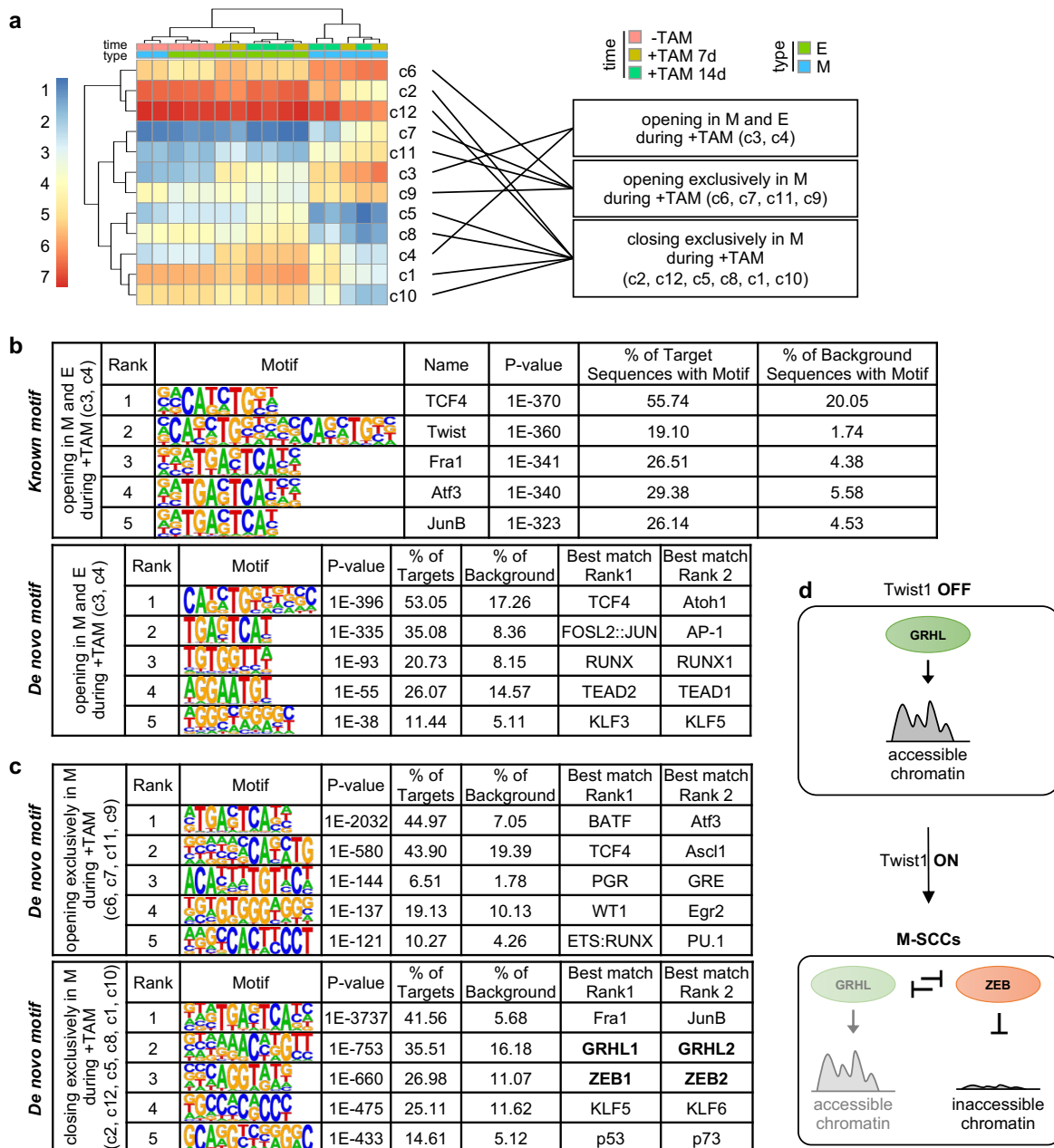
Doing so revealed two main clusters: a cluster formed by M-SCCs during Twist1-activation (+TAM 7d, +TAM 14d, Figure 4-7a) and another cluster that sub-clustered into two additional clusters. These two additional sub-clusters were formed, on the one hand, by untreated E-SCCs and M-SCCs (-TAM) and, on the other hand, by E-SCCs during Twist1-activation (+TAM7d, +TAM 14d, Figure 4-7a).

These results indicated that the chromatin accessibility in M-SCCs during Twist1-activation was very distinct to the accessibility in E-SCCs during Twist1-activation and to both E-SCCs and M-SCCs before Twist1-activation.

A closer inspection of the peak clustering revealed three different types of changes in chromatin accessibility during Twist1-activation (+TAM, Figure 4-7a): regions of chromatin in M-SCCs and E-SCCs with an increase in accessibility upon Twist1-activation (clusters c3 and c4), chromatin regions with an increase in accessibility exclusively in M-SCCs (clusters c6, c7, c11, and c9), and regions of chromatin with a decrease in accessibility only in M-SCCs but not in E-SCCs (clusters c2, c12, c5, c8, c1, and c10, Figure 4-7a).

For further analysis steps, the single clusters were combined according to the type of change in chromatin accessibility (Figure 4-7a). As I wished to identify specific TF motifs, the combined clusters were further analysed using the *Homer* TF motif analysis. Doing so revealed different TF-motifs for each cluster type (Figure 4-7b, c). First of all, *known motif* and *de novo motif* TF analysis of chromatin regions with increased accessibility in both E-SCCs and M-SCCs during Twist1-activation revealed motifs for several TFs (c3, c4, Figure 4-7b). For example, a Twist1 motif was detected in chromatin regions that showed an increase in accessibility when Twist1 was activated, indicating that Twist1 itself could bind at these loci and activate transcription either directly or indirectly (Figure 4-7b). Moreover, the same analysis showed a motif for TCF4 (transcription factor 4), a basic helix-loop-helix TF which was shown to interact with TWIST1 (Huttlin et al., 2017; Kotlyar et al., 2015). In addition, the TF analyses of these chromatin regions (c3, c4) revealed further binding motifs for the TFs Fra1 (also known as Fosl1), Atf3, and JunB (Figure 4-7b), which are members of the FOS, ATF, and JUN family, respectively. Proteins belonging to these three families can dimerize and form the activator protein-1 (AP-1) transcription factor complex, which is involved in several biological processes like differentiation or cell proliferation, depending on the proteins that form the dimer (Shaulian & Karin, 2002). Moreover, the same TF analysis revealed a binding motif for TEAD (Figure 4-7b), a TF that has already been shown to interact with AP-1 and YAP/TAZ to promote tumour growth of mammary epithelial cells (Zanconato et al., 2015), hinting towards the activation of cancer-related mechanisms in E-SCCs and M-SCCs upon Twist1-activation.

a

time: -TAM, +TAM 7d, +TAM 14d
type: E, M

opening in M and E during +TAM (c3, c4)

opening exclusively in M during +TAM (c6, c7, c11, c9)

closing exclusively in M during +TAM (c2, c12, c5, c8, c1, c10)

b

Known motif — opening in M and E during +TAM (c3, c4)

Rank	Motif	Name	P-value	% of Target Sequences with Motif	% of Background Sequences with Motif
1		TCF4	1E-370	55.74	20.05
2		Twist	1E-360	19.10	1.74
3		Fra1	1E-341	26.51	4.38
4		Atf3	1E-340	29.38	5.58
5		JunB	1E-323	26.14	4.53

De novo motif — opening in M and E during +TAM (c3, c4)

Rank	Motif	P-value	% of Targets	% of Background	Best match Rank1	Best match Rank 2
1		1E-396	53.05	17.26	TCF4	Atoh1
2		1E-335	35.08	8.36	FOSL2::JUN	AP-1
3		1E-93	20.73	8.15	RUNX	RUNX1
4		1E-55	26.07	14.57	TEAD2	TEAD1
5		1E-38	11.44	5.11	KLF3	KLF5

c

De novo motif — opening exclusively in M during +TAM (c6, c7, c11, c9)

Rank	Motif	P-value	% of Targets	% of Background	Best match Rank1	Best match Rank 2
1		1E-2032	44.97	7.05	BATF	Atf3
2		1E-580	43.90	19.39	TCF4	Ascl1
3		1E-144	6.51	1.78	PGR	GRE
4		1E-137	19.13	10.13	WT1	Egr2
5		1E-121	10.27	4.26	ETS:RUNX	PU.1

De novo motif — closing exclusively in M during +TAM (c2, c12, c5, c8, c1, c10)

Rank	Motif	P-value	% of Targets	% of Background	Best match Rank1	Best match Rank 2
1		1E-3737	41.56	5.68	Fra1	JunB
2		1E-753	35.51	16.18	**GRHL1**	**GRHL2**
3		1E-660	26.98	11.07	**ZEB1**	**ZEB2**
4		1E-475	25.11	11.62	KLF5	KLF6
5		1E-433	14.61	5.12	p53	p73

d

Twist1 **OFF**

GRHL → accessible chromatin

Twist1 **ON**

M-SCCs

GRHL ⊣⊢ ZEB

GRHL → accessible chromatin

ZEB ⊣ inaccessible chromatin

Figure 4-7: Clustering and transcription factor motif analysis revealed differences between EMT-resistant and EMT-competent cells during Twist1-activation

The results in this figure have been published in a pre-print at bioRxiv (Eichelberger et al., 2020).

a) Heatmap of chromatin accessibility of 12 clusters of ATAC-sequencing peaks of E-SCCs (E) and M-SCCs (M), before (-TAM) or at day 7 (+TAM 7d) or day 14 of Tamoxifen treatment (+TAM 14d). **b)** Top 5 Homer *known motif* (upper table) and *de novo motif* (lower table) transcription factor motif analysis of grouped clusters opening in both M-SCCs (M) and E-SCCs (E) upon Tamoxifen treatment (+TAM). **c)** Top 5 hits of Homer *de novo motif* transcription factor motif analysis of grouped clusters opening exclusively in M-SCCs (M, upper table) or closing exclusively in M-SCCs (M, lower table) upon Tamoxifen treatment (+TAM). **d)** Schematic depiction of how ZEB and GRHL proteins could regulate chromatin accessibility before (Twist1 OFF) and during Twist1-activation (Twist1 ON) in M-SCCs.

Together, these results showed that in chromatin regions that revealed an increase in accessibility upon Twist1-activation in both E-SCCs and M-SCCs, a range of motifs for different TFs were present. The increased accessibility at these regions suggested that the TFs that bind to these motifs could activate transcription of these genes either directly or indirectly. The presence of a Twist1 motif within the chromatin regions that showed increased accessibility upon Twist1-activation in both E-SCCs and M-SCCs indicated that Twist1 was similarly active in E-SCCs and M-SCCs. This finding, in turn, underpinned the hypothesis that the differential behaviour of E-SCCs and M-SCCs with regard to EMT is rather caused by Twist1 downstream factors than Twist1 itself.

To unravel which factors could regulate the differential EMT-response of E-SCCs and M-SCCs, clusters that displayed differential changes in chromatin accessibility in E-SCCs and M-SCCs were analysed. In detail, Homer *de novo motif* analyses were performed on chromatin regions with an increase in accessibility (clusters c6, c7, c11, and c9) or with a decrease in accessibility (clusters c2, c12, c5, c8, c1, and c10) in M-SCCs but not E-SCCs upon Twist1-activation (Figure 4-7a, c). Interestingly, the TF motif analysis of these chromatin regions again included motifs for AP-1 members (BATF/Atf3 and Fra1/JunB, Figure 4-7c).

This finding indicated that upon Twist1-activation, AP-1 could further change the expression of additional chromatin regions in M-SCCs either directly or indirectly. Of note, the accessibility of these regions was not changed upon Twist1-activation in E-SCCs.

In chromatin regions where the accessibility decreased exclusively in M-SCCs during Twist1-activation (c2, c12, c5, c8, c1, c10), *de novo* TF analysis revealed motifs for KLF5/KLF6 and p53/p63/p73 (Figure 4-7c). As TP53 is inhibited in HMLE-Twist1-ER cells due to expression of the SV40 large T, it is more likely that TP63 and TP73 were related to the motif in this context. As mentioned earlier, the motifs for the TFs KLF5/KLF6 and p63/p73 were found in chromatin regions of M-SCCs where the accessibility was reduced upon Twist1-activation. Based on that, two possibilities could exist for how these TFs regulated gene expression. Either these TFs directly reduced the accessibility of these regions, or these TFs usually act as activators but were pushed aside as the chromatin accessibility was decreased upon Twist1-activation. Interestingly, KLF5 and TP63/TP73 have already been shown to inhibit EMT via activating the expression of micro RNAs from the miR-200 family (Knouf et al., 2012; Zhang et al., 2013). Moreover, in a study that has investigated the EMT transition states in skin squamous cell carcinoma, KLF5 was also associated with an epithelial state (Pastushenko et al., 2018). Of note, the chromatin accessibility in regions containing motifs for these specific TFs was reduced in M-SCCs upon Twist1-activation.

This result suggested that in M-SCCs, these TFs acted as activators in the epithelial cell state but could no longer bind to these chromatin regions when Twist1 was activated, which led to the loss of expression of genes associated with an epithelial state exclusively in M-SCCs.

In addition, in chromatin regions that showed a decrease in accessibility in M-SCCs upon Twist1-activation, the TF analysis revealed motifs for the EMT-TFs ZEB1/ZEB2 and GRHL1/GRHL2 (Figure 4-7c). These findings were particularly the focal point of my attention, as ZEB and GRHL proteins are prominent factors implicated in the regulation of EMT.

ZEB1, for example, is a TF downstream of Twist1 and is implicated in EMT, as it represses the expression of epithelial genes like *CDH1* (Sánchez-Tilló et al., 2010) and *EPCAM* (Vannier et al., 2013) and additionally regulates *GRHL2* expression in a double negative feedback-loop (Cieply et al., 2012; Werner et al., 2013). For grainy-head like factors (GRHL1-3), it is known that they can prime enhancers of epithelial genes and thereby regulate the activation of transcription (Jacobs et al., 2018). That implies that GRHL proteins have the opposite effect of ZEB proteins. More precisely, ZEB proteins repress the expression of epithelial genes, whereas GRHL proteins enhance the expression of epithelial genes.

As motifs for both ZEB and GRHL proteins were found in clusters with reduced chromatin accessibility exclusively in M-SCCs upon Twist1-activation (Figure 4-7c), I assumed that these factors could regulate the transition from an epithelial to a mesenchymal state in M-SCCs but not E-SCCs. I hypothesised that during Twist1-activation, Twist1 could trigger the expression of ZEB proteins in M-SCCs (Figure 4-7d). In these cells, ZEB could, in turn, repress the expression of GRHL proteins and, as a result, GRHL proteins could no longer prime epithelial enhancers (Figure 4-7d). That, in turn, could facilitate the ZEB-mediated repression of epithelial genes and the transition of M-SCCs from an epithelial to a mesenchymal state.

Taken together, the TF motif analysis of chromatin regions with changing accessibility upon Twist1-activation revealed that binding motifs of several TFs were implicated.

In chromatin regions with increased accessibility upon Twist1-activation in both E-SCCs and M-SCCs, a motif for Twist1 was found, which supported the finding that Twist1 itself regulated gene expression in E-SCCs and M-SCCs similarly. However, this further indicated that factors or mechanisms downstream of Twist1 regulated whether cells underwent or resisted EMT. Analysis of TF motifs in chromatin regions that changed accessibility upon Twist1-activation in M-SCCs only revealed factors including ZEB and GRHL proteins. These results hinted towards a role of these factors in regulating the differential EMT-behaviour of E-SCCs and M-SCCs during Twist1-activation.

4.8 Twist1-activation triggered differential *ZEB1* expression levels in EMT-resistant and EMT-competent cells

As the previous results hinted towards a role of GRHL and ZEB proteins in the regulation of EMT-resistance and EMT-competence, I set out to analyse *GRHL* and *ZEB* expression levels in E-SCCs and M-SCCs during Twist1-activation. For that purpose, I used the RNA-

sequencing data obtained from E-SCCs and M-SCCs at different time points of transient Twist1-activation.

Analysis of FPKMs (Fragments Per Kilobase Million) revealed that before Twist1-activation (-TAM), transcripts of *GRHL1*, *GRHL2*, and *GRHL3* were detected in E-SCCs and M-SCCs (Figure 4-8a). However, upon Twist1-activation (+7d and +14d), *GRHL3* transcript levels were reduced in both E-SCCs and M-SCCs (Figure 4-8a), indicating that *GRHL3* was not implicated in the differential EMT-response of E-SCCs and M-SCCs. In contrast, *GRHL1* and *GRHL2* transcripts were differentially regulated in E-SCCs and M-SCCs upon Twist1-activation. E-SCCs maintained expression of both GRHL1 and GRHL2, whereas expression levels of both genes were reduced in M-SCCs upon Twist1-activation (+7d and +14d, Figure 4-8a).

Of note, a strong difference was especially recognisable for the expression levels of *GRHL2*. In untreated conditions (-TAM), E-SCCs and M-SCCs expressed the highest levels of *GRHL2*, whereas Twist1-activation (+7d, +14d) resulted in downregulated expression levels of *GRHL2* in M-SCCs only. In contrast, E-SCCs maintained the expression of *GRHL2* despite activation of the EMT-TF Twist1 (Figure 4-8a).

Analysis of *ZEB* transcript levels showed that *ZEB* was expressed oppositely (Figure 4-8b). In detail, *ZEB1* and *ZEB2* transcript levels were low in E-SCCs and M-SCCs before Twist1-activation (-TAM) but were upregulated during Twist1-activation (+7d, +14d, Figure 4-8b). Remarkably, especially expression levels of *ZEB1* were strongly upregulated upon Twist1-activation in E-SCCs and M-SCCs. Besides that, I observed differential expression dynamics in E-SCCs and M-SCCs (Figure 4-8b). In detail, ZEB1 FPKM levels in E-SCCs after 14 days of Twist1-activation (+14d) were comparable to levels of M-SCCs after seven days (+7d) of Twist1-activation (Figure 4-8b). Moreover, in M-SCCs, ZEB1 transcript levels increased even more when Twist1 was activated for additional seven days (+14d, Figure 4-8b).

Besides the observation of differential *ZEB1* expression dynamics in E-SCCs and M-SCCs upon Twist1-activation, ZEB1 is a well-known EMT-TF. Therefore, I set out to analyse the expression of *ZEB1* in more detail. In addition to the RNA-sequencing analysis, I investigated *ZEB1* expression by qPCR and immunoblot (Figure 4-8c, d). Doing so revealed a similar trend as observed in the FPKM analysis. M-SCCs expressed high *ZEB1* transcript levels already after seven days of Twist1-activation (+TAM 7d, Figure 4-8c). After 14 days of Tamoxifen treatment, an increase in expression levels was also detected in E-SCCs (+TAM 14d, Figure 4-8c). However, at this time point, *ZEB1* expression levels in M-SCCs were significantly higher than those in E-SCCs (Figure 4-8c).

Interestingly, analysis of ZEB1 protein revealed similar results (Figure 4-8d). Before Twist1-activation, no ZEB1 protein was detected in E-SCCs and M-SCCs (-TAM, Figure 4-8d). However, after 7 days of Twist1-activation, ZEB1 protein was detected in M-SCCs but not E-SCCs (+TAM 7d, Figure 4-8d). On day 14 of Twist1-activation, ZEB1 protein was detected in

both E-SCCs and M-SCCs. However, at this time point, ZEB1 protein levels were higher in M-SCCs than E-SCCs (+TAM 14d, Figure 4-8d).
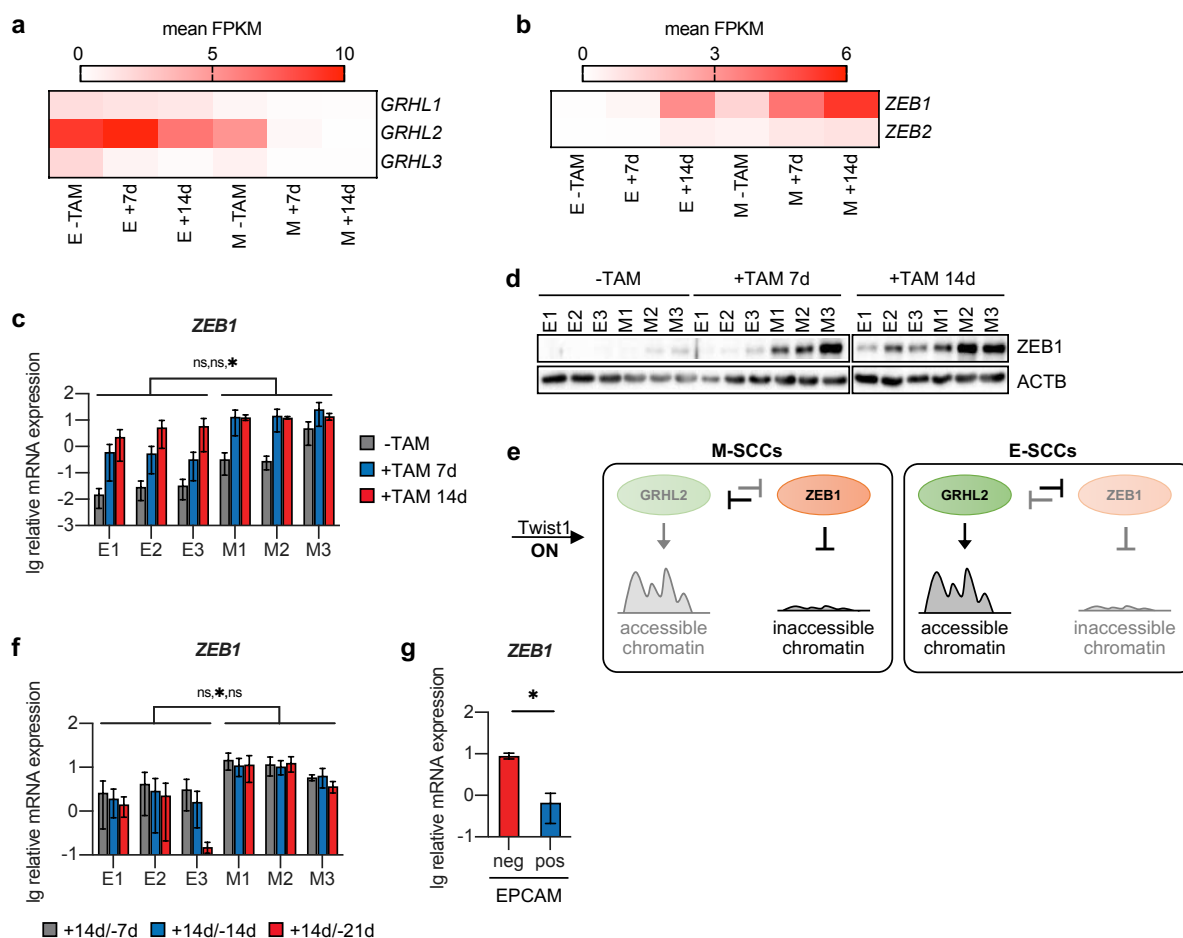
Figure 4-8: EMT-resistant and EMT-competent cells expressed different *ZEB1* levels upon Twist1-activation

The results in this figure have been published in a pre-print at bioRxiv (Eichelberger et al., 2020).

a) Heatmap of mean *GRHL1*, *GRHL2*, and *GRHL3* FPKM values of RNA-sequencing data of E-SCCs (E) and M-SCCs (M), before (-TAM), or at day 7 (+7d), or day 14 day of Tamoxifen treatment (+14d). n=3. **b)** Heatmap of mean *ZEB1* and *ZEB2* FPKM values of RNA-sequencing data of E-SCCs (E) and M-SCCs (M), treated as described in a). n=3. **c)** Lg relative *ZEB1* mRNA expression levels in E-SCCs and M-SCCs treated as described in a. n=3. Mean $\pm$ SEM. Multiple t-testing with Holm-Sidak correction. p_{adj}-values: ∗<0.05, ns: not significant. **d)** Immunoblot of ZEB1 and ACTB of E-SCCs and M-SCCs treated as described in a). **e)** Schematic depiction of how GRHL2 and ZEB1 regulate chromatin accessibility in M-SCCs and E-SCCs during Twist1-activation (Twist1 ON). **f)** Lg relative *ZEB1* mRNA expression levels in E-SCCs and M-SCCs, treated with Tamoxifen for 14 days followed by Tamoxifen withdrawal for 7 days (+14d/-7d), 14 days (+14d/-14d), or 21 days (+14d/-21d). n=3. Mean $\pm$ SEM. Multiple t-testing with Holm-Sidak correction. p_{adj}-values: ∗<0.05, ns: not significant. **g)** Lg relative *ZEB1* mRNA expression levels in sorted EPCAM[pos] (pos) and EPCAM[neg] (neg) HMLE-Twist1-ER cells, treated with Tamoxifen for 21 days followed by Tamoxifen withdrawal for ten days (+TAM21d/-TAM10d).

Taken together, these results indicated that GRHL2 and ZEB1 were part of the differential response of E-SCCs and M-SCCs upon Twist1-activation with regard to EMT. In detail, upon Twist1-activation in M-SCCs, these cells expressed high levels of *ZEB1* transcript and protein and repressed the expression of *GRHL2* entirely. However, upon Twist1-activation in E-SCCs, these cells expressed lower *ZEB1* transcript and protein levels and maintained the expression of *GRHL2*.

Based on these findings, I hypothesised that upon Twist1-activation, high ZEB1 levels in M-SCCs repressed the expression of *GRHL2*, which resulted in reduced chromatin accessibility of epithelial genes. In contrast, lower ZEB1 levels in E-SCCs were not sufficient to repress *GRHL2*, which in turn allowed the maintenance of chromatin accessibility and expression of epithelial genes (Figure 4-8e). Besides that, this hypothesis involved the assumption that the expression of high ZEB1 levels is essential for EMT. Previous results showed that cells that have undergone a Twist1-mediated EMT remained in a mesenchymal state, despite the withdrawal of Twist1-activation.

Therefore, I wished to analyse the *ZEB1* expression levels of cells treated with Tamoxifen to activate Twist1 followed by Tamoxifen withdrawal to stop Twist1-activation (+TAM/-TAM). I performed qPCR to detect *ZEB1* transcript levels in E-SCCs and M-SCCs (Figure 4-8f) and both subsets of EPCAM-sorted HMLE-Twist1-ER cells (Figure 4-8g) after transient Twist1-activation.

This analysis showed that after transient activation of Twist1 followed by Twist1 deactivation, M-SCCs expressed higher levels of *ZEB1* compared to E-SCCs (Figure 4-8f). Moreover, a similar trend was observed for EPCAM-sorted cells. After transient activation of Twist1 (+TAM 21d/-TAM 10d) in HMLE-Twist1-ER cells, EPCAM^{neg}-sorted cells displayed significantly higher *ZEB1* transcript levels than EPCAM^{pos}-sorted cells (Figure 4-8g).

These results indicated that despite the withdrawal of the EMT-stimulus, high expression levels of *ZEB1* were maintained in cells that have previously undergone an EMT.

Together, these results showed that upon Twist1-activation, M-SCCs expressed higher *ZEB1* transcript and protein levels than E-SCCs. In M-SCCs, high levels of ZEB1 resulted in repression of *GRHL2* expression and, in turn, allowed the repression of epithelial gene expression. However, in E-SCCs, the increase in ZEB1 levels was insufficient to efficiently repress *GRHL2* and epithelial gene expression. Moreover, in cells that have previously undergone a Twist1-mediated EMT, high *ZEB1* expression levels were maintained after transient Twist1-activation. Based on these observations, I hypothesised that a self-enforcing ZEB1 expression mechanism might have been induced in M-SCCs upon Twist1-activation, which determined the irreversibility of EMT.

4.9 ZEB1 targets were repressed exclusively in EMT-competent cells

The previous results showed that *ZEB1* was differentially expressed in M-SCCs and E-SCCs upon Twist1-activation. In detail, M-SCCs expressed higher levels of *ZEB1* transcript and protein upon Twist1-activation. Moreover, high *ZEB1* gene expression levels were maintained in cells that have undergone an EMT, although Twist1-activation was withdrawn. Besides that, during Twist1-activation, repression of the epithelial gene *GRHL2* occurred only in cells with high ZEB1 levels (M-SCCs). Based on these findings, my next aim was to investigate further known downstream targets of ZEB1 to unravel whether their expression levels are changed accordingly to changes in ZEB1 expression levels (Figure 4-9a).
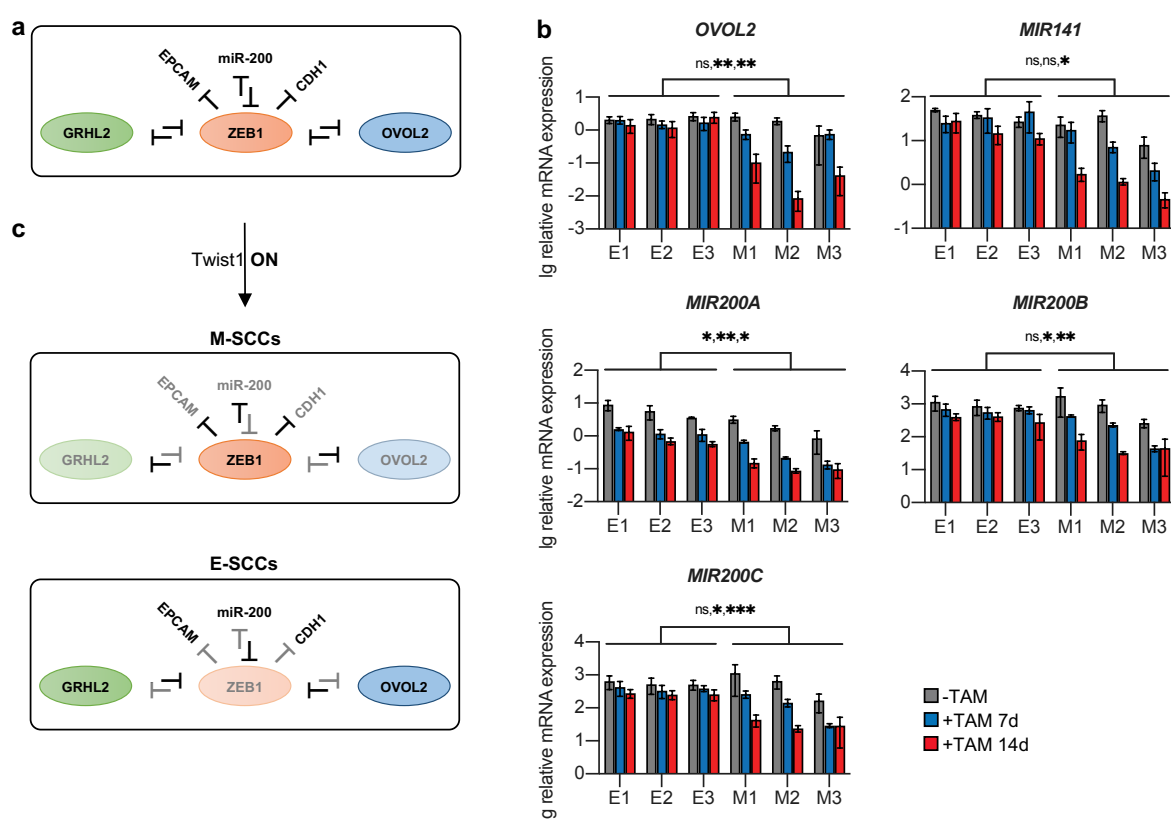
Figure 4-9: ZEB1 downstream targets were repressed in M-SCCs but not E-SCCs

The results in this figure have been published in a pre-print at bioRxiv (Eichelberger et al., 2020).

a) Schematic depiction of the regulatory network of ZEB1, GRHL2, OVOL2, miR-200, EPCAM, and CDH1. **b)** Lg relative *OVOL2* mRNA and *MIR141*, *MIR200A*, *MIR200B*, and *MIR200C* miRNA expression levels in E-SCCs and M-SCCs, before (-TAM), or at day 7 (+7d), or day 14 of Tamoxifen treatment (+14d). n=3. Mean $\pm$ SEM. Multiple t-testing with Holm-Sidak correction. p_{adj}-values: $*<0.05$, $**<0.005$, $***<0.0005$, ns: not significant. **c)** Schematic depiction of the regulatory network of ZEB1, GRHL2, OVOL2, miR-200, EPCAM, and CDH1 upon Twist1-activation (Twist1 ON) in M-SCCs (top) and E-SCCs (bottom).

Previous analyses in this study revealed a decrease in the expression levels of *CDH1* and *EPCAM* in M-SCCs but not E-SCCs when Twist1 was activated transiently (Figure 4-4a-c, Figure 4-5b). Both epithelial genes, *CDH1* and *EPCAM*, are repressed as ZEB1 binds to the promoter region (Eger et al., 2005; Sánchez-Tilló et al., 2010; Vannier et al., 2013). In addition to these two target genes, expression of the transcription factor ovo like zinc finger 2 (*OVOL2*) gene is also repressed by ZEB1. Of note, ZEB1 and OVOL2 can repress each other in a double-negative feedback loop by mutual binding to the promoter region (Hong et al., 2015; Watanabe et al., 2014).

The analysis of *OVOL2* expression levels by qPCR revealed E-SCCs and M-SCCs expressed similar transcript levels before Twist1-activation (-TAM, Figure 4-9b). However, upon Twist1-activation (+TAM 7d, +TAM 14d), *OVOL2* transcript levels decreased exclusively in M-SCCs but not E-SCCs. This decrease in M-SCCs resulted in significantly lower levels of *OVOL2* transcript in M-SCCs compared to E-SCCs ($p_{adj}<0.005$, Figure 4-9b).

Besides analysing *OVOL2* expression, I also investigated the expression levels of micro RNAs (miRNAs) from the miR-200 family (*MIR141*, *MIR200A*, *MIR200B*, and *MIR200C*). These miRNAs reduce the abundance of *ZEB1* transcript by binding to the three prime untranslated region (3'UTR) of the *ZEB1* messenger RNA (mRNA) (Gregory et al., 2008; Park et al., 2008). However, the expression of these miRNAs is also controlled by ZEB1. In detail, ZEB1 can repress the expression of miR-200 family members by binding to the promoter regions of the genes coding for these miRNAs (Bracken et al., 2008; Burk et al., 2008).

Analysis of the expression levels of the miR-200 family members by qPCR revealed that, except for *MIR200A*, the expression levels of each miRNA were similar in E-SCCs and M-SCCs before Twist1-activation (-TAM, Figure 4-9b). However, upon Twist1-activation (+TAM 7d, +TAM 14d), the expression levels decreased exclusively in M-SCCs, which resulted in significantly lower levels in M-SCCs compared to E-SCCs ($p_{adj}<0.0005$-0.05, Figure 4-9b).

Interestingly, the TF motif analysis of chromatin regions that displayed reduced accessibility upon Twist1-activation in M-SCCs only revealed factors (KLF5 and TP63/TP73, Figure 4-7c) that are known to activate the expression of miR-200 family members (Knouf et al., 2012; Zhang et al., 2013). Remarkably, when Twist1 was activated, these chromatin regions were less accessible exclusively in M-SCCs. This finding indicated that these TFs could no longer bind to *MIR200* genes, which results in loss of *MIR200* gene expression in M-SCCs. In addition, the same TF motif analysis also revealed a motif for ZEB1 (Figure 4-7c). As ZEB1 is known to repress the expression of the *MIR200* genes (Bracken et al., 2008; Burk et al., 2008), the reduced accessibility of chromatin regions that harboured a ZEB1 motif indicated that ZEB1 was involved in the repression of these genes. However, this occurred exclusively in M-SCCs when Twist1 was activated.

Together, these results indicated that upon Twist1-activation, *ZEB1* expression levels in M-SCCs were sufficient to repress downstream targets, like *CDH1* and *EPCAM*, as well as *ZEB1* repressors including miR-200 and OVOL2, but also GRHL2 (Figure 4-9c). E-SCCs, on the contrary, maintained the expression of the epithelial genes *CDH1* and *EPCAM* because Twist1-activation resulted in *ZEB1* expression levels that were not sufficient to successfully repress downstream targets as well as *ZEB1* repressors (Figure 4-9c).

4.10 EMT-resistance was maintained despite long-term Twist1-activation

The previous investigations showed that upregulation of *ZEB1* expression was delayed in E-SCCs and that *ZEB1* gene expression levels were lower in E-SCCs than M-SCCs. Based on these findings, I wondered whether a more extended period of Twist1-activation in E-SCCs could trigger higher expression levels of *ZEB1*, which in turn could repress epithelial genes, like *CDH1* and *OVOL2*. To investigate this hypothesis, I activated Twist1 in E-SCCs and M-SCCs by Tamoxifen treatment for 21 (+TAM 21d) and 28 days (+TAM 28d). Upon long-term Twist1-activation, E-SCCs showed slight changes in morphology; however, the epithelial cobblestone-like morphology was still visible (Figure 4-10a). In contrast, M-SCCs trans-differentiated from an epithelial to a mesenchymal phenotype (Figure 4-10a).
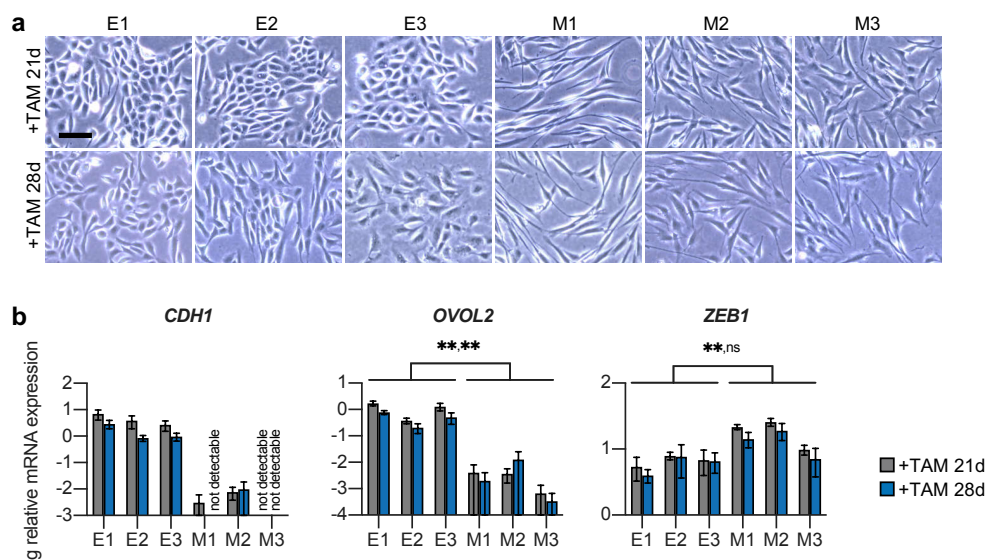
Figure 4-10: Long-term Twist1-activation did not induce EMT of EMT-resistant cells

The results in this figure have been published in a pre-print at bioRxiv (Eichelberger et al., 2020).

a) Bright-field images of E-SCCs and M-SCCs treated with Tamoxifen for 21 (+TAM 21d) or 28 days (+TAM 28d). Scale bar: 100 μm. **b)** Lg relative *CDH1*, *OVOL2*, and *ZEB1* expression levels in E-SCCs and M-SCCs treated as described in a). n=3. Mean ± SEM. Multiple t-testing with Holm-Sidak correction. p_{adj}-values: ∗∗<0.005, ns: not significant.

Moreover, analysis of *CDH1* and *OVOL2* transcript levels by qPCR revealed that the expression of both epithelial genes was repressed exclusively in M-SCCs, whereas E-SCCs maintained the expression of both *CDH1* and *OVOL* 2 (Figure 4-10b).

Interestingly, although significant only after 21 days of Twist1-activation (+TAM 21d), *ZEB1* expression levels were higher in M-SCCs compared to E-SCCs, despite long-term Twist1-activation (+TAM 21d, +TAM 28d, Figure 4-10b).

Taken together, these data showed that the activation of Twist1 for a longer period was not sufficient to induce high *ZEB1* expression levels in E-SCCs. Instead, *ZEB1* expression levels in E-SCCs were lower compared to those in M-SCCs. Consequently, E-SCCs maintained the expression of epithelial genes and persisted in an epithelial state. Furthermore, these findings indicated that exceeding a certain *ZEB1* expression threshold could play a crucial role for a successful and complete EMT. However, as these results suggested, *ZEB1* expression levels were high in M-SCCs only.

4.11 Twist1 binding within the *ZEB1 gene* is not essential for Twist1-mediated EMT

The previous results indicated that high expression levels of *ZEB1* were important for the successful trans-differentiation of M-SCCs from an epithelial to a mesenchymal state upon Twist1-activation. Interestingly, it has already been shown that Twist1 binds to a double E-box motif (CANNTG) spaced by exactly five nucleotides within the *ZEB1* (NM00128128) gene (Chang et al., 2015). Moreover, others showed that Tamoxifen treatment of HMLE-Twist1-ER bulk cells increased the amount of Twist1 binding to this specific chromatin region (Dragoi et al., 2016). In addition, the results in this study showed that long-term Twist1-activation was not sufficient to obtain high *ZEB1* expression levels and EMT in E-SCCs.

Based on these facts, I hypothesised that differential Twist1 binding to the *ZEB1* gene might determine the differential *ZEB1* expression levels in E-SCCs and M-SCCs and, in turn, control whether cells resist or undergo Twist1-mediated EMT, respectively.

To investigate this hypothesis, I first set out to address whether Twist1 binding to the double E-box motif within the *ZEB1* gene is necessary for the trans-differentiation of EMT-competent M-SCCs. For this purpose, I manipulated the Twist1 binding motif within the *ZEB1* gene of an M-SCC using a transient CRISPR/Cas9 approach (Breunig et al., 2018).

To obtain a range of different DNA modifications, I used two sets of guide RNAs (gRNAs). One set of gRNAs consisted of two gRNAs that spanned the Twist1 binding site (Figure 4-11a). To analyse the DNA modifications that were introduced by CRISPR/Cas9, single-cell clones (SCCs) were isolated, expanded, and screened. In detail, PCRs were performed to analyse the chromatin region that was targeted by CRISPR/Cas9. Doing so revealed a variety of DNA

modifications (Figure 4-11b). For further analyses, I used SCCs that harboured different lengths of DNA deletions. For example, I used two SCCs (*a1* and *a2*) obtained by performing CRISPR/Cas9 with a pair of gRNAs spaced by 90 nucleotides (Figure 4-11a). In SCC *a1*, the sequencing analysis showed that approximately half of the Twist1 binding motif was deleted. In detail, SCC *a1* still harboured one of the two E-boxes that define the Twist1 binding motif. In contrast, SCC *a2* still harboured the double E-box motif but revealed DNA deletions surrounding the Twist1 motif (Figure 4-11b).
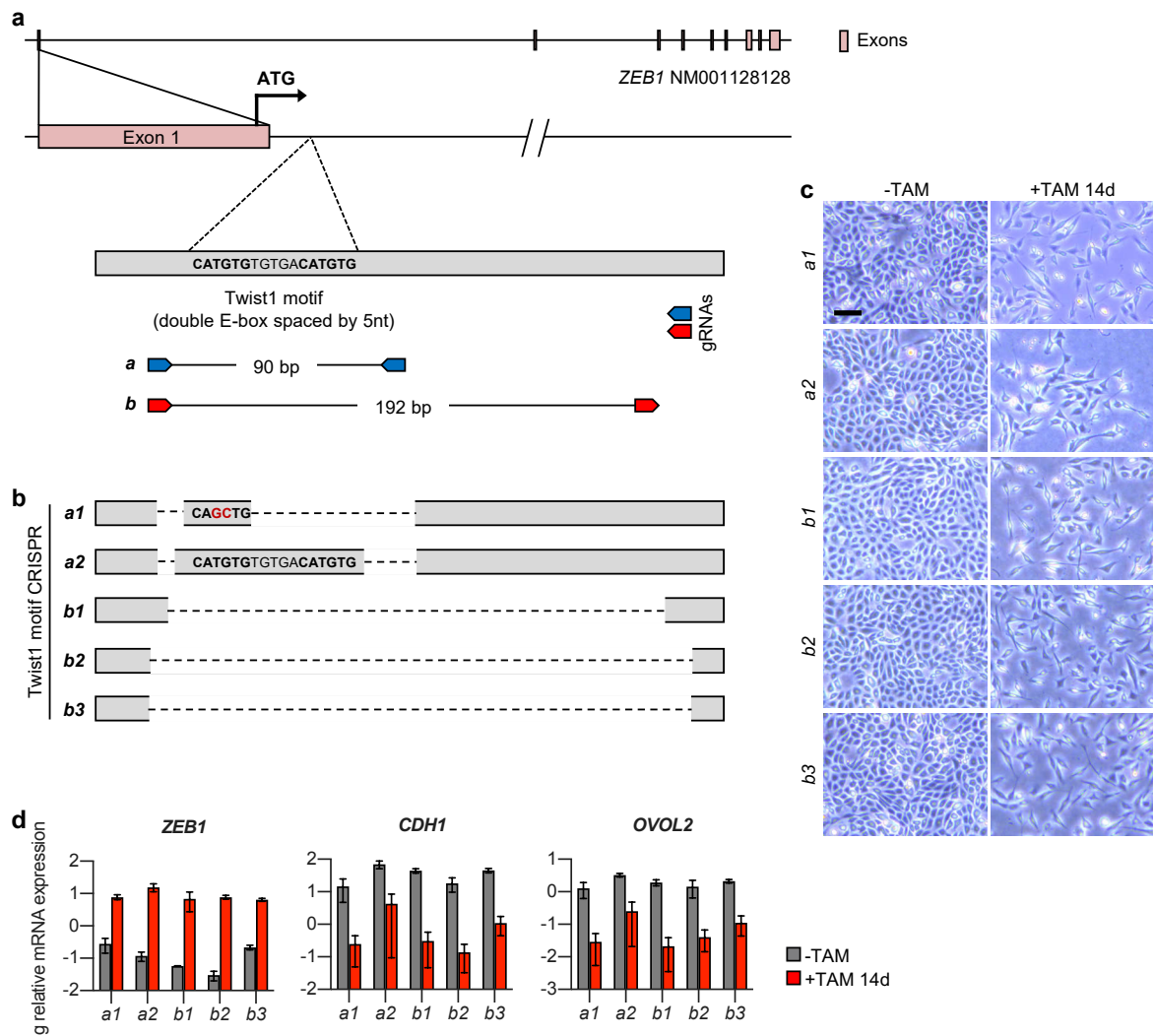
Figure 4-11: Deletion of the Twist1 binding motif within the *ZEB1* genes did not impact Twist1-mediated EMT of EMT-competent cells

a) Schematic depiction of the *ZEB1* (NM00128128) gene harbouring a Twist1 motif targeted with different guide RNAs (gRNAs). bp: base pairs. **b)** Schematic depiction of DNA modifications in different Twist1 binding motif CRISPR single-cell clones (SCCs: *a1-2*, *b1-3*). **c)** Bright-field images of Twist1 binding motif CRISPR SCCs (*a1-2* and *b1-3*), before (-TAM) or at day 14 of Tamoxifen treatment (+TAM 14d). Scale bar: 100 µm. **d)** Lg relative *ZEB1*, *CDH1*, and *OVOL2* expression levels in Twist1 binding motif CRISPR SCCs (*a1-2* and *b1-3*) treated as described in c). n=2. Mean ± SEM.

Using a pair of gRNAs spaced by 192 base pairs (Figure 4-11a) generated SCCs that harboured more extensive deletions. These deletions, in turn, resulted in a complete loss of the Twist1 binding motif in these SCCs (SCC *b1-3*, Figure 4-11b).

To analyse whether the deletion of the Twist1 binding motif within the *ZEB1* gene of an M-SCC could impact Twist1-mediated EMT, Twist1 was activated in all CRISPR/Cas9 SCCs by Tamoxifen treatment for 14 days (+TAM 14d). Upon doing so, all SCCs trans-differentiated from an epithelial to a mesenchymal morphology (Figure 4-11c), indicating that all SCCs underwent EMT. Besides, the expression levels of *ZEB1* increased in all SCCs upon Twist1-activation, regardless of whether the Twist1 binding motif was missing or intact (Figure 4-11d). Moreover, the expression levels of ZEB1 downstream targets *CDH1* and *OVOL2* decreased in all SCCs when Twist1 was activated (Figure 4-11d).

Taken together, these results indicated that deletion of the Twist1 binding motif within the *ZEB1* gene did not inhibit Twist1-mediated EMT of M-SCCs. At the same time, these results suggested that binding of Twist1 to the double E-box motif within the *ZEB1* gene was not essential for Twist1-mediated upregulation of *ZEB1* expression and EMT of M-SCCs.

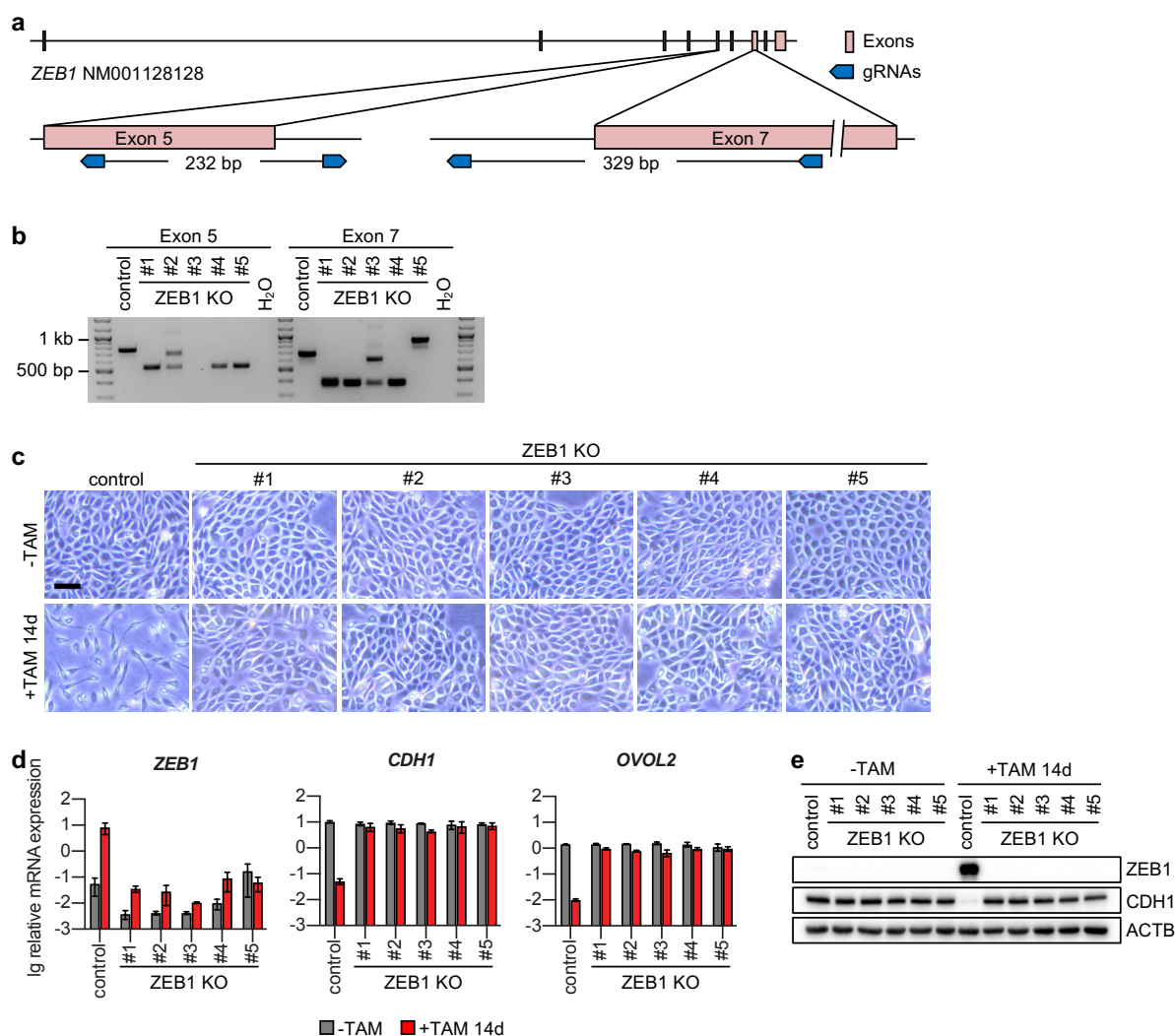## 4.12 Loss of *ZEB1* in EMT-competent cells inhibited Twist1-mediated EMT

As previously shown, ZEB1 was differentially expressed in E-SCCs and M-SCCs upon transient activation of Twist1. In detail, M-SCCs expressed higher ZEB1 levels compared to E-SCCs upon Twist1-activation. Moreover, previous results indicated that high ZEB1 levels in M-SCCs repressed epithelial genes and induced EMT.

To analyse this further, I set out to investigate whether upregulation of *ZEB1* expression is essential for the repression of epithelial genes, like *CDH1* and *OVOL2*, and the Twist1-mediated EMT of M-SCCs.

For this purpose, a *ZEB1* knockout was performed in an M-SCC by a transient CRISPR/Cas9 approach (Breunig et al., 2018). As *ZEB1* has nine different transcript variants, four gRNAs were used to ensure the deletion of all variants. Two gRNAs were paired, and targeted DNA regions within the *ZEB1* gene represented in all nine transcript variants. One pair of gRNAs was spaced by 232 base pairs and targeted Exon 5 of the *ZEB1* (NM00128128) gene. The second pair of gRNAs targeted Exon 7 of the *ZEB1* (NM00128128) gene, and these gRNAs were spaced by 329 base pairs (Figure 4-12a).

To analyse DNA modifications, *ZEB1* knockout (ZEB1 KO) SCCs were again isolated, expanded, and screened. For the latter, PCRs spanning the targeted regions were performed to determine changes in the length of the DNA sections. In wild type cells, the PCRs for Exon 5 and Exon 7 should generate products with a length of 769 and 693 base pairs, respectively.

Figure 4-12: Loss of *ZEB1* expression prevented Twist1-mediated EMT of EMT-competent cells

The results in this figure have been published in a pre-print at bioRxiv (Eichelberger et al., 2020).

a) Schematic depiction of the *ZEB1* (NM00128128) gene and paired guide RNAs (gRNAs) targeting Exon 5 and Exon 7. bp: base pairs. **b)** DNA fragments obtained by PCRs spanning the CRISPR/Cas9 targeted Exon 5 and 7 of M-SCC ZEB1 knockout single-cell clones (ZEB1 KO #1-5) and an M-SCC control clone. bp: base pairs. kb: kilobase. H_2O: water control. **c)** Bright-field images of M-SCC ZEB1 knockout single-cell clones (ZEB1 KO #1-5) and an M-SCC control clone, before (-TAM) or on day 14 of Tamoxifen treatment (+TAM 14d). Scale bar: 100 μm. **d)** Lg relative *ZEB1*, *CDH1*, and *OVOL2* mRNA expression levels of M-SCC ZEB1 knockout single-cell clones (ZEB1 KO #1-5) and an M-SCC control clone treated as described in c). n=2. Mean ± SEM. **e)** Immunoblot of ZEB1, CDH1, and ACTB of M-SCC ZEB1 knockout single-cell clones (ZEB1 KO #1-5) and an M-SCC control clone treated as described in c).

Indeed, analysis of PCR fragments obtained from the DNA of the control SCC showed that the length of the PCRs corresponded to the expected lengths of approximately 800 and 700 base pairs for Exon 5 and Exon 7, respectively (Figure 4-12b).

In contrast, analysis of the PCR fragments obtained from DNA of the *ZEB1* knockout SCCs revealed homozygous and heterozygous DNA modifications. In detail, fragment lengths that

were shorter or longer than those of control cells were detected, indicating that the CRISPR/Cas9 approach induced deletions (for example, Exon 5 of ZEB1 KO #1 and #4) as well as insertions of DNA (Exon 7 of ZEB1 KO #5, Figure 4-12b), respectively.

As I aimed to investigate whether knockout of *ZEB1* impacts Twist1-mediated EMT in M-SCCs, I activated Twist1 in all ZEB1 KO SCCs (#1-5) and the control SCC by treating the cells with Tamoxifen for 14 days (+TAM 14d). Doing so revealed that the control SCC, which still harboured the *ZEB1* wildtype gene, trans-differentiated from an epithelial morphology to a mesenchymal morphology (+TAM 14d, Figure 4-12c). In contrast, *ZEB1* KO SCCs did not undergo trans-differentiation but maintained an epithelial, cobblestone-like morphology despite the activation of Twist1 for 14 days (+TAM 14d, Figure 4-12c), indicating that *ZEB1* knockout in M-SCCs inhibited EMT.

To show that the CRISPR/Cas9 approach resulted in complete loss of *ZEB1* transcript and protein, all SCCs were analysed by qPCR and immunoblot before (-TAM) and after two weeks of Twist1-activation (+TAM 14d, Figure 4-12d, e). These analyses revealed that upon Twist1-activation (+TAM 14d), *ZEB1* transcript and protein were exclusively upregulated in the control SCC, whereas no ZEB1 protein was detectable in ZEB1 KO SCCs #1-5 (Figure 4-12d, e).

On this basis, I set out to analyse whether knockout of *ZEB1* impacts the expression of epithelial genes. Previously, I have shown that high levels of ZEB1 correlated with repression of *CDH1* and *OVOL2*. To investigate whether ZEB1 directly impacts the expression of both genes, I analysed *CDH1* and *OVOL2* transcript and protein levels in control and ZEB1 KO SCCs by qPCR (Figure 4-12d, e). This analysis showed that *OVOL2* and *CDH1* transcript levels decreased in the control SCC upon Twist1-activation, whereas ZEB1 KO SCCs maintained the expression of both epithelial genes despite the activation of Twist1 for 14 days (Figure 4-12d). Moreover, immunoblot analysis revealed that CDH1 protein was expressed in all SCCs before Twist1-activation (-TAM, Figure 4-12e). However, upon Twist1-activation, the expression of CDH1 was lost exclusively in the control SCC (+TAM 14d, Figure 4-12e). In contrast, CDH1 protein could be detected in all ZEB1 KO SCCs despite the activation of Twist1 for 14 days (+TAM 14d, Figure 4-12e).

Taken together, these results showed that knockout of *ZEB1* inhibited EMT of M-SCCs. That further indicated that *ZEB1* gene expression was essential for Twist1-mediated EMT of M-SCCs. Moreover, *ZEB1* knockout prevented the repression of epithelial genes, like *CDH1* and *OVOL2*, and inhibited the trans-differentiation from an epithelial to a mesenchymal cell state.

4.13 EMT-resistant and EMT-competent cells displayed morphological changes upon *ZEB1* overexpression

The previous experiments revealed that M-SCCs expressed higher *ZEB1* transcript and protein levels compared to E-SCCs when Twist1 was activated. That, in turn, allowed M-SCCs to repress epithelial genes and undergo EMT. Moreover, a *ZEB1* knockout experiment indicated that the induction of *ZEB1* expression upon Twist1-activation was essential for M-SCCs to undergo EMT.

Based on these findings, I hypothesised that *ZEB1* overexpression could trigger EMT in EMT-resistant E-SCCs. To investigate this, I set out to overexpress *ZEB1* in E-SCCs. For this purpose, a *TetOn-ZEB1-IRES2-mCherry* construct (Figure 4-13a) was introduced in an E-SCC using PiggyBac. Besides that, the same *ZEB1* construct was introduced in an M-SCC to obtain cells known to be able to undergo EMT. After doing so, SCCs were isolated and expanded.

My first aim was to investigate whether *ZEB1* can be overexpressed in the isolated SCCs (E-ZEB1 #1-4 and M-ZEB1 #1-2). For this purpose, the expression of *ZEB1* was induced by treating the SCCs (E-ZEB1 #1-4, M-ZEB1 #1-2) with doxycycline for one week (+1 µg/ml dox 7d). As a control, SCCs that harboured a doxycycline-inducible β-glucuronidase (*GUS*) construct (E-GUS #1-2, M-GUS #1-2) were also treated with doxycycline for one week (+1 µg/ml dox 7d). *GUS* is frequently used as a control gene, and its expression does not impact EMT. Moreover, to obtain an uninduced reference point, all SCCs were treated with DMSO simultaneously.

Analysis of transcript levels by qPCR revealed that *ZEB1* was upregulated in all E-ZEB1 and M-ZEB1 SCCs upon treatment with doxycycline for one week (+1 µg/ml dox 7d), compared to cells treated with DMSO only (Figure 4-13b). Importantly, E-GUS and M-GUS SCCs did not upregulate the expression of *ZEB1* expression when treated with doxycycline (+1 µg/ml dox 7d, Figure 4-13b). In theory, upon doxycycline treatment, cells containing the *ZEB1* construct should express the *ZEB1* RNA fused to the *IRES2* and *mCherry* RNA (Figure 4-13a). Therefore, the qPCR analysis of the *mCherry* transcript should provide evidence of whether the exogenous *ZEB1* is expressed. Doing so revealed that all E-ZEB1 and M-ZEB1 SCCs upregulated the expression of *mCherry* transcript upon treatment with doxycycline for one week (+1 µg/ml dox 7d), compared to cells treated with DMSO only (Figure 4-13b), indicating that exogenous *ZEB1* was overexpressed in all E-ZEB1 and M-ZEB1 SCCs upon doxycycline treatment. Besides *ZEB1* transcript levels, I also wished to analyse the expression and localisation of ZEB1 protein in doxycycline-induced E-ZEB1 and M-ZEB1 SCCs. Immunoblot analysis revealed that in SCCs treated with DMSO, no ZEB1 protein was detected (DMSO, Figure 4-13c). However, the abundance of ZEB1 protein increased in all E-ZEB1 and M-ZEB1 SCCs when these cells were treated with doxycycline (+1 µg/ml dox 7d, Figure 4-13c).
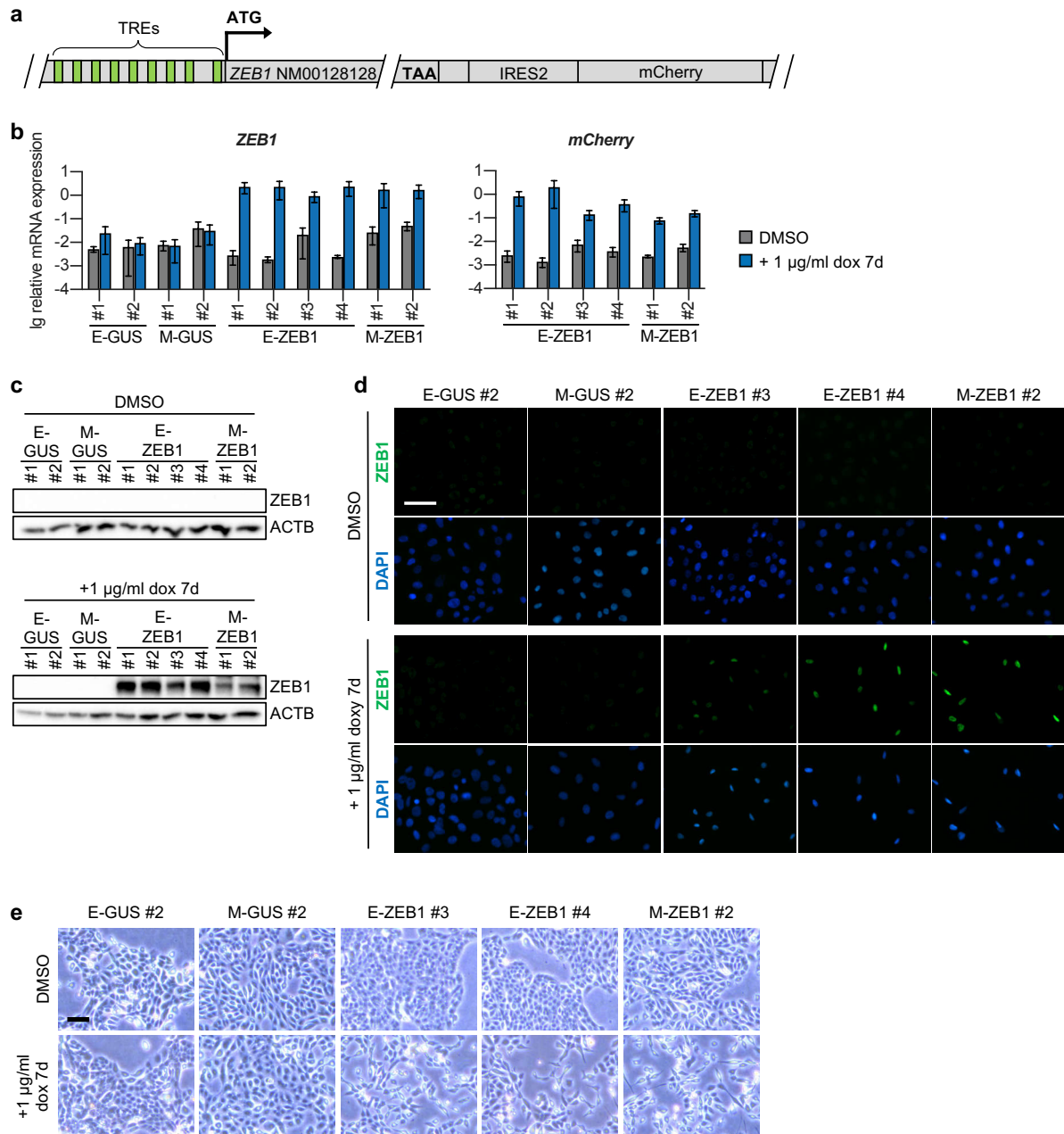
Figure 4-13: *ZEB1* overexpression induced morphological changes of both EMT-competent and EMT-resistant cells

a) Schematic depiction of the doxycycline-inducible *ZEB1* construct. TREs: Tetracycline response elements. ATG: Start codon. TAA: Stop codon. IRES2: Internal ribosome entry site type II. **b)** Lg relative *ZEB1* and *mCherry* mRNA expression levels of E-GUS #1-2 and M-GUS #1-2 and or E-ZEB1 #1-4 and M-ZEB1 #1-2 SCCs treated with DMSO (DMSO) or with 1 μg/ml doxycycline for 7 days (+1 μg/ml dox 7d). n=2. Mean ± SEM. **c)** Immunoblot of ZEB1 and ACTB of E-GUS #1-2, M-GUS #1-2, E-ZEB1 #1-4, and M-ZEB1 #1-2 SCCs treated as described in b). **d)** Images of immunofluorescent staining of ZEB1 (green) and DAPI (blue) of E-GUS #2, M-GUS #2, and or E-ZEB1 #3 and #4, and M-ZEB1 #1 SCCs treated as described in b). Scale bar: 50 μm. **e)** Bright-field images of E-GUS #2, M-GUS #2, and or E-ZEB1 #3 and #4, and M-ZEB1 #1 SCCs treated as described in b). Scale bar: 100 μm.

Importantly, immunofluorescent staining proved that upon doxycycline treatment (+1 µg/ml dox 7d), ZEB1 protein was localised in the nucleus of all cells of E-ZEB1 and M-ZEB1 SCCs (Figure 4-13d). Of note, in E-GUS and M-GUS SCCs, no ZEB1 protein was detected when these cells were treated with doxycycline (+1 µg/ml dox 7d, Figure 4-13c, d). These results indicated that doxycycline treatment induced the overexpression of *ZEB1* transcript and protein in E-ZEB1 and M-ZEB1 SCCs. Moreover, in these cells, the ZEB1 protein was located in the nucleus, where ZEB1 regulates the transcription of genes.

As these analyses revealed that *ZEB1* transcript and protein were successfully overexpressed in M-SCCs and E-SCCs upon doxycycline treatment, I set out to analyse whether both E-ZEB1 and M-ZEB1 SCCs maintain an epithelial, cobblestone-like morphology or trans-differentiate to a mesenchymal morphology.

For this purpose, the morphology of the cells was investigated before (DMSO) and after seven days of *ZEB1* induction (+1 µg/ml dox 7d). Doing so revealed that upon doxycycline treatment and concomitant induction of *ZEB1* overexpression (+1 µg/ml dox 7d), M-ZEB1 SCCs and E-ZEB1 SCCs changed from an epithelial to a more mesenchymal morphology (Figure 4-13e). Of note, M-GUS and E-GUS SCCs remained epithelial when treated with doxycycline (+1 µg/ml dox 7d, Figure 4-13e).

These results indicated that *ZEB1* overexpression triggered the trans-differentiation of M-ZEB1 and E-ZEB1 SCCs from an epithelial to a mesenchymal morphology.

Together, these results showed that *ZEB1* could be overexpressed in both E-ZEB1 and M-ZEB1 SCCs. Moreover, in these SCCs, ZEB1 localised within the nucleus, the site where it could act as a TF. Furthermore, both M-ZEB1 and E-ZEB1 SCCs underwent morphological changes, which suggested that overexpression of *ZEB1* initiated EMT in both EMT-competent M-SCCs and EMT-resistant E-SCCs. Importantly, these results indicated that EMT-resistance of E-SCCs could be eliminated by *ZEB1* overexpression.

4.14 *ZEB1* overexpression triggered the repression of target genes in both EMT-competent and EMT-resistant cells

Based on the previous finding that *ZEB1* overexpression triggered trans-differentiation from an epithelial to a mesenchymal phenotype in EMT-resistant cells, I wished to further analyse the effects of *ZEB1* overexpression in more detail. I first set out to unravel whether the overexpression of *ZEB1* induces similar changes in M-SCCs and E-SCCs or whether changes in M-SCCs differ from those in E-SCCs. In detail, I wished to analyse the chromatin accessibility by ATAC-sequencing to unravel whether differences between E-SCCs and M-SCCs evolve upon ZEB1 overexpression. For this purpose, E-ZEB1 #1-4 and M-ZEB1 #1-2 were treated with doxycycline for seven days (+1 µg/ml dox 7d) and analysed by ATAC-

sequencing. Previous observations indicated that E-GUS and M-GUS SCCs maintained an epithelial state when treated with doxycycline. Therefore, E-GUS #1-2 and M-GUS #1-2 treated with doxycycline for one week (+1 µg/ml dox 7d) were used as reference points.
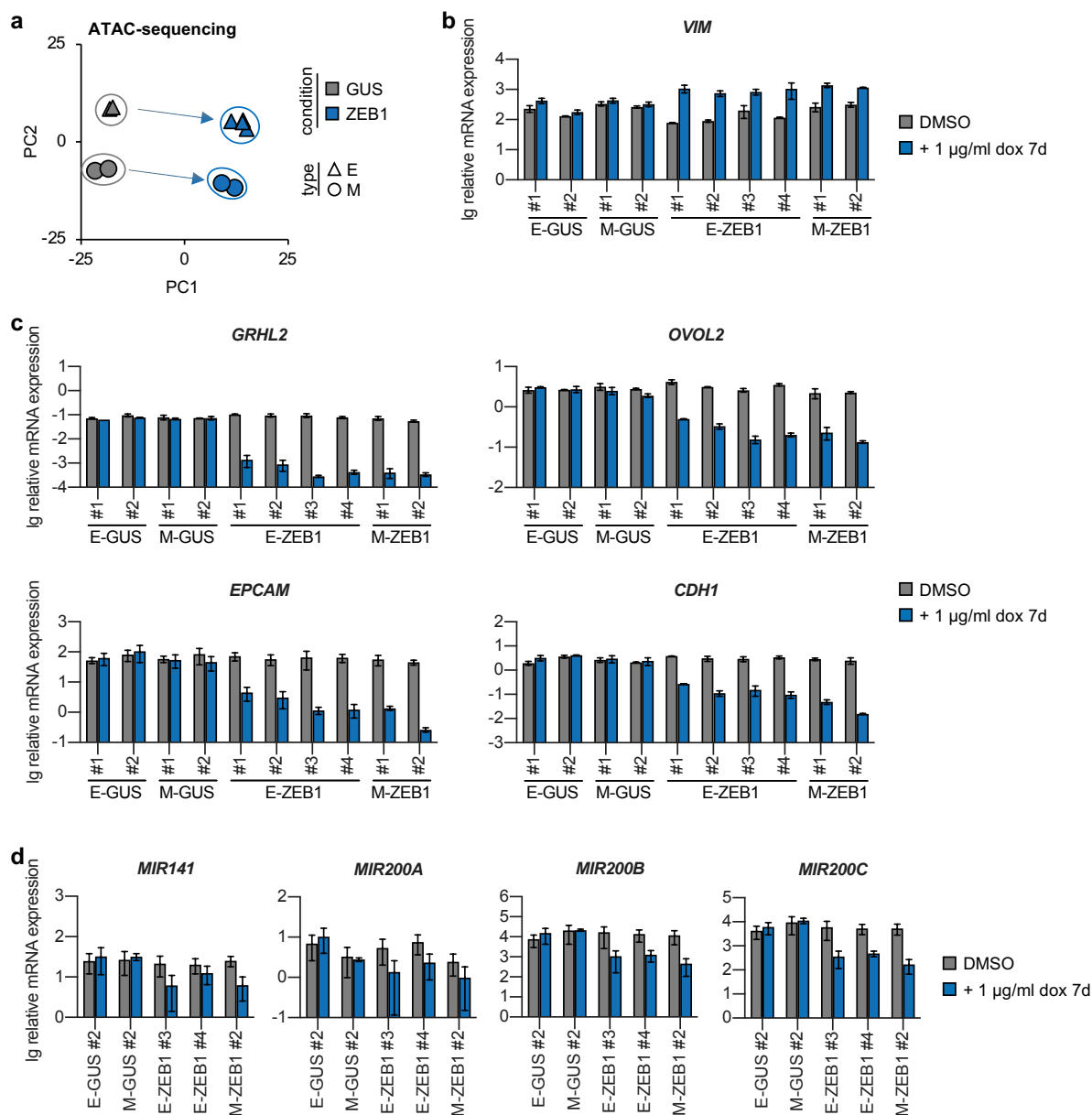
Figure 4-14: Overexpression of *ZEB1* repressed target genes in EMT-resistant and EMT-competent cells
a) Principal component (PC) analysis of ATAC-sequencing results of EMT-resistant (E, △) and EMT-competent (M, ○) SCCs harbouring an inducible *GUS* (E-GUS #1-2, M-GUS #1-2) or *ZEB1* (E-ZEB1 #1-4, M-ZEB1 #1-2) construct treated with 1 µg/ml doxycycline for 7 days. Each data point represents one SCC. **b)** Lg relative *VIM* mRNA expression levels of E-GUS #1-2, M-GUS #1-2, E-ZEB1 #1-4, and M-ZEB1 #1-2 SCCs treated with DMSO (DMSO) or with 1 µg/ml doxycycline for 7 days (+1 µg/ml dox 7d). n=2. Mean ± SEM. **c)** Lg relative *GRHL2*, *OVOL2*, *EPCAM*, and *CDH1* mRNA expression levels of E-GUS #1-2, M-GUS #1-2, E-ZEB1 #1-4, and M-ZEB1 #1-2 SCCs treated as described in b). n=2. Mean ± SEM. **d)** Lg relative *MIR141*, *MIR200A*, *MIR200B*, and *MIR200C* miRNA expression levels of E-GUS #2, M-GUS #2, E-ZEB1 #3 and #4, and M-ZEB1 #2 SCCs treated as described in b). n=2. Mean ± SEM.

Principal component analysis (PCA) of the ATAC-sequencing data revealed slight differences in chromatin accessibility between E-GUS and M-GUS SCCs on the PC2 axis; however, no differences were detected on the PC1 axis (Figure 4-14a). Upon *ZEB1* overexpression, a shift on the PC1 axis was detected for both E-ZEB1 and M-ZEB1 SCCs when compared to E-GUS and M-GUS SCCs, respectively (Figure 4-14a). On the PC2 axis, a minor shift occurred for both E-ZEB1 and M-ZEB1 SCCs, compared to E-GUS and M-GUS SCCs, respectively (Figure 4-14a). However, differences on the PC2 axis between E-ZEB1 and M-ZEB1 SCCs were comparable to differences between E-GUS and M-GUS SCCs, respectively (Figure 4-14a). Together, these findings showed that the induction of *ZEB1* overexpression induced similar changes in chromatin accessibility in E-ZEB1 and M-ZEB1 SCCs. That further indicated that both E-SCCs and M-SCCs responded similarly to *ZEB1* overexpression.

To investigate this observation further, I analysed the expression levels of the ZEB1 target genes *VIM* (mesenchymal) and *GRHL2*, *OVOL2*, *EPCAM*, *CDH1*, and *MIR200* (epithelial). Analysis of *VIM* transcript levels by qPCR revealed that, compared to cells treated with DMSO only, transcript levels of *VIM* were upregulated in both E-ZEB1 and M-ZEB1 SCCs when cells were treated with doxycycline for one week (+1 µg/ml dox 7d) to induce overexpression of *ZEB1* (Figure 4-14b). Furthermore, transcript levels of the ZEB1 target genes *GRHL2*, *OVOL2*, *EPCAM*, and *CDH1* decreased in E-ZEB1 and M-ZEB1 SCCs upon treatment with doxycycline and concomitant *ZEB1* overexpression (+1 µg/ml dox 7d, Figure 4-14c). In addition, expression levels of the miR-200 family members *MIR141*, *MIR200A*, *MIR200B*, and *MIR200C* were reduced in all E-ZEB1 and M-ZEB1 SCCs when *ZEB1* was induced (+1 µg/ml dox 7d, Figure 4-14d). Of note, the expression levels of these genes (*VIM*, *GRHL2*, *OVOL2*, *EPCAM*, *CDH1*, *MIR200*) were comparable in all E-GUS and M-GUS SCCs (DMSO, Figure 4-14b-d). Moreover, expression levels did not change when E-GUS and M-GUS SCCs were treated with doxycycline (+1 µg/ml dox 7d, Figure 4-14b-d).

These results indicated that the expression levels of mesenchymal and epithelial genes upon *ZEB1* overexpression changed similarly in E-ZEB1 and M-ZEB1 SCCs. In detail, expression of mesenchymal gene VIM increased, whereas epithelial gene expression decreased in all E-ZEB1 and M-ZEB1 SCCs when *ZEB1* overexpression was induced.

Taken together, these data showed that overexpression of the EMT-transcription factor *ZEB1* changed the overall chromatin accessibility in E-ZEB1 and M-ZEB1 SCCs similarly. The decrease in expression levels of epithelial genes and the increase of mesenchymal gene expression in both E-ZEB1 and M-ZEB1 SCCs indicated that the EMT-resistance of E-SCCs was eliminated when *ZEB1* was overexpressed. That, in turn, supported the hypothesis that *ZEB1* expression levels determined whether cells undergo or resist EMT. Overall, these findings hinted at the existence of a certain *ZEB1* expression threshold that needs to be exceeded for the successful trans-differentiation from an epithelial to a mesenchymal state.

4.15 EMT-competent cells remained mesenchymal, and EMT-resistant cells reverted to an epithelial state after transient overexpression of *ZEB1*

As I have observed that both E-SCCs and M-SCCs underwent EMT upon *ZEB1* overexpression, I wondered whether these cells remain in a mesenchymal cell state or revert to an epithelial state when the induction of *ZEB1* overexpression is withdrawn.

Therefore, I transferred E-ZEB1 and M-ZEB1 SCCs treated with doxycycline for seven days to doxycycline-free conditions for seven (+dox 7d/DMSO 7d) and 14 days (+dox 7d/DMSO 14d). As controls, E-GUS and M-GUS SCCs were treated identically.

Doing so revealed that M-ZEB1 SCCs maintained a mesenchymal morphology despite doxycycline withdrawal (+dox 7d/DMSO 7d and +dox 7d/14d, Figure 4-15a). Interestingly, E-ZEB1 SCCs reverted from a mesenchymal to an epithelial morphology upon withdrawal of doxycycline (+dox 7d/DMSO 7d or 14d, Figure 4-15a).

For a more detailed analysis, I set out to investigate the expression of the epithelial cell surface protein EPCAM by flow cytometry at four different time points: before treatment (DMSO) after one week of *ZEB1* overexpression (+1 µg/ml dox 7d) and after one week of *ZEB1* overexpression followed by one week (+dox 7d/DMSO 7d) or two weeks of withdrawal (+dox 7d/DMSO 14d, Figure 4-15b).

This analysis showed that before ZEB1 overexpression, E-ZEB1 and M-ZEB1 SCCs, as well as E-GUS and M-GUS SCCs, displayed EPCAM protein (EPCAM$^+$) on their surface (DMSO, Figure 4-15b). However, after one week of doxycycline treatment and concomitant activation of *ZEB1* overexpression (+1 µg/ml dox 7d), E-ZEB1 and M-ZEB1 SCCs lost the expression of surface EPCAM protein (EPCAM$^-$, Figure 4-15b). Of note, *EPCAM* transcript levels were also decreased in these cells upon *ZEB1* overexpression (+1 µg/ml dox 7d, Figure 4-14c).

Interestingly, when doxycycline was withdrawn for seven (+dox 7d/DMSO 7d) or 14 days (+dox 7d/DMSO 14d), M-ZEB1 SCCs remained EPCAM-negative (EPCAM$^-$), whereas E-ZEB1 SCCs started to re-establish the expression of surface EPCAM protein (EPCAM$^+$, Figure 4-15b). In detail, after seven days of withdrawal (+dox 7d/DMSO 7d), the majority of cells of E-ZEB1 #1 and #2 was already EPCAM-positive (EPCAM$^+$, 89% and 75%, respectively, Figure 4-15b). When the withdrawal was extended to 14 days (+dox 7d/DMSO 14d), the proportion of EPCAM$^+$ cells increased even further (93% and 84%, Figure 4-15b). Interestingly, also E-ZEB1 #3 and #4 reverted to a more epithelial phenotype (EPCAM$^+$, Figure 4-15b), though, these cells revealed a slower reversion process than E-ZEB1 #1 and #2. Briefly, after seven days of withdrawal (+dox 7d/DMSO 7d), a small number of cells of E-ZEB1 #3 and #4 displayed EPCAM protein on their surface (EPCAM$^+$, 38% and 42%, respectively, Figure 4-15b). Again, after an additional week of withdrawal (+dox 7d/DMSO 14d), the proportion of EPCAM$^+$ cells of E-ZEB1 #3 and #4 increased (45% and 50%, Figure 4-15b).
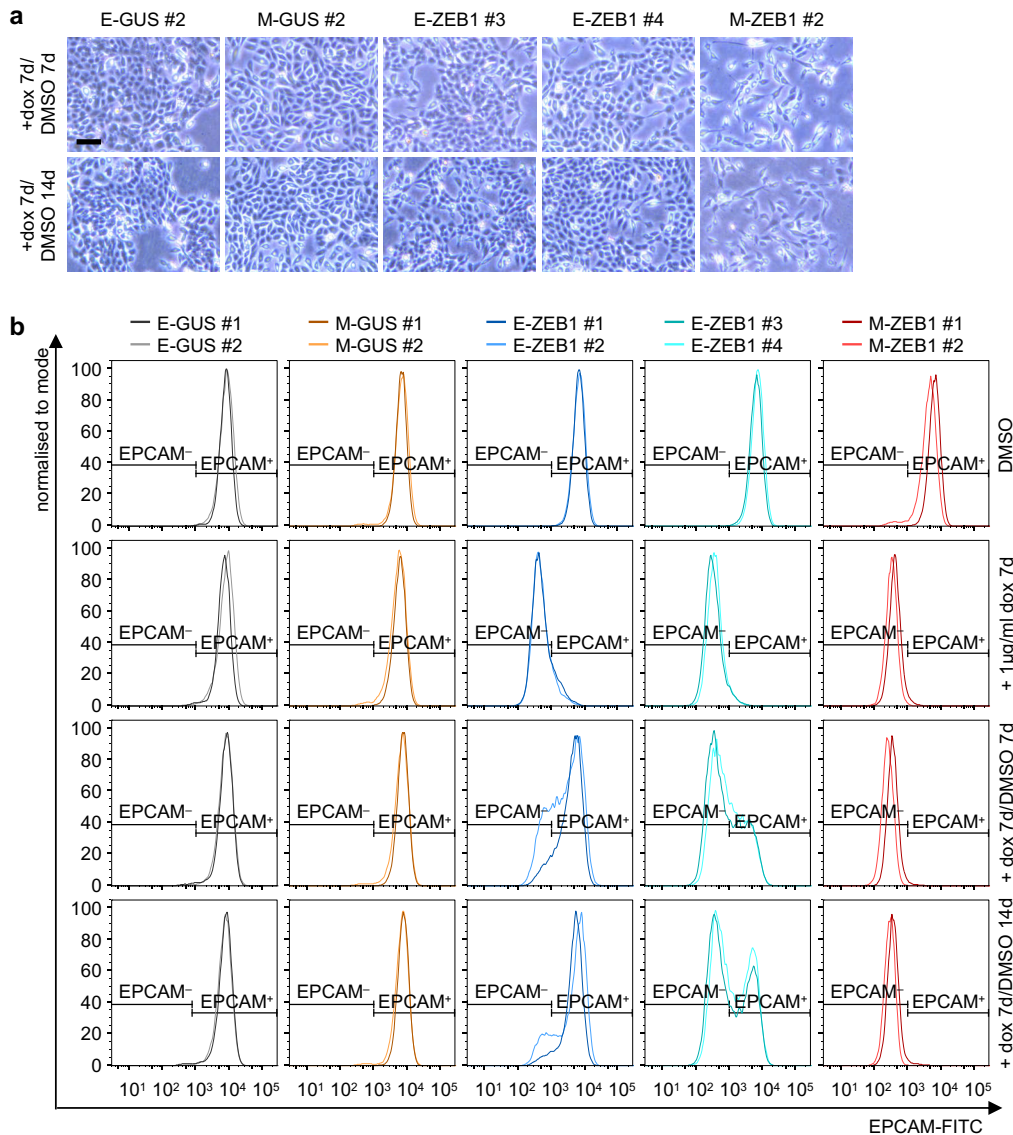
Figure 4-15: Re-expression of EPCAM after transient *ZEB1* induction appeared only in EMT-resistant cells
a) Bright-field images of E-GUS #2, M-GUS #2, E-ZEB1 #3 and #4, and M-ZEB1 #2 treated with 1 µg/ml doxycycline for 7 days followed by treatment with DMSO for 7 (+dox 7d/DMSO 14d) or 14 days (+dox 7d/DMSO 14d). Scale bar: 100 µm. **b)** Flow cytometric analysis of EPCAM of E-GUS #1-2, M-GUS #1-2, E-ZEB #1-4, and M-ZEB #1-2 treated with DMSO or with 1 µg/ml doxycycline for 7 days (+ 1 µg/ml dox 7d) or treated with 1 µg/ml doxycycline for 7 days followed by treatment with DMSO for 7 (+dox 7d/DMSO 14d) or 14 days (+dox 7d/DMSO 14d).

Together, these results revealed that upon withdrawal of *ZEB1*-induction, E-ZEB1 SCCs started re-expressing the epithelial cell-cell junction protein EPCAM and reverting to an epithelial morphology. Of note, cells within each E-ZEB1 SCC displayed heterogeneity, as not all cells were equally fast in reverting to an epithelial state. In contrast, M-ZEB1 SCCs remained EPCAM-negative and maintained a mesenchymal morphology despite the withdrawal of *ZEB1*-induction. These results showed that EMT induced by transient *ZEB1* overexpression was reversible in E-SCCs and irreversible in M-SCCs.

4.16 Decrease of *ZEB1* expression levels below a certain threshold allows reversion of EMT-resistant cells to an epithelial state

The previous analyses showed that after transient overexpression of *ZEB1*, M-SCCs remained mesenchymal, whereas E-SCCs started re-expressing epithelial markers and reverting to an epithelial state. Based on these findings, I aimed to investigate the differences between cells that reverted to an epithelial state and cells that remained mesenchymal. For this purpose, I made use of the fact that after transient overexpression of *ZEB1*, M-ZEB1 SCCs maintained a mesenchymal state whereas E-ZEB1 SCCs revealed heterogeneity in terms of how fast cells re-expressed EPCAM protein on the surface.

I used FACS to sort cells after seven days of transient *ZEB1* induction followed by 14 days of withdrawal (+dox 7d/DMSO 14d). In detail, EPCAM^{pos} and EPCAM^{neg} fractions of E-ZEB1 #3 and #4, as well as EPCAM^{neg} cells of M-ZEB1 #2, were separated (Figure 4-16a). To obtain fully epithelial cells that could be used as control, EPCAM^{pos}-sorted cells of E-GUS #2 and M-GUS #2 were used (+dox 7d/DMSO 14d, Figure 4-16a). As I wished to determine the expression of different epithelial and mesenchymal genes in the sorted EPCAM^{pos} and EPCAM^{neg} cells, freshly sorted cells were processed directly for qPCR analyses.

Analysis of *ZEB1* transcript levels in sorted cells revealed 6- to 8-fold higher *ZEB1* transcript levels in the EPCAM^{neg}-sorted cells of E-ZEB1 SCCs compared to the EPCAM^{pos}-sorted cells of E-ZEB1 SCCs (Figure 4-16b). Of note, *ZEB1* levels of EPCAM^{neg}-sorted cells of M-ZEB1 #2 were similar to those in the EPCAM^{neg}-sorted cells of E-ZEB1 SCCs (Figure 4-16b). Moreover, *ZEB1* levels in EPCAM^{neg}-sorted cells of E-ZEB1 SCCs and EPCAM^{neg}-sorted cells of M-ZEB1 #2 were comparable to the levels observed in bulk cells treated with doxycycline for seven days (Figure 4-16b, Figure 4-13b). Importantly, the qPCR analysis showed that *mCherry* levels in EPCAM^{neg}-sorted and EPCAM^{pos}-sorted cells of E-ZEB1 SCCs, as well as M-ZEB1 #2 (Figure 4-16b), were comparable to *mCherry* levels of E-ZEB1 and M-ZEB1 SCCs before induction of *ZEB1* overexpression (DMSO, Figure 4-13b).

These results showed that after transient induction of *ZEB1* overexpression, M-ZEB1 SCCs maintained high levels of ZEB1 despite the withdrawal of doxycycline. Interestingly, high *ZEB1* transcript levels were also observed in the EPCAM^{neg}-sorted cells of E-ZEB1 SCCs. Together, these observations suggested that the maintenance of high *ZEB1* expression levels was responsible for the mesenchymal state of these cells. In contrast, the EPCAM^{pos}-sorted cells of E-ZEB1 SCCs displayed *ZEB1* levels that were lower than those in EPCAM^{neg}-sorted cells, indicating that the downregulation of *ZEB1* expression was essential for the reversion to an epithelial state. Of note, the withdrawal of doxycycline treatment resulted in a downregulation of *mCherry* expression in all cells harbouring the *ZEB1* construct, indicating that the expression of the *ZEB1-mCherry* construct was no longer induced. Nevertheless, high *ZEB1* transcript levels were detected in the EPCAM^{neg}-sorted cells. Based on this observation, I hypothesised

that activation of the exogenous *ZEB1* construct in E-ZEB1 and M-ZEB1 SCCs triggered the expression of the endogenous *ZEB1*. Importantly, previous results already indicated that ZEB1 might maintain its expression by a self-enforcing feedback loop.

However, in the EPCAM^{pos}-sorted cells of E-ZEB1 SCCs, *ZEB1* transcript levels decreased upon doxycycline withdrawal, which indicated that the self-enforcing ZEB1 mechanism was disturbed in these cells.

Based on these findings, I analysed additional mesenchymal and epithelial markers in the sorted cells by qPCR. These analyses revealed that the expression levels of the mesenchymal gene *VIM* were 3-fold higher in the EPCAM^{neg}-sorted cells compared to EPCAM^{pos}-sorted sorted cells of E-ZEB1 SCCs (Figure 4-16b). Moreover, expression levels of EPCAM^{neg}-sorted cells of M-ZEB1 #2 were similar to levels detected in EPCAM^{neg}-sorted cells of E-ZEB1 SCCs (Figure 4-16b).

These results showed that after transient induction of *ZEB1* overexpression, M-ZEB1 SCCs maintained the expression of mesenchymal markers. In contrast, the expression levels of mesenchymal markers decreased E-ZEB1 SCCs when the induction of ZEB1 overexpression was ceased.

Interestingly, members of the miR-200 family, more precisely *MIR200B* and *MIR200C*, were also differentially expressed in the sorted EPCAM^{neg} and EPCAM^{pos} subpopulations. In detail, 4- to 7-fold higher expression levels were observed in EPCAM^{pos}-sorted cells compared to EPCAM^{neg}-sorted cells of E-ZEB1 SCCs (Figure 4-16c). Again, levels in EPCAM^{neg}-sorted cells of M-ZEB1 #2 were similar to those observed in EPCAM^{neg}-sorted cells of E-ZEB1 SCCs (Figure 4-16c). Intriguingly, no differences between EPCAM^{pos}-sorted cells and EPCAM^{neg}-sorted cells were observed for the expression levels of *MIR141* and *MIR200A* (Figure 4-16c). However, I hypothesised that this could potentially be traced back to a differential affinity between miRNAs and ZEB1, as one base within the seed sequence slightly differs between *MIR141*/*MIR200C* and *MIR200A*/*MIR200B* (Cano & Nieto, 2008).

Substantial differences in expression levels were also observed for the ZEB1 downstream target genes *OVOL2* and *GRHL2*. In EPCAM^{pos}-sorted cells of E-ZEB1 SCCs, expression levels of *OVOL2* were 17- to 25-fold higher compared to EPCAM^{neg}-sorted cells (Figure 4-16d). Remarkably, more substantial differences were detected for *GRHL2.* In EPCAM^{pos}-sorted cells of E-ZEB1 SCCs, expression levels of *GRHL2* were 434- to 456-fold higher compared to EPCAM^{neg}-sorted cells of E-ZEB1 SCCs (Figure 4-16d). Again, in M-ZEB1 #2 EPCAM^{neg}-sorted cells, expression levels of *OVOL2* and *GRHL2* were comparable to those in EPCAM^{neg}-sorted cells of E-ZEB1 SCCs (Figure 4-16d). Interestingly, expression levels of *OVOL2* and *GRHL2* in EPCAM^{neg}-sorted cells of E-ZEB1 SCCs, as well as M-ZEB1 #2 (+dox 7d/DMSO 14d, Figure 4-16d), were similar to the expression levels detected in M-ZEB1 and E-ZEB1 bulk cells after seven days of *ZEB1* overexpression (+1 µg/ml dox 7d, Figure 4-14c).
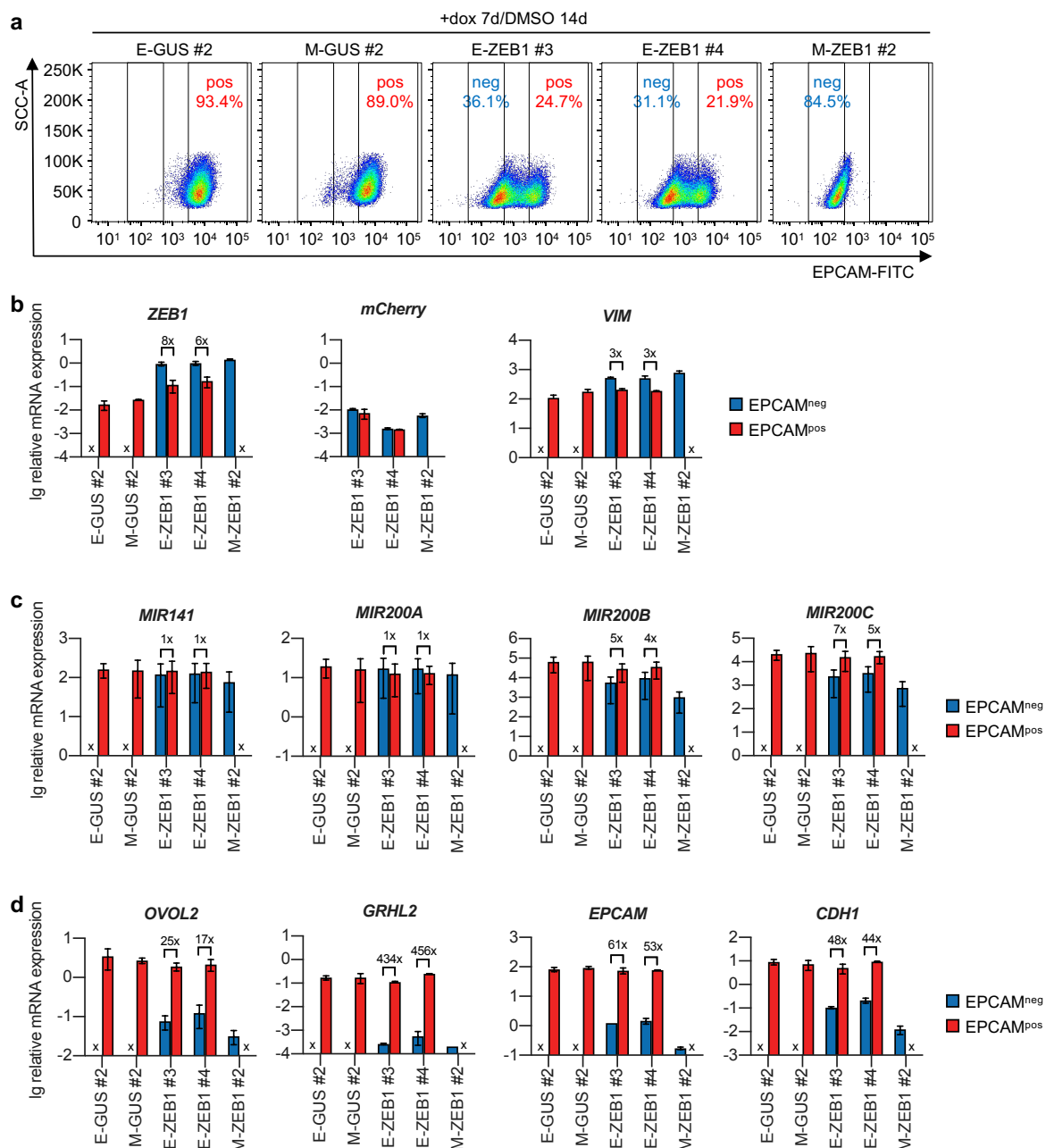
Figure 4-16: EMT-resistant cells re-expressed epithelial genes upon decrease of *ZEB1* expression levels
a) Flow cytometric analysis of EPCAM of E-GUS #2, M-GUS #2, E-ZEB1 #3 and #4, and M-ZEB1 #2 treated with 1 µg/ml doxycycline for 7 days followed by doxycycline withdrawal for 14 days (+dox 7d/DMSO 14d). Gates for fluorescent activated cell sorting are shown in black. neg: EPCAM[neg], pos: EPCAM[pos]. **b)** Lg relative *ZEB1*, *mCherry*, and *VIM* mRNA expression levels of EPCAM[neg] and EPCAM[pos] sorted cells of E-GUS #2 and M-GUS #2 and or E-ZEB1 #3 and #4 and M-ZEB1 #2 treated as described in a). n=2. Mean ± SEM. nx indicates differential expression between EPCAM[neg] and EPCAM[pos] cells. **c)** Lg relative *MIR141*, *MIR200A*, *MIR200B*, and *MIR200C* miRNA expression levels of EPCAM[neg] and EPCAM[pos] sorted cells of E-GUS #2, M-GUS #2, E-ZEB1 #3 and #4, and M-ZEB1 #2 treated as described in a). nx indicates differential expression between EPCAM[neg] and EPCAM[pos] cells. **d)** Lg relative *OVOL2*, *GRHL2*, *EPCAM*, and *CDH1* mRNA expression levels of EPCAM[neg] and EPCAM[pos] sorted cells of E-GUS #2, M-GUS #2, E-ZEB1 #3 and #4, and M-ZEB1 #2 treated as described in a). n=2. Mean ± SEM. nx indicates differential expression between EPCAM[neg] and EPCAM[pos] cells.

In contrast, levels of *OVOL2* and *GRHL2* transcript levels in EPCAM^{pos}-sorted cells of both E-ZEB1 SCCs (+dox 7d/DMSO 14d, Figure 4-16d) were comparable to transcript levels observed in E-ZEB1 and M-ZEB1 bulk cells before the induction of *ZEB1* overexpression (DMSO, Figure 4-14c). The same was observed for ZEB1 downstream target genes *CDH1* and *EPCAM*. In detail, expression levels of both epithelial genes were 44- to 61-fold higher in EPCAM^{pos}-sorted cells compared to EPCAM^{neg}-sorted cells of E-ZEB1 SCCs (Figure 4-16d). Also, *CDH1* and *EPCAM* transcript levels of EPCAM^{neg}-sorted M-ZEB1 #2 were similar to levels in EPCAM^{neg}-sorted cells of E-ZEB1 SCCs (Figure 4-16d).

These results showed that upon withdrawal of *ZEB1* induction, M-ZEB1 SCCs maintained low expression levels of epithelial genes, like *GRHL2*, *CDH1* and *EPCAM*, whereas E-ZEB1 SCCs started to re-establish the expression of epithelial genes.

Taken together, these results indicated that induction of *ZEB1* overexpression triggered the activation of a self-enforcing *ZEB1* expression loop E-ZEB1 and M-ZEB1 SCCs. Besides that, transient *ZEB1* overexpression in M-ZEB1 SCCs resulted in trans-differentiation to an irreversible, mesenchymal state. I hypothesised that the maintenance of high *ZEB1* expression levels in these cells resulted in the repression of ZEB1 downstream targets, like *EPCAM* and *CDH1*. In addition, expression levels of *ZEB1* repressors, like *GRHL2* and *OVOL2*, remained repressed, which contributed to the maintenance of the mesenchymal cell state. In contrast, after transient *ZEB1* overexpression in E-ZEB1 SCCs, *ZEB1* expression levels started to decrease, which allowed the re-expression of epithelial genes and the re-establishment of the feedback loops between ZEB1 and its mutual repressors like GRHL2 and OVOL2. That, in turn, lead to the reversion of E-ZEB1 SCCs to an epithelial cell state.

5 Summarising overview

Previous i*n vivo* studies have revealed that cancer cells expressing the cell-cell junction protein EPCAM are more efficient in tumorigenesis and metastasis (Baccelli et al., 2013; Hiraga et al., 2016). Furthermore, others have shown that activation of the EMT-TF Twist1 in breast cancer cells (HMLE-Twist1-ER) triggered EMT. However, upon withdrawal of Twist1-activation, only a subset of these cells has reverted to an epithelial cell state and displayed tumorigenic traits (Schmidt et al., 2015). Most distant metastases display an epithelial phenotype. Moreover, the outgrowth of these distant metastases causes the majority of cancer deaths. Therefore, the present study aimed to unravel what traits or factors determine that breast cancer cells that have undergone an EMT can regain epithelial and proliferative traits.

For that, I used HMLE-Twist1-ER cells, in which EMT can be induced in a targeted and, if required, a transient manner by activating the EMT-TF Twist1.

By doing so, I revealed that a subset of cells was EMT-competent and underwent trans-differentiation from an epithelial to a mesenchymal phenotype during Twist1-activation. Interestingly, this trans-differentiation was not reversible, but cells maintained a mesenchymal state, which in turn repressed the colonising capacity of the cells in a three-dimensional environment (EMT-competent, Figure 5-1). On the contrary, and unexpectedly, I revealed that the other subset of cells was resistant to Twist1-mediated EMT. These cells remained in an epithelial state and, in turn, were able to maintain their colonising ability in three-dimensional conditions (EMT-resistant, Figure 5-1).

Analysis of the chromatin accessibility in both subsets of cells showed that Twist1-activation generated differences in chromatin accessibility in EMT-competent and EMT-resistant cells.

Specifically, in EMT-competent cells, chromatin accessibility decreased at regions that contained binding motifs for GRHL2, a transcription factor (TF) that primes enhancers of epithelial genes for the expression. Interestingly the findings in the present study showed that before Twist1-activation, both subsets of cells expressed GRHL2. However, upon Twist1-activation, *GRHL2* expression decreased exclusively in EMT-competent cells. On the contrary, EMT-resistant cells could maintain the expression of *GRHL2* (Figure 5-1).

Interestingly, analysis of the chromatin landscape further revealed that Twist1-activation reduced the chromatin accessibility in EMT-competent cells at regions that contain a binding motif for ZEB1, an EMT-TF that is associated with repression of epithelial genes. Indeed, EMT-competent cells expressed higher levels of *ZEB1* transcript and protein than EMT-resistant cells. Furthermore, high levels of ZEB1 allowed the repression of epithelial genes associated with an epithelial identity, including *GRHL2*, *CDH1*, *EPCAM*, *OVOL2*, and *miR-200* family members (EMT-competent, Figure 5-1). On the contrary, ZEB1 levels in EMT-resistant cells were not sufficient to repress epithelial genes. Therefore, EMT-resistant cells remained epithelial despite Twist1-activation (EMT-resistant, Figure 5-1).
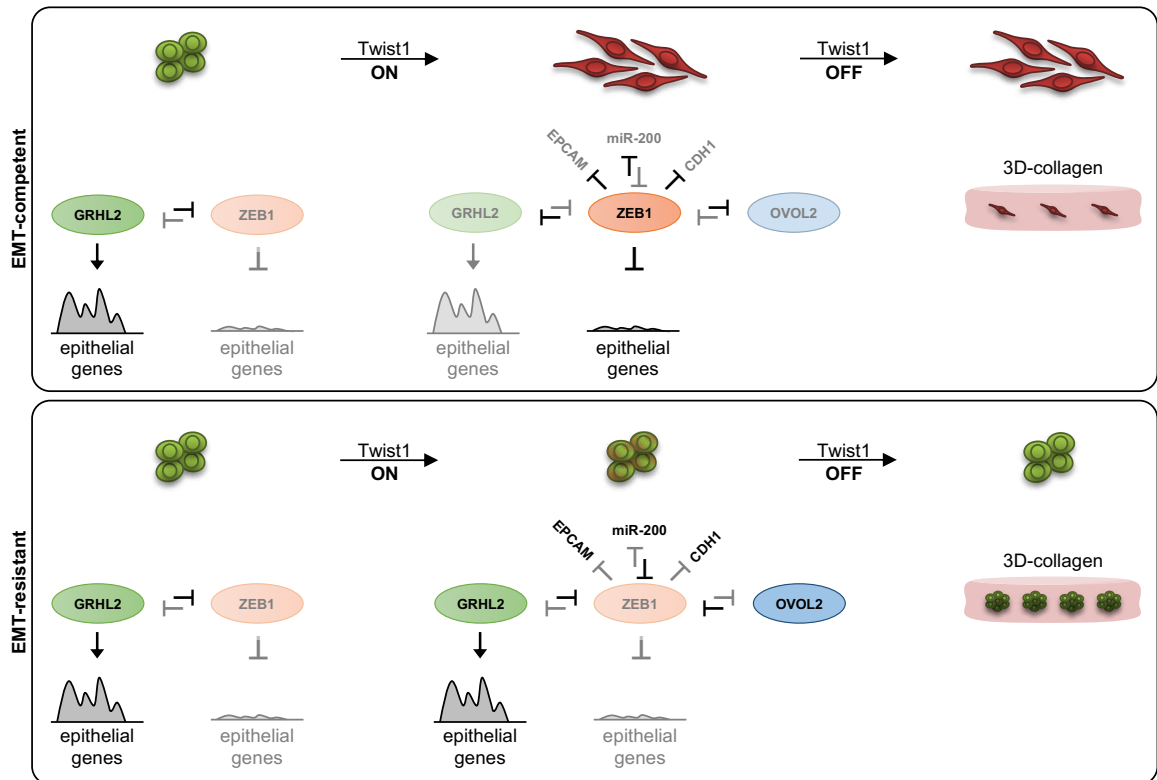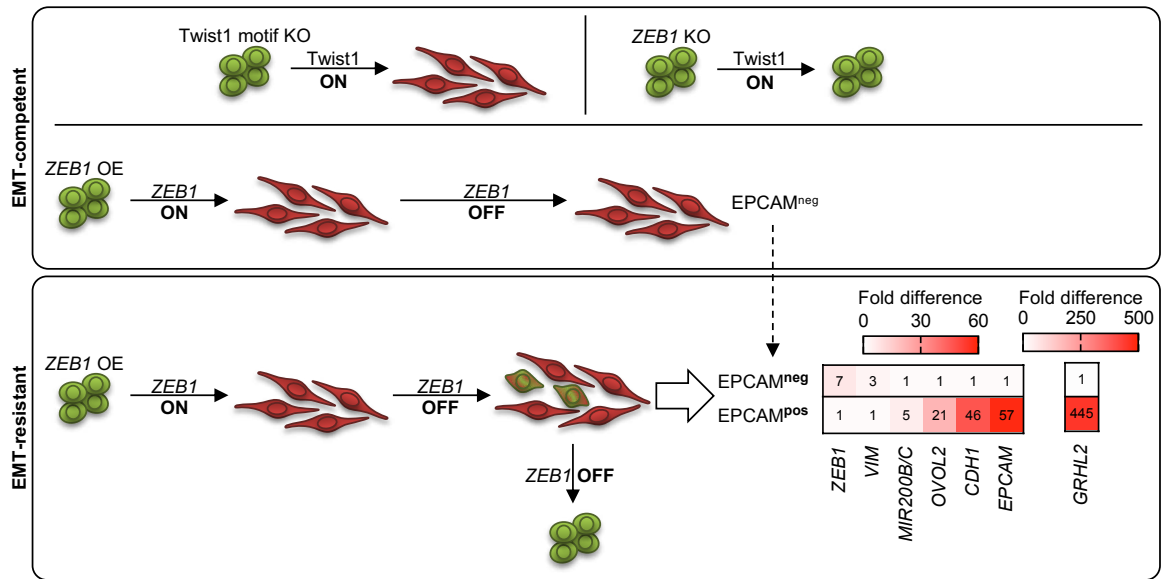
Figure 5-1: Feedback loops in EMT-competent and EMT-resistant cells

In normal conditions, when Twist1 activity is off, EMT-competent and EMT-resistant cells express GRHL2, which can prime epithelial genes for transcription. Upon Twist1-activation (Twist1 ON), ZEB1 expression is upregulated, and its expression levels in EMT-susceptible cells exceed a certain threshold resulting in the repression of epithelial genes like EPCAM, CDH1, miR-200 family members, OVOL2, and GRHL2. In EMT-resistant cells, levels of ZEB1 during Twist1-activation are too low to repress epithelial genes efficiently. The transient activation of Twist1 allows the outgrowth of EMT-resistant cells in 3D, whereas the colonising activity is inhibited in EMT-competent cells as they maintain a mesenchymal state.

The importance of ZEB1 for EMT was further underpinned by knockout experiments in EMT-competent cells. Loss of *ZEB1* in EMT-competent cells inhibited Twist1-mediated trans-differentiation of these cells (Figure 5-2). Remarkably, inducible *ZEB1* overexpression in EMT-resistant and EMT-compatible cells triggered EMT in both subsets of cells. However, after transient *ZEB1* overexpression, only EMT-resistant cells could revert to an epithelial state, as they could decrease *ZEB1* expression levels again. The results suggested that in these cells, *ZEB1* levels decreased below a certain threshold, leading to the re-expression of epithelial genes and the reversion from a mesenchymal to an epithelial state (Figure 5-2). Strikingly, the experiments in this study indicated that high ZEB1 levels could trigger the expression of *ZEB1* itself by a self-enforcing feedback loop.

EMT-competent

Twist1 motif KO — Twist1 **ON** →

ZEB1 KO — Twist1 **ON** →

ZEB1 OE — ZEB1 **ON** → — ZEB1 **OFF** → EPCAM^neg

EMT-resistant

ZEB1 OE — ZEB1 **ON** → — ZEB1 **OFF** → → EPCAM^neg / EPCAM^pos

ZEB1 **OFF** →

Fold difference						Fold difference	
0	30	60				0	250	500

	ZEB1	VIM	MIR200B/C	OVOL2	CDH1	EPCAM	GRHL2
EPCAM^neg	7	3	1	1	1	1	1
EPCAM^pos	1	1	5	21	46	57	445

Figure 5-2: *ZEB1* knockout and overexpression in EMT-competent and or EMT-resistant cells
Twist1 binding to its motif within the ZEB1 gene of EMT-competent cells is not essential for the upregulation of *ZEB1* expression upon Twist1-activation. However, expression of the EMT-TF ZEB1 is essential for EMT in EMT-competent cells, as deletion of *ZEB1* inhibits Twist1-mediated EMT. *ZEB1* overexpression induces EMT in EMT-susceptible cells as well as EMT-resistant cells. However, upon transient activation of *ZEB1* overexpression, only EMT-resistant cells can reduce *ZEB1* expression levels and thereby allow the upregulation of epithelial gene expression and the reversion back to an epithelial state.

In conclusion, this study revealed that a subset of breast cancer cells was competent to undergo Twist1-mediated EMT. The trans-differentiation from a mesenchymal to an epithelial state was irreversible and dependent on a certain expression level of the EMT-TF ZEB1. Remarkably, the other subset of breast cancer cells maintained epithelial traits despite the activation of the EMT-TF Twist1. Within the latter, Twist1 was not sufficient to drive the expression of *ZEB1* to a high level. However, the expression of high ZEB1 levels was an essential factor for the trans-differentiation to a mesenchymal state.

Interestingly, *ZEB1* overexpression triggered EMT in EMT-competent but also EMT-resistant cells. However, only EMT-resistant cells reverted to an epithelial state after transient *ZEB1* overexpression. In contrast, EMT-competent cells remained a mesenchymal cell state.

Most importantly, the results of this study suggested that repression of chromatin accessibility of epithelial genes and concomitant trans-differentiation from an epithelial to an irreversible, mesenchymal state in a subset of cells inhibited the outgrowth of cells in a three-dimensional environment. At the same time, some breast cancer cells preserved the epithelial program, which allowed the colonisation in a three-dimensional environment.

Breast cancer progression is a complicated mechanism, and several factors can influence tumour growth, invasiveness, and the formation of metastases. The findings of this study show

that heterogeneity with regard to EMT exists within breast cancer cells. Besides that, this study gives further insights into the mechanism of epithelial plasticity during breast cancer progression. In the future, the findings of this study could be the basis for further research and the development of therapeutics targeting the EMT signalling cascade to prevent the dissemination of tumour cells and target tumour cells at distant sites.

6 Discussion

The epithelial-mesenchymal transition (EMT) is an essential process during development. Cells with an epithelial phenotype lose cell-cell adhesions and trans-differentiate to cells with mesenchymal traits (Kalluri & Weinberg, 2009; Yang & Weinberg, 2008). Besides its implication during development, EMT is also observed in a process known as the metastatic cascade, where it is one of the first steps that promote the formation of metastases at distant sites (Scheel & Weinberg, 2012). Despite numerous studies on EMT and its role during metastasis formation, several opinions range from EMT as indispensable to EMT as an unessential factor during cancer progression.

Breast cancer belongs to the most common cancer types in women, and survival rates of patients with localised cancer are just below 100%. However, if patients are diagnosed with metastasised breast cancer, the chances of survival are meagre (Howlader et al., 2020). In general, around 90% of all cancer-related deaths result from metastases at distant organs (Chaffer & Weinberg, 2011; Fidler, 2002). In order to develop treatments that can increase patient survival, it is essential to understand what types of cancer cells can colonise at distant sites and what mechanisms promote the outgrowth of these metastases.

6.1 Heterogeneity with regard to EMT emerges already during Twist1-activation

Experiments with immortalised human mammary epithelial cells (HMLE) that express the Tamoxifen-inducible Twist1 protein (Twist1-ER) have shown that these cells undergo EMT when Twist1 is activated (Schmidt et al., 2015). However, when Twist1-activation was followed by withdrawal of Twist1-activation, only a subset of cells reverted to an epithelial state (Schmidt et al., 2015), suggesting that these cells underwent the reverse process of EMT, which is known as the mesenchymal-epithelial transition (MET).

In the present study, I used the same HMLE-Twist1-ER cells (CD24^{high}) and induced EMT in a targeted manner by activating Twist1 via treatment with Tamoxifen (TAM). To unravel the previously observed heterogeneity after transient Twist1-activation by Schmidt and colleagues (Schmidt et al., 2015), I set out to analyse the cells for their expression of EPCAM.

In the past, it has already been shown that CTCs obtained from luminal breast cancer patients that express EPCAM can metastasise *in vivo* (Baccelli et al., 2013). Additionally, high EPCAM levels of tumour cells isolated from pleural effusions of patients with different breast cancer subtypes correlate with higher tumorigenicity and metastatic capacity in mice (data obtained by M. Saini, pre-printed at Eichelberger et al., 2020).

Interestingly, within the present study, Twist1-activation revealed the existence of two subpopulations within the HMLE-Twist1-ER bulk: One population of cells that lost EPCAM expression, which indicated that these cells underwent EMT, and one subset of cells that

maintained EPCAM expression, which suggested that these cells resisted EMT despite Twist1-activation. Furthermore, using single-cell clones (SCCs) isolated from the HMLE-Twist1-ER bulk population confirmed the existence of cells that could resist Twist1-mediated EMT (E-SCCs) and remain epithelial, and cells that were EMT-competent and underwent trans-differentiation to a mesenchymal state (M-SCCs) when Twist1 was activated.

These results showed that the heterogeneity with regard to EMT already evolved during Twist1-activation and not, as previously shown by Schmidt and colleagues, after transient Twist1-activation. However, this raises the question of why this was not observed beforehand in several studies that used HMLE-Twist1-ER bulk cells (Casas et al., 2011; Dragoi et al., 2016; Schmidt et al., 2015). Of note, the proportion of cells that maintained EPCAM expression during Twist1-activation was very small, which might explain why analyses using the whole bulk population rather represented the properties of the EPCAM^{neg} cells than displaying heterogeneity. Of note, mixing experiments, where labelled EMT-competent and labelled EMT-resistant SCCs were cultured together, revealed that during Twist1-activation, EMT-resistant cells could not be distinguished from EMT-competent cells based on morphology (data not shown, obtained by J. Bartsch). This observation could explain why the heterogeneity in terms of EMT was not detected earlier. Therefore, it is important that the results obtained in this study will be considered in future studies that use HMLE-Twist1-ER bulk cells.

6.2 Maintenance of epithelial traits is essential for colonisation

As described previously, I observed that one subset of HMLE-Twist1-ER cells was EMT-competent, whereas the other subset of cells was resistant to Twist-mediated EMT. Using SCCs isolated from the untreated HMLE-Twist1-ER bulk population, I analysed the changes and differences that evolved in both subpopulations during Twist1-activation more precisely. In EMT-resistant SCCs (E-SCCs), the expression of epithelial genes like *CDH1* (E-cadherin) and *EPCAM* was maintained despite Twist1-activation. In contrast, in EMT-competent SCCs (M-SCCs), the expression of epithelial genes was downregulated upon Twist1-activation, resulting in loss of EPCAM and CDH1 protein abundance. Interestingly, upon Twist1-activation followed by withdrawal of Twist1-activation, M-SCCs did not undergo a mesenchymal-epithelial transition (MET) but remained in a mesenchymal state.

To better understand how this could impact tumour progression, cells were seeded into a 3D-collagen gel assay which has been developed to assess the regenerative capacity of healthy mammary epithelial cells (Linnemann et al., 2015) but has also been used to assess proliferation and invasive capacities of HMLE-Twist1-ER cells (Schmidt et al., 2015).

Doing so revealed that before Twist1-activation, both E-SCCs and M-SCCs could grow epithelial organoids in the three-dimensional environment. However, after transient Twist1-

activation, only EMT-resistant cells (E-SCCs) were able to form organoids expressing the epithelial cell-cell junction protein CDH1.

The loss of CDH1 expression is crucial during the metastatic cascade and cancer progression (Lamouille et al., 2014; Thiery et al., 2009). In the past, it was proposed that loss of the epithelial marker CDH1 is essential for metastasis. For example, Onder and colleagues have shown that loss of CDH1 in HMLE-Ras cells using an shRNA approach supports the formation of metastases *in vivo* (Onder et al., 2008). However, more recent studies have suggested that complete loss of epithelial traits hinders tumour growth. Kröger and colleagues, for example, have recently revealed that a spontaneous and irreversible EMT in a subset of HMLE-Ras cells reduces the tumorigenicity of these cells (Kröger et al., 2019). Interestingly, the same result was observed for the EMT-competent cells (M-SCCs) in the present study. Furthermore, Padmanaban and colleagues have shown that loss of CDH1 protein in models of invasive breast cancers leads to an increase in the invasion, but at the same time to a decrease in the outgrowth of metastases at distant sites, indicating that *CDH1* expression is important for metastasis (Padmanaban et al., 2019).

In the present study, I have observed that upon transient activation of EMT, only EMT-resistant cells could grow as epithelial structures as they were able to maintain the expression of cell-cell junction proteins, including CDH1. In conclusion, the results obtained in this study harmonise with the findings of Padmanaban and colleagues and further indicate that the expression of epithelial genes is important for epithelial colonisation.

Furthermore, metastases of breast cancers typically display epithelial traits (Kowalski et al., 2003). Therefore, the findings of this study could indicate that EMT-resistant E-SCCs represent tumour cells that actively from metastases at distant sites. On the contrary, EMT-competent M-SCCs had lost the colonising capacity after transient activation of Twist1. Therefore, I hypothesised that M-SCCs represent tumour cells with reduced tumorigenicity. In line with these findings, Fischer and colleagues have hypothesised that epithelial tumour cells better colonise at distant sites as these cells outcompete cells with a reduced proliferative capacity due to having undergone an EMT (Fischer et al., 2015).

6.3 A transient epithelial-mesenchymal hybrid sate – important for metastasis?

When analysing EMT-resistant cells (E-SCCs) and EMT-competent cells (M-SCCs) for their expression of mesenchymal markers, I observed that upon Twist1-activation, M-SCCs and E-SCCs upregulated the expression of mesenchymal genes like Vimentin (*VIM*) and Fibronectin (*FN1*). However, at the same time, E-SCCs maintained the expression of epithelial genes, which indicated that these cells obtained an epithelial-mesenchymal hybrid state. Recent analyses of EMT-transition states of *in vivo* models of mammary or cutaneous squamous

carcinoma have suggested that the acquisition of a hybrid epithelial-mesenchymal state is important for tumour progression (Pastushenko et al., 2018). Besides, another recent study has revealed that the epithelial-mesenchymal hybrid state is essential for the tumorigenicity of HMLE-Ras cells (Kröger et al., 2019).

However, in the present study, I observed that E-SCCs downregulated mesenchymal markers again when Twist1-activation was withdrawn, which indicated that these cells did not maintain the epithelial-mesenchymal hybrid state but instead reverted to an epithelial state. Moreover, these cells could colonise as epithelial organoids in a three-dimensional environment. Therefore, the results obtained in this study showed that a transient expression of epithelial and mesenchymal markers rather than a persistent epithelial-mesenchymal hybrid cell state is important for metastasis. Of note, this finding is further supported by other studies that have shown that a transient activation of EMT-transcription factors is essential for metastasis (Ocaña et al., 2012; Schmidt et al., 2015; Tsai et al., 2012). Nevertheless, to better understand whether a transient or a continuous epithelial-mesenchymal hybrid state is essential for metastasis, future studies could investigate whether EMT or rather EMT-activating stimuli are present or whether the activation of EMT is entirely ceased at the metastatic site.

6.4 Maintenance of epithelial gene expression in EMT-resistant cells correlates with tumour progression

In the present study, I observed that irreversible EMT repressed epithelial colonisation, whereas the transient acquisition of an epithelial-mesenchymal hybrid state allowed the formation of epithelial organoids in a three-dimensional environment. Therefore, I wished to understand what factors determine whether cells undergo EMT or resist trans-differentiation to a mesenchymal state and instead gain an epithelial-mesenchymal hybrid state.

For this purpose, ATAC-sequencing was performed in EMT-resistant cells (E-SCCs) and EMT-competent cells (M-SCCs) during different time points of transient Twist1-activation. Doing so revealed that changes in chromatin accessibility that occurred in E-SCCs upon Twist1-activation were reversed when Twist1-activation was withdrawn, which further supported the finding that E-SCCs transiently acquired an epithelial-mesenchymal hybrid state. In contrast to that, changes of chromatin accessibility that occurred in M-SCCs upon Twist1-activation were reversed only marginally when Twist1-activation was withdrawn in these cells, which further indicated that M-SCCs trans-differentiated to an irreversible mesenchymal state.

Analysis of chromatin regions that displayed a decrease in accessibility exclusively in M-SCCs but not E-SCCs upon Twist1-activation revealed the transcription factors ZEB1 and GRHL2, amongst others, as important factors of Twist1-mediated EMT. Further analyses indeed revealed that both *ZEB1* and *GRHL2* were differentially expressed in E-SCCs and M-SCCs

during Twist1-activation. In detail, after 14 days of Twist1-activation, *ZEB1* transcript and protein expression levels were significantly higher in M-SCCs than E-SCCs, and *GRHL2* expression levels were exclusively downregulated in M-SCCs. GRHL2 is known to prime epithelial enhancers for transcription and is regulated by ZEB1 in a mutual feedback loop (Cieply et al., 2012; Werner et al., 2013). Moreover, GRHL2 expression has already been linked to tumour growth and progression. For example, Xiang and colleagues have shown that high GRHL2 levels are associated with an increased risk of metastasis and poor relapse-free survival of breast cancer patients (Xiang et al., 2012). Additionally, the same study has revealed that overexpression of *GRHL2* promotes tumour growth and metastasis *in vivo* (Xiang et al., 2012). Moreover, the analyses of human pancreatic carcinoma samples by Wang and colleagues have revealed a correlation between strong GRHL2 expression and a worse prognosis for pancreatic carcinoma patients (Wang, Pan, Zhang, & Wang, 2019). Remarkably, the results of these studies correlate with the findings in the present study. In detail, in this study, the maintenance of high *GRHL2* expression levels despite transient activation of Twist1, as observed in E-SCCs, correlated with the capacity to colonise in a three-dimensional environment as epithelial organoids.

Concerning ZEB1, it has been shown that this transcription factor represses the expression of the epithelial genes *CDH1* (Sánchez-Tilló et al., 2010) and *EPCAM* (Vannier et al., 2013). In the present study, for both genes, *CDH1* and *EPCAM*, loss of chromatin accessibility and concomitant repression of gene expression appeared exclusively in M-SCCs, whereas in E-SCCs, chromatin accessibility and expression levels of these genes were maintained despite Twist1-activation. Moreover, cells that maintained the expression of epithelial proteins during transient Twist1-activation were able to colonise as epithelial organoids in a three-dimensional environment. In line with these observations, the correlation between CDH1 and EPCAM expression and tumorigenicity and metastasis has already been shown in other studies (Baccelli et al., 2013; Hiraga et al., 2016; Kowalski et al., 2003; Padmanaban et al., 2019).

Furthermore, other ZEB1 downstream factors like miR-200 family members and OVOL2, which ZEB1 represses in a mutual feedback loop (Bracken et al., 2008; Burk et al., 2008; Hong et al., 2015), were also repressed exclusively in M-SCCs during Twist1-activation. In line with that, a recent study that reviewed the results of several publications has revealed that high expression levels of miR-200 family members in various cancers correlate with worse overall survival of patients (Liu et al., 2018). Remarkably, in the present study, E-SCCs maintained the expression of all of these ZEB1 downstream factors associated with an epithelial state and, as shown in the studies mentioned above, a worse prognosis of cancer patients. Of note, only E-SCCs were able to colonise as epithelial organoids after transient Twist1-activation. Therefore, the findings by Liu, Wang, Xiang, and colleagues further support the hypothesis that was previously made in this study: HMLE-Twist1-ER cells that maintain high expression

levels of epithelial genes, like *EPCAM*, *CDH1*, *GRHL2* and *MIR200* despite transient Twist1-activation (E-SCCs) represent breast cancer cells that can propagate a tumour. Nevertheless, future studies could, for example, investigate whether loss of *GRHL2* expression in EMT-resistant cells (E-SCCs) could induce Twist1-mediated trans-differentiation to a mesenchymal state. If loss of *GRHL2* expression allows Twist1-mediated EMT of E-SCCs, this would indicate that GRHL2 mediates the maintenance of epithelial traits in E-SCCs. That, in turn, would further suggest that GRHL2 has an essential role in tumour progression.

6.5   The importance of *ZEB1* expression for EMT and metastasis

As previously mentioned, I observed that expression levels of *ZEB1* were significantly higher in M-SCCs than E-SCCs during Twist1-activation. Since ZEB1 target genes were repressed exclusively in M-SCCs, I hypothesised that a specific ZEB1 level was essential for the successful repression of epithelial genes. However, further investigations need to be performed in the future to unravel why only a subset of cells can express high ZEB1 levels and undergo EMT (M-SCCs), whereas ZEB1 levels in other cells remain low, and these cells resist trans-differentiation despite Twist1-activation.

Given this, one should keep in mind that Chaffer and colleagues have identified that different methylations at the *ZEB1* promoter of cells before an EMT-stimulus can determine whether cells are suitable to undergo EMT or not (Chaffer et al., 2013). In detail, a poised chromatin status, which is determined by the bivalent chromatin methylation patterns H3K4me3 and H3K27me3 at the *ZEB1* promoter region, enables breast cancer cells to undergo EMT. However, if the promoter region exclusively displays the repressive H3K27me3 methylation patterns, cells are prevented from undergoing EMT (Chaffer et al., 2013). Although I have observed that Twist1 binding to the promoter region is not essential for *ZEB1* expression and Twist1-mediated EMT in M-SCCs, further experiments should be performed to analyse the methylation state of the *ZEB1* gene in E-SCCs and M-SCCs before and during Twist1-activation. This analysis could unravel whether the methylation state determines differential *ZEB1* expression levels and the ability to undergo or resist Twist1-mediated EMT.

Nevertheless, by performing a CRISPR/Cas9-mediated loss of *ZEB1* in EMT-competent M-SCCs, I showed that *ZEB1* expression was essential for Twist1-mediated EMT in M-SCCs, as its loss inhibited Twist1-mediated EMT in these cells.

In the past, other studies have suggested that ZEB1 is important for the metastasis of cancer cells. For example, Spaderna and colleagues have shown that knockdown of *ZEB1* reduces the migration and invasion of breast and colorectal cancer cells as well as their metastasising capacity (Spaderna et al., 2008). Moreover, a study using a mouse model that develops pancreatic tumours driven by a Pdx1-cre-mediated activation of mutated Tp53 and Kras has

shown that *Zeb1* deletion results in a decrease of invasion of primary tumour cells and, in turn, reduces metastasis *in vivo*, suggesting that Zeb1 is essential for metastasis (Krebs et al., 2017). However, the present study shows that ZEB1 was essential for Twist1-mediated EMT. That means that loss of *ZEB1* prevents the gain of migratory traits because cells maintain an epithelial state. Hence, this could explain why migration and invasion were reduced in the studies by Spaderna, Krebs, and colleagues.

However, in contrast to the studies mentioned above, Fischer and colleagues have shown that inhibition of EMT and maintenance of an epithelial state by overexpressing *Mir200* in cells from a metastatic luminal breast cancer model does not affect lung metastasis (Fischer et al., 2015). That suggests that EMT and the expression of the *Mir200* target gene *Zeb1* are not essential for metastasis. In line with the finding of Fischer and colleagues, the present study also shows that for the outgrowth of cells, trans-differentiation to a mesenchymal state was not essential but instead inhibited colonisation in a three-dimensional environment. In consistence with this observation, it has been shown that epithelial prostate cancer cells have a higher metastatic potential than mesenchymal prostate cancer cells (Celià-Terrassa et al., 2012).

Nevertheless, this raises the question of how cancer cells that actively maintain epithelial characteristics can disseminate and reach distant sites. Interestingly, a study that used skin squamous cell carcinoma models has revealed that tumours need to display at least some degree of EMT for successful metastasis (Revenco et al., 2019). At the same time, the same study has shown that metastases in mice display epithelial features (Revenco et al., 2019), suggesting that cells undergo a mesenchymal-epithelial-transition (MET) at the site of metastasis. However, in the present study, cells that colonised after transient activation of EMT (E-SCCs) maintained epithelial traits and temporarily upregulated the expression of mesenchymal genes during transient Twist1-activation. Furthermore, these cells showed a slightly increased migratory behaviour when Twist1 was activated, although migratory traits were not as pronounced as in M-SCCs upon Twist1-activation.

Based on these observations, one could adopt the hypothesis by Celià-Terrassa and colleagues who have suggested that mesenchymal cancer cells facilitate the escape of epithelial cancer cells from the tumour so that the latter can form metastases (Celià-Terrassa et al., 2012). However, to further shed light on this hypothesis, future studies could investigate if breast cancer cells with differential response to an EMT-stimulus cooperate to escape from the primary tumour and whether that impacts the outgrowth of metastases at distant sites.

6.6   A self-enforcing *ZEB1* expression loop maintains the mesenchymal state of EMT-competent cells

The finding that M-SCCs expressed higher *ZEB1* levels than E-SCCs led to the hypothesis that overexpression of *ZEB1* might trigger EMT in E-SCCs. Indeed, in both E-SCCs and M-SCCs, *ZEB1* overexpression induced EMT, indicated by the loss of expression of epithelial genes like *EPCAM*, *CDH1*, *OVOL2*, and *GRHL2* and the acquisition of a mesenchymal morphology. Interestingly, the performed analyses also indicated that upon induction of *ZEB1* overexpression, the expression of the endogenous *ZEB1* was upregulated in both E-SCCs (E-ZEB1) and M-SCCs (M-ZEB1). In line with this observation, Preca and colleagues have shown that in breast and pancreatic cancer, ZEB1 can maintain its expression by repressing the Epithelial Splicing Regulatory Protein 1 (ESRP1) (Preca et al., 2015). That further leads to the synthesis of the CD44s isoform that is, in turn, able to activate the expression of *ZEB1* (Preca et al., 2015). Moreover, Preca and colleagues have discovered a positive feedback loop between ZEB1 and Hyaluronic Acid Synthase 2 (HAS2) (Preca et al., 2017).

Nevertheless, to unravel which factors could trigger the self-enforcing expression of *ZEB1* in E-SCCs and M-SCCs upon *ZEB1* overexpression, future studies need to be performed. Future analyses might also unravel how M-SCCs evolve and maintain high *ZEB1* expression levels upon transient Twist1-activation.

Upon withdrawal of *ZEB1* overexpression, I observed that E-ZEB1 cells re-expressed epithelial genes and trans-differentiated back to an epithelial state. In contrast, M-ZEB1 cells remained in a mesenchymal state despite the withdrawal of *ZEB1* overexpression. Further analyses revealed that in E-ZEB1 cells that have already reverted to an epithelial state, *ZEB1* transcript levels were downregulated again. Moreover, compared to E-ZEB1 cells that were still mesenchymal, *ZEB1* expression levels of already reverted epithelial E-ZEB1 cells were lower. Based on this observation, I hypothesised that as soon as *ZEB1* overexpression was withdrawn in E-ZEB1 cells, the endogenous *ZEB1* expression level decreased to a level, which was not sufficient to repress the expression of epithelial genes efficiently. Therefore, these cells reverted to an epithelial state. Interestingly, upon withdrawal of doxycycline in M-ZEB1 cells, the cells maintained high *ZEB1* expression levels. Remarkably, the maintenance of high *ZEB1* expression levels was also observed after transient Twist1-activation in M-SCCs. The maintenance of high *ZEB1* expression levels in M-SCCs despite the withdrawal of *ZEB1* overexpression or Twist1-activation further indicated that a self-enforcing *ZEB1* expression loop was activated in EMT-competent cells, which kept the cells in a mesenchymal state.

However, it remains an open question how M-SCCs maintain high ZEB1 levels and how E-ZEB1 can decrease high ZEB1 levels after transient *ZEB1* overexpression. To investigate that, future studies need to be performed. Thereby, one should consider that within the genome, ZEB1 protein can also bind to tandem repeats (TRs) that harbour several ZEB1 binding motifs.

Recently, Balestrieri and colleagues have shown that upon binding of ZEB1 at these TRs, ZEB1 can control the expression of genes, including miR-200 family members (Balestrieri et al., 2018). Interestingly, this regulation is controlled in a *cis*-regulatory manner, meaning that these TRs often localise several kilobases up- or downstream of the regulated genes (Balestrieri et al., 2018). Besides that, other stimuli that can induce EMT in cells should, of course, be taken into consideration when performing further investigations (see chapter 6.7).

6.7 Is autocrine TGFβ essential for high *ZEB1* expression levels and the stable mesenchymal cell state?

This study showed that upon Twist1-activation, E-SCCs maintained the expression of epithelial markers, but and at the same time, upregulated the expression of mesenchymal markers, which indicated that these cells obtained an epithelial-mesenchymal hybrid cell state. Several studies have suggested that a hybrid epithelial-mesenchymal phenotype is essential and the most competent for forming metastases. For example, Pastushenko and colleagues have shown that in mouse models of breast or squamous skin carcinoma, a hybrid epithelial-mesenchymal cell state has the highest metastatic capacity (Pastushenko et al., 2018). Furthermore, Kröger and colleagues have revealed that dwelling in an epithelial-mesenchymal state is essential for tumour initiation of HMLE-Ras cells (Kröger et al., 2019).

However, in the present study, E-SCCs could revert to an epithelial state once the EMT-stimulus Twist1 or ZEB1 was ceased. Furthermore, after transient Twist1-activation, E-SCCs colonised in a three-dimensional environment and formed epithelial organoids. Therefore, the present study indicates that it is not necessarily the maintenance of an epithelial-mesenchymal hybrid state but rather the maintenance of epithelial traits that is important for tumour progression. Of note, there was an apparent difference between Twist1-activation and *ZEB1* overexpression in E-SCCs. Upon transient Twist1-activation, E-SCCs resisted full trans-differentiation as *ZEB1* expression levels were not high enough to repress epithelial genes efficiently. On the contrary, during transient *ZEB1* overexpression, epithelial genes were repressed in E-SCCs, and cells transiently gained a mesenchymal state, as ZEB1 levels were high enough to repress epithelial genes. On the contrary, M-SCCs trans-differentiated and resided in a mesenchymal cell state, irrelevant of the EMT-stimulus, as these cells obtained and maintained high *ZEB1* expression levels in both setups.

Of course, these observations raise the question of whether other EMT-inducing stimuli could impact the upregulation and maintenance of *ZEB1* expression levels. In light of this, E-SCCs and M-SCCs were treated with TGFβ (Transforming Growth Factor beta) in a preliminary pilot experiment (data not shown in this study). This experiment revealed that M-SCCs could undergo EMT, whereas E-SCCs resisted trans-differentiation (data not shown in this study),

similar to what was observed during transient Twist1-activation. Besides that, a study by Dragoi and colleagues has shown that Twist1-mediated upregulation of *ZEB1* expression and concomitant EMT of HMLE-Twist1-ER bulk cells is dependent on the activation of the TGFβ receptor 1 (TGFBRI). However, the same study has observed that after full trans-differentiation to a mesenchymal state, cells no longer depend on TGFBRI activity when Twist1-activation is maintained (Dragoi et al., 2016). However, Dragoi and colleagues have not investigated whether TGFβ signalling is important for maintaining the mesenchymal state after transient Twist1-activation. Nevertheless, Gregory and colleagues have addressed this question by treating Madin-Darby Canine Kidney (MDCK) cells that had previously undergone a TGFβ-mediated EMT with a TGFBRI-inhibitor (Gregory et al., 2011). By doing so, they have revealed that autocrine TGFβ signalling is necessary to maintain high *Zeb1* and low *Mir200* expression levels and, in turn, the mesenchymal state of the MDCK cells that had previously undergone an EMT (Gregory et al., 2011).

Based on this knowledge, future studies could also investigate whether differences in autocrine TGFβ signalling determine whether cells obtain and, most importantly, maintain high *ZEB1* expression levels and, in turn, a mesenchymal cell state.

6.8   The importance of treatment options for both epithelial and mesenchymal cancer cells

In the past, breast cancer cells that underwent EMT and obtained a mesenchymal state were thought to be the most tumorigenic population (Al-Hajj, Wicha, Benito-Hernandez, Morrison, & Clarke, 2003; Mani et al., 2008). Therefore, studies have focused on identifying compounds that specifically target mesenchymal breast cancer cells (Gupta et al., 2009; Kai et al., 2015). However, the results of the present study point out that epithelial traits are important for colonisation, whereas residing in a mesenchymal state hinders epithelial outgrowth. This observation is further supported by the finding that epithelial breast cancer cells that express the epithelial cell-cell junction protein EPCAM have a higher capacity in initiating tumours and metastases *in vivo* (Baccelli et al., 2013; Sikandar et al., 2017). Moreover, it has been shown that the expression of EMT-associated genes in breast cancer cells does not necessarily correlate with enhanced tumorigenicity of these cells (Sikandar et al., 2017).

Of course, this raises the question of why mesenchymal tumour cells with low or absent expression of EPCAM can be found within the circulation of patients with advanced cancer. To answer this question, one could consider the hypothesis by Celià-Terrassa and colleagues that mesenchymal cancer cells might simply facilitate the escape and the dissemination of epithelial cancer cells (Celià-Terrassa et al., 2012).

Although the present study suggests that a mesenchymal state is permanent and could be favourable for patient survival due to a low colonisation activity, it should be kept in mind that the mesenchymal cells could persist at distant sites and, nevertheless, slowly proliferate as single mesenchymal cells. That should also be taken into account when new compounds that target metastasised cancer cells are developed. In particular, because Fischer and colleagues have shown that cells that had undergone an EMT are resistant to chemotherapies (Fischer et al., 2015). However, the fact that in the same study, epithelial cells were shown to be sensitive to treatment with chemotherapeutics (Fischer et al., 2015) is an important finding, considering that epithelial traits seem to be essential for the propagation of breast cancers.

7 Appendix

Gating strategy for HMLE-Twist1-ER bulk cells (+TAM 21d)
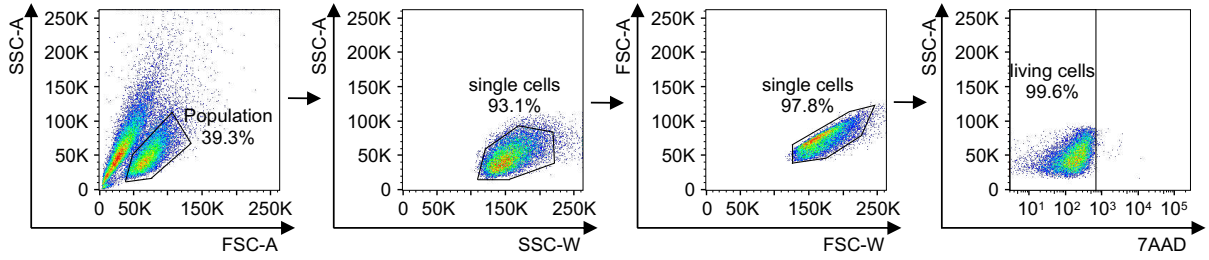
Gating strategy for *ZEB1* or *GUS* overexpression SCCs (+ 1µg/ml doxycycline 7d)
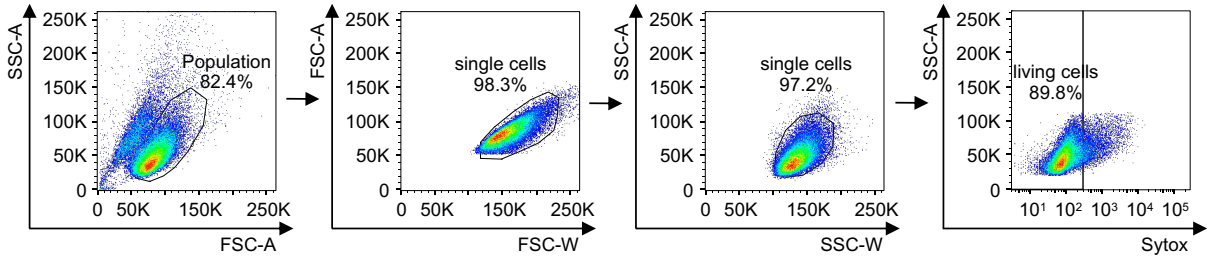
Figure 7-1: Gating strategies for HMLE-Twist1-ER cells and *ZEB1* or *GUS* overexpression SCCs

8 References

Aceto, N., Bardia, A., Miyamoto, D. T., Donaldson, M. C., Wittner, B. S., Spencer, J. A., … Maheswaran, S. (2014). Circulating tumor cell clusters are oligoclonal precursors of breast cancer metastasis. *Cell*, *158*(5), 1110–1122. https://doi.org/10.1016/j.cell.2014.07.013

Al-Hajj, M., Wicha, M. S., Benito-Hernandez, A., Morrison, S. J., & Clarke, M. F. (2003). Prospective identification of tumorigenic breast cancer cells. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(7), 3983–3988. https://doi.org/10.1073/pnas.0530291100

Alkabban, F., & Ferguson, T. (2020). *Breast Cancer. [Updated 2020 Nov 10]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2020 Jan-.* Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK482286/

Allard, W. J., Matera, J., Miller, M. C., Repollet, M., Connelly, M. C., Rao, C., … Terstappen, L. W. M. M. (2004). Tumor cells circulate in the peripheral blood of all major carcinomas but not in healthy subjects or patients with nonmalignant diseases. *Clinical Cancer Research : An Official Journal of the American Association for Cancer Research*, *10*(20), 6897–6904. https://doi.org/10.1158/1078-0432.CCR-04-0378

American Cancer Society. (2020). American Cancer Society. Retrieved from https://www.cancer.org/

Anders, S., Pyl, P. T., & Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics (Oxford, England)*, *31*(2), 166–169. https://doi.org/10.1093/bioinformatics/btu638

Baccelli, I., Schneeweiss, A., Riethdorf, S., Stenzinger, A., Schillert, A., Vogel, V., … Trumpp, A. (2013). Identification of a population of blood circulating tumor cells from breast cancer patients that initiates metastasis in a xenograft assay. *Nature Biotechnology*, *31*(6), 539–544. https://doi.org/10.1038/nbt.2576

Balestrieri, C., Alfarano, G., Milan, M., Tosi, V., Prosperini, E., Nicoli, P., … Natoli, G. (2018). Co-optation of Tandem DNA Repeats for the Maintenance of Mesenchymal Identity. *Cell*, *173*(5), 1150-1164.e14. https://doi.org/10.1016/j.cell.2018.03.081

Bane, A. (2013). Ductal Carcinoma In Situ : What the Pathologist Needs to Know and Why. *International Journal of Breast Cancer*, *2013*, 1–7. https://doi.org/10.1155/2013/914053

Beerling, E., Seinstra, D., de Wit, E., Kester, L., van der Velden, D., Maynard, C., … van Rheenen, J. (2016). Plasticity between Epithelial and Mesenchymal States Unlinks EMT from Metastasis-Enhancing Stem Cell Capacity. *Cell Reports*, *14*(10), 2281–2288. https://doi.org/10.1016/j.celrep.2016.02.034

Bergholtz, H., Lien, T. G., Swanson, D. M., Frigessi, A., Daidone, M. G., Tost, J., … Sørlie, T. (2020). Contrasting DCIS and invasive breast cancer by subtype suggests basal-like

DCIS as distinct lesions. *Npj Breast Cancer*, *6*(1), 26. https://doi.org/10.1038/s41523-020-0167-x

Bolós, V., Peinado, H., Pérez-Moreno, M. A., Fraga, M. F., Esteller, M., & Cano, A. (2003). The transcription factor Slug represses E-cadherin expression and induces epithelial to mesenchymal transitions: a comparison with Snail and E47 repressors. *Journal of Cell Science*, *116*(Pt 3), 499–511. https://doi.org/10.1242/jcs.00224

Bracken, C. P., Gregory, P. A., Kolesnikoff, N., Bert, A. G., Wang, J., Shannon, M. F., & Goodall, G. J. (2008). A double-negative feedback loop between ZEB1-SIP1 and the microRNA-200 family regulates epithelial-mesenchymal transition. *Cancer Research*, *68*(19), 7846–7854. https://doi.org/10.1158/0008-5472.CAN-08-1942

Breunig, C. T., Durovic, T., Neuner, A. M., Baumann, V., Wiesbeck, M. F., Köferle, A., … Stricker, S. H. (2018). One step generation of customizable gRNA vectors for multiplex CRISPR approaches through string assembly gRNA cloning (STAgR). *PloS One*, *13*(4), e0196015. https://doi.org/10.1371/journal.pone.0196015

Bronner, M. E. (2012). Formation and migration of neural crest cells in the vertebrate embryo. *Histochemistry and Cell Biology*, *138*(2), 179–186. https://doi.org/10.1007/s00418-012-0999-z

Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y., & Greenleaf, W. J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nature Methods*, *10*(12), 1213–1218. https://doi.org/10.1038/nmeth.2688

Burk, U., Schubert, J., Wellner, U., Schmalhofer, O., Vincan, E., Spaderna, S., & Brabletz, T. (2008). A reciprocal repression between ZEB1 and members of the miR-200 family promotes EMT and invasion in cancer cells. *EMBO Reports*, *9*(6), 582–589. https://doi.org/10.1038/embor.2008.74

Cancer Genome Atlas Network. (2012). Comprehensive molecular portraits of human breast tumours. *Nature*, *490*(7418), 61–70. https://doi.org/10.1038/nature11412

Cano, A., & Nieto, M. A. (2008). Non-coding RNAs take centre stage in epithelial-to-mesenchymal transition. *Trends in Cell Biology*, *18*(8), 357–359. https://doi.org/10.1016/j.tcb.2008.05.005

Casas, E., Kim, J., Bendesky, A., Ohno-Machado, L., Wolfe, C. J., & Yang, J. (2011). Snail2 is an essential mediator of Twist1-induced epithelial mesenchymal transition and metastasis. *Cancer Research*, *71*(1), 245–254. https://doi.org/10.1158/0008-5472.CAN-10-2330

Celià-Terrassa, T., Meca-Cortés, O., Mateo, F., Martínez de Paz, A., Rubio, N., Arnal-Estapé, A., … Thomson, T. M. (2012). Epithelial-mesenchymal transition can suppress major attributes of human epithelial tumor-initiating cells. *The Journal of Clinical Investigation*,

122(5), 1849–1868. https://doi.org/10.1172/JCI59218

Chaffer, C. L., Marjanovic, N. D., Lee, T., Bell, G., Kleer, C. G., Reinhardt, F., … Weinberg, R. A. (2013). Poised chromatin at the ZEB1 promoter enables breast cancer cell plasticity and enhances tumorigenicity. *Cell*, *154*(1), 61–74. https://doi.org/10.1016/j.cell.2013.06.005

Chaffer, C. L., & Weinberg, R. A. (2011). A perspective on cancer cell metastasis. *Science (New York, N.Y.)*, *331*(6024), 1559–1564. https://doi.org/10.1126/science.1203543

Chang, A. T., Liu, Y., Ayyanathan, K., Benner, C., Jiang, Y., Prokop, J. W., … Yang, J. (2015). An evolutionarily conserved DNA architecture determines target specificity of the TWIST family bHLH transcription factors. *Genes and Development*, *29*(6), 603–616. https://doi.org/10.1101/gad.242842.114

Cho, E. H., Wendel, M., Luttgen, M., Yoshioka, C., Marrinucci, D., Lazar, D., … Kuhn, P. (2012). Characterization of circulating tumor cell aggregates identified in patients with epithelial tumors. *Physical Biology*, *9*(1), 016001. https://doi.org/10.1088/1478-3975/9/1/016001

Cieply, B., Riley, P., Pifer, P. M., Widmeyer, J., Addison, J. B., Ivanov, A. V., … Frisch, S. M. (2012). Suppression of the epithelial-mesenchymal transition by Grainyhead-like-2. *Cancer Research*, *72*(9), 2440–2453. https://doi.org/10.1158/0008-5472.CAN-11-4038

Cohen, S. J., Punt, C. J. A., Iannotti, N., Saidman, B. H., Sabbath, K. D., Gabrail, N. Y., … Meropol, N. J. (2008). Relationship of circulating tumor cells to tumor response, progression-free survival, and overall survival in patients with metastatic colorectal cancer. *Journal of Clinical Oncology : Official Journal of the American Society of Clinical Oncology*, *26*(19), 3213–3221. https://doi.org/10.1200/JCO.2007.15.8923

Corces, M. R., Trevino, A. E., Hamilton, E. G., Greenside, P. G., Sinnott-Armstrong, N. A., Vesuna, S., … Chang, H. Y. (2017). An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nature Methods*, *14*(10), 959–962. https://doi.org/10.1038/nmeth.4396

Cristofanilli, M., Budd, G. T., Ellis, M. J., Stopeck, A., Matera, J., Miller, M. C., … Hayes, D. F. (2004). Circulating tumor cells, disease progression, and survival in metastatic breast cancer. *The New England Journal of Medicine*, *351*(8), 781–791. https://doi.org/10.1056/NEJMoa040766

Cristofanilli, M., Hayes, D. F., Budd, G. T., Ellis, M. J., Stopeck, A., Reuben, J. M., … Terstappen, L. W. M. M. (2005). Circulating tumor cells: a novel prognostic factor for newly diagnosed metastatic breast cancer. *Journal of Clinical Oncology : Official Journal of the American Society of Clinical Oncology*, *23*(7), 1420–1430. https://doi.org/10.1200/JCO.2005.08.140

de Bono, J. S., Scher, H. I., Montgomery, R. B., Parker, C., Miller, M. C., Tissing, H., …

Raghavan, D. (2008). Circulating tumor cells predict survival benefit from treatment in metastatic castration-resistant prostate cancer. *Clinical Cancer Research : An Official Journal of the American Association for Cancer Research*, *14*(19), 6302–6309. https://doi.org/10.1158/1078-0432.CCR-08-0872

De Craene, B., & Berx, G. (2013). Regulatory networks defining EMT during cancer initiation and progression. *Nature Reviews. Cancer*, *13*(2), 97–110. https://doi.org/10.1038/nrc3447

de Herreros, A. G., Peiró, S., Nassour, M., & Savagner, P. (2010). Snail family regulation and epithelial mesenchymal transitions in breast cancer progression. *Journal of Mammary Gland Biology and Neoplasia*, *15*(2), 135–147. https://doi.org/10.1007/s10911-010-9179-8

Deutsche Krebsgesellschaft. (2020). *Jahresbericht der zertifizierten Brustkrebszentren, Kennzahlenauswertung 2020, Auditjahr 2019 / Kennzahlenjahr 2018*. 36.

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., … Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics (Oxford, England)*, *29*(1), 15–21. https://doi.org/10.1093/bioinformatics/bts635

Dongre, A., & Weinberg, R. A. (2019). New insights into the mechanisms of epithelial-mesenchymal transition and implications for cancer. *Nature Reviews. Molecular Cell Biology*, *20*(2), 69–84. https://doi.org/10.1038/s41580-018-0080-4

Dragoi, D., Krattenmacher, A., Mishra, V. K., Schmidt, J. M., Kloos, U. J., Meixner, L. K., … Scheel, C. H. (2016). Twist1 induces distinct cell states depending on TGFBR1-activation. *Oncotarget*, *7*(21), 30396–30407. https://doi.org/10.18632/oncotarget.8878

Eger, A., Aigner, K., Sonderegger, S., Dampier, B., Oehler, S., Schreiber, M., … Foisner, R. (2005). DeltaEF1 is a transcriptional repressor of E-cadherin and regulates epithelial plasticity in breast cancer cells. *Oncogene*, *24*(14), 2375–2385. https://doi.org/10.1038/sj.onc.1208429

Eichelberger, L., Saini, M., Moreno, H. D., Klein, C., Bartsch, J. M., Falcone, M., … Scheel, C. H. (2020). Maintenance of epithelial traits and resistance to mesenchymal reprogramming promote proliferation in metastatic breast cancer. *BioRxiv*, 2020.03.19.998823. https://doi.org/10.1101/2020.03.19.998823

Elenbaas, B., Spirio, L., Koerner, F., Fleming, M. D., Zimonjic, D. B., Donaher, J. L., … Weinberg, R. a. (2001). Human breast cancer cells generated by oncogenic transformation of primary mammary epithelial cells. *Genes & Development*, *15*(1), 50–65. https://doi.org/10.1101/gad.828901

Fackenthal, J. D., & Olopade, O. I. (2007). Breast cancer risk associated with BRCA1 and BRCA2 in diverse populations. *Nature Reviews. Cancer*, *7*(12), 937–948. https://doi.org/10.1038/nrc2054

References

Fidler, I. J. (2002). Critical determinants of metastasis. *Seminars in Cancer Biology, 12*(2), 89–96. https://doi.org/10.1006/scbi.2001.0416

Fischer, K. R., Durrans, A., Lee, S., Sheng, J., Li, F., Wong, S. T. C., … Gao, D. (2015). Epithelial-to-mesenchymal transition is not required for lung metastasis but contributes to chemoresistance. *Nature, 527*(7579), 472–476. https://doi.org/10.1038/nature15748

Furusawa, T., Moribe, H., Kondoh, H., & Higashi, Y. (1999). Identification of CtBP1 and CtBP2 as corepressors of zinc finger-homeodomain factor deltaEF1. *Molecular and Cellular Biology, 19*(12), 8581–8590. https://doi.org/10.1128/mcb.19.12.8581

Geiger, T. R., & Peeper, D. S. (2009). Metastasis mechanisms. *Biochimica et Biophysica Acta, 1796*(2), 293–308. https://doi.org/10.1016/j.bbcan.2009.07.006

Gorges, T. M., Tinhofer, I., Drosch, M., Röse, L., Zollner, T. M., Krahn, T., & von Ahsen, O. (2012). Circulating tumour cells escape from EpCAM-based detection due to epithelial-to-mesenchymal transition. *BMC Cancer, 12*(1), 178. https://doi.org/10.1186/1471-2407-12-178

Gorodetska, I., Kozeretska, I., & Dubrovska, A. (2019). BRCA Genes: The Role in Genome Stability, Cancer Stemness and Therapy Resistance. *Journal of Cancer, 10*(9), 2109–2127. https://doi.org/10.7150/jca.30410

Gregory, P. A., Bert, A. G., Paterson, E. L., Barry, S. C., Tsykin, A., Farshid, G., … Goodall, G. J. (2008). The miR-200 family and miR-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1. *Nature Cell Biology, 10*(5), 593–601. https://doi.org/10.1038/ncb1722

Gregory, P. A., Bracken, C. P., Smith, E., Bert, A. G., Wright, J. A., Roslan, S., … Goodall, G. J. (2011). An autocrine TGF-beta/ZEB/miR-200 signaling network regulates establishment and maintenance of epithelial-mesenchymal transition. *Molecular Biology of the Cell, 22*(10), 1686–1698. https://doi.org/10.1091/mbc.E11-02-0103

Gupta, P. B., Onder, T. T., Jiang, G., Tao, K., Kuperwasser, C., Weinberg, R. A., & Lander, E. S. (2009). Identification of selective inhibitors of cancer stem cells by high-throughput screening. *Cell, 138*(4), 645–659. https://doi.org/10.1016/j.cell.2009.06.034

Gusterson, B. A., & Stein, T. (2012). Human breast development. *Seminars in Cell & Developmental Biology, 23*(5), 567–573. https://doi.org/10.1016/j.semcdb.2012.03.013

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y. C., Laslo, P., … Glass, C. K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular Cell, 38*(4), 576–589. https://doi.org/10.1016/j.molcel.2010.05.004

Herschkowitz, J. I., Simin, K., Weigman, V. J., Mikaelian, I., Usary, J., Hu, Z., … Perou, C. M. (2007). Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. *Genome Biology, 8*(5), R76.

https://doi.org/10.1186/gb-2007-8-5-r76

Hiraga, T., Ito, S., & Nakamura, H. (2016). EpCAM expression in breast cancer cells is associated with enhanced bone metastasis formation. *International Journal of Cancer*, *138*(7), 1698–1708. https://doi.org/10.1002/ijc.29921

Hong, T., Watanabe, K., Ta, C. H., Villarreal-Ponce, A., Nie, Q., & Dai, X. (2015). An Ovol2-Zeb1 Mutual Inhibitory Circuit Governs Bidirectional and Multi-step Transition between Epithelial and Mesenchymal States. *PLoS Computational Biology*, *11*(11), e1004569. https://doi.org/10.1371/journal.pcbi.1004569

Hortobagyi, G. N., Connolly, J. L., D'Orsi, C. J., Edge, S. B., Mittendorf, E. A., Rugo, H. S., … Giuliano, A. (2017). Breast. In *AJCC Cancer Staging Manual* (Vol. 57, pp. 589–636). https://doi.org/10.1007/978-3-319-40618-3_48

Howlader, N., AM, N., M, K., D, M., A, B., M, Y., … KA, C. (eds). (2020). SEER Cancer Statistics Review, 1975-2017, National Cancer Institute. Bethesda, MD, https://seer.cancer.gov/csr/1975_2017/, based on November 2019 SEER data submission, posted to the SEER web site, April 2020.

Huttlin, E. L., Bruckner, R. J., Paulo, J. A., Cannon, J. R., Ting, L., Baltier, K., … Harper, J. W. (2017). Architecture of the human interactome defines protein communities and disease networks. *Nature*, *545*(7655), 505–509. https://doi.org/10.1038/nature22366

Ince, T. A., Richardson, A. L., Bell, G. W., Saitoh, M., Godar, S., Karnoub, A. E., … Weinberg, R. A. (2007). Transformation of different human breast epithelial cell types leads to distinct tumor phenotypes. *Cancer Cell*, *12*(2), 160–170. https://doi.org/10.1016/j.ccr.2007.06.013

Jacobs, J., Atkins, M., Davie, K., Imrichova, H., Romanelli, L., Christiaens, V., … Aerts, S. (2018). The transcription factor Grainy head primes epithelial enhancers for spatiotemporal activation by displacing nucleosomes. *Nature Genetics*, *50*(7), 1011–1020. https://doi.org/10.1038/s41588-018-0140-x

Kai, M., Kanaya, N., Wu, S. V., Mendez, C., Nguyen, D., Luu, T., & Chen, S. (2015). Targeting breast cancer stem cells in triple-negative breast cancer using a combination of LBH589 and salinomycin. *Breast Cancer Research and Treatment*, *151*(2), 281–294. https://doi.org/10.1007/s10549-015-3376-5

Kalluri, R., & Weinberg, R. a. (2009). The basics of epithelial-mesenchymal transition. *Journal of Clinical Investigation*, *119*(6), 1420–1428. https://doi.org/10.1172/JCI39104

Kim, S.-I., Oceguera-Yanez, F., Sakurai, C., Nakagawa, M., Yamanaka, S., & Woltjen, K. (2016). Inducible Transgene Expression in Human iPS Cells Using Versatile All-in-One piggyBac Transposons. *Methods in Molecular Biology (Clifton, N.J.)*, *1357*, 111–131. https://doi.org/10.1007/7651_2015_251

Knouf, E. C., Garg, K., Arroyo, J. D., Correa, Y., Sarkar, D., Parkin, R. K., … Tewari, M. (2012).

An integrative genomic approach identifies p73 and p63 as activators of miR-200 microRNA family transcription. *Nucleic Acids Research*, *40*(2), 499–510. https://doi.org/10.1093/nar/gkr731

Kotlyar, M., Pastrello, C., Pivetta, F., Lo Sardo, A., Cumbaa, C., Li, H., … Jurisica, I. (2015). In silico prediction of physical protein interactions and characterization of interactome orphans. *Nature Methods*, *12*(1), 79–84. https://doi.org/10.1038/nmeth.3178

Kowalski, P. J., Rubin, M. a, & Kleer, C. G. (2003). E-cadherin expression in primary carcinomas of the breast and its distant metastases. *Breast Cancer Research : BCR*, *5*(6), R217-22. https://doi.org/10.1186/bcr651

Krebs, A. M., Mitschke, J., Losada, M. L., Schmalhofer, O., Boerries, M., Busch, H., … Brabletz, T. (2017). The EMT-activator Zeb1 is a key factor for cell plasticity and promotes metastasis in pancreatic cancer. *Nature Cell Biology*, *19*(5), 518–529. https://doi.org/10.1038/ncb3513

Kröger, C., Afeyan, A., Mraz, J., Eaton, E. N., Reinhardt, F., Khodor, Y. L., … Weinberg, R. A. (2019). Acquisition of a hybrid E/M state is essential for tumorigenicity of basal breast cancer cells. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(15), 7353–7362. https://doi.org/10.1073/pnas.1812876116

Lamouille, S., Xu, J., & Derynck, R. (2014). Molecular mechanisms of epithelial-mesenchymal transition. *Nature Reviews. Molecular Cell Biology*, *15*(3), 178–196. https://doi.org/10.1038/nrm3758

Langmead, B. (2010). Aligning Short Sequencing Reads with Bowtie. *Current Protocols in Bioinformatics*, *32*(1), 11.7.1-11.7.14. https://doi.org/doi:10.1002/0471250953.bi1107s32

Leitlinienprogramm Onkologie. (2020). *Interdisziplinäre S3 Leitlinie für die Früherkennung, Diagnostik, Therapie und Nachsorge des Mammakarzinoms, Langversion 4.3*. (Februar).

Linnemann, J. R., Miura, H., Meixner, L. K., Irmler, M., Kloos, U. J., Hirschi, B., … Scheel, C. H. (2015). Quantification of regenerative potential in primary human mammary epithelial cells. *Development*, *142*(18), 3239–3251. https://doi.org/10.1242/dev.123554

Liu, W., Zhang, K., Wei, P., Hu, Y., Peng, Y., Fang, X., … Wang, J. (2018). Correlation between miR-200 Family Overexpression and Cancer Prognosis. *Disease Markers*, *2018*, 6071826. https://doi.org/10.1155/2018/6071826

Macias, H., & Hinck, L. (2012). Mammary gland development. *Wiley Interdisciplinary Reviews. Developmental Biology*, *1*(4), 533–557. https://doi.org/10.1002/wdev.35

Mahdavi, M., Nassiri, M., Kooshyar, M. M., Vakili-Azghandi, M., Avan, A., Sandry, R., … Gopalan, V. (2019). Hereditary breast cancer; Genetic penetrance and current status with BRCA. *Journal of Cellular Physiology*, *234*(5), 5741–5750. https://doi.org/10.1002/jcp.27464

Mani, S. A., Guo, W., Liao, M. J., Eaton, E. N., Ayyanan, A., Zhou, A. Y., … Weinberg, R. A.

(2008). The Epithelial-Mesenchymal Transition Generates Cells with Properties of Stem Cells. *Cell*, *133*(4), 704–715. https://doi.org/10.1016/j.cell.2008.03.027

Martin, T. A., Goyal, A., Watkins, G., & Jiang, W. G. (2005). Expression of the transcription factors snail, slug, and twist and their clinical significance in human breast cancer. *Annals of Surgical Oncology*, *12*(6), 488–496. https://doi.org/10.1245/ASO.2005.04.010

Massari, M. E., & Murre, C. (2000). Helix-loop-helix proteins: regulators of transcription in eucaryotic organisms. *Molecular and Cellular Biology*, *20*(2), 429–440. https://doi.org/10.1128/mcb.20.2.429-440.2000

Murre, C., Bain, G., van Dijk, M. A., Engel, I., Furnari, B. A., Massari, M. E., … Stuiver, M. H. (1994). Structure and function of helix-loop-helix proteins. *Biochimica et Biophysica Acta*, *1218*(2), 129–135. https://doi.org/10.1016/0167-4781(94)90001-9

Nahta, R., & Esteva, F. J. (2006). Herceptin: mechanisms of action and resistance. *Cancer Letters*, *232*(2), 123–138. https://doi.org/10.1016/j.canlet.2005.01.041

Nieto, M. A., Sargent, M. G., Wilkinson, D. G., & Cooke, J. (1994). Control of cell behavior during vertebrate development by Slug, a zinc finger gene. *Science (New York, N.Y.)*, *264*(5160), 835–839. https://doi.org/10.1126/science.7513443

Ocaña, O. H., Córcoles, R., Fabra, A., Moreno-Bueno, G., Acloque, H., Vega, S., … Nieto, M. A. (2012). Metastatic colonization requires the repression of the epithelial-mesenchymal transition inducer Prrx1. *Cancer Cell*, *22*(6), 709–724. https://doi.org/10.1016/j.ccr.2012.10.012

Onder, T. T., Gupta, P. B., Mani, S. A., Yang, J., Lander, E. S., & Weinberg, R. A. (2008). Loss of E-Cadherin Promotes Metastasis via Multiple Downstream Transcriptional Pathways. *Cancer Research*, *68*(10), 3645–3654. https://doi.org/10.1158/0008-5472.CAN-07-2938

Padmanaban, V., Krol, I., Suhail, Y., Szczerba, B. M., Aceto, N., Bader, J. S., & Ewald, A. J. (2019). E-cadherin is required for metastasis in multiple models of breast cancer. *Nature*, *573*(7774), 439–444. https://doi.org/10.1038/s41586-019-1526-3

Paget, S. (1889). THE DISTRIBUTION OF SECONDARY GROWTHS IN CANCER OF THE BREAST. *The Lancet*, *133*(3421), 571–573. https://doi.org/10.1016/S0140-6736(00)49915-0

Park, S.-M., Gaur, A. B., Lengyel, E., & Peter, M. E. (2008). The miR-200 family determines the epithelial phenotype of cancer cells by targeting the E-cadherin repressors ZEB1 and ZEB2. *Genes & Development*, *22*(7), 894–907. https://doi.org/10.1101/gad.1640608

Parmar, H., & Cunha, G. R. (2004). Epithelial-stromal interactions in the mouse and human mammary gland in vivo. *Endocrine-Related Cancer*, *11*(3), 437–458. https://doi.org/10.1677/erc.1.00659

Pastushenko, I., D'Haene, N., Fioramonti, M., Blanpain, C., Minguijón, E., Van Keymeulen, A., … Dubois, C. (2018). Identification of the tumour transition states occurring during EMT.

Nature, *556*(7702), 463–468. https://doi.org/10.1038/s41586-018-0040-3

Peinado, H., Ballestar, E., Esteller, M., & Cano, A. (2004). Snail mediates E-cadherin repression by the recruitment of the Sin3A/histone deacetylase 1 (HDAC1)/HDAC2 complex. *Molecular and Cellular Biology*, *24*(1), 306–319. https://doi.org/10.1128/mcb.24.1.306-319.2004

Perou, C. M., Sørlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. a, … Botstein, D. (2000). Molecular portraits of human breast tumours. *Nature*, *406*(6797), 747–752. https://doi.org/10.1038/35021093

Prat, A., Parker, J. S., Karginova, O., Fan, C., Livasy, C., Herschkowitz, J. I., … Perou, C. M. (2010). Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. *Breast Cancer Research : BCR*, *12*(5), R68. https://doi.org/10.1186/bcr2635

Preca, B.-T., Bajdak, K., Mock, K., Lehmann, W., Sundararajan, V., Bronsert, P., … Stemmler, M. P. (2017). A novel ZEB1/HAS2 positive feedback loop promotes EMT in breast cancer. *Oncotarget*, *8*(7), 11530–11543. https://doi.org/10.18632/oncotarget.14563

Preca, B.-T., Bajdak, K., Mock, K., Sundararajan, V., Pfannstiel, J., Maurer, J., … Stemmler, M. P. (2015). A self-enforcing CD44s/ZEB1 feedback loop maintains EMT and stemness properties in cancer cells. *International Journal of Cancer*, *137*(11), 2566–2577. https://doi.org/10.1002/ijc.29642

Revenco, T., Nicodème, A., Pastushenko, I., Sznurkowska, M. K., Latil, M., Sotiropoulou, P. A., … Blanpain, C. (2019). Context Dependency of Epithelial-to-Mesenchymal Transition for Metastasis. *Cell Reports*, *29*(6), 1458-1468.e3. https://doi.org/10.1016/j.celrep.2019.09.081

Riethdorf, S., O'Flaherty, L., Hille, C., & Pantel, K. (2018). Clinical applications of the CellSearch platform in cancer patients. *Advanced Drug Delivery Reviews*, *125*, 102–121. https://doi.org/10.1016/j.addr.2018.01.011

Robert-Koch-Institut. (2019). Krebs in Deutschland für 2015 / 2016 Krebs in Deutschland. *Robert Koch Institut*, 160. Retrieved from https://www.krebsdaten.de/Krebs/DE/Content/Publikationen/Krebs_in_Deutschland/kid_2019/krebs_in_deutschland_2019.pdf?__blob=publicationFile

Roy, R., Chun, J., & Powell, S. N. (2011). BRCA1 and BRCA2: different roles in a common pathway of genome protection. *Nature Reviews. Cancer*, *12*(1), 68–78. https://doi.org/10.1038/nrc3181

Russo, J., & Russo, I. H. (2004). Development of the human breast. *Maturitas*, *49*(1), 2–15. https://doi.org/10.1016/j.maturitas.2004.04.011

Sánchez-Tilló, E., Lázaro, A., Torrent, R., Cuatrecasas, M., Vaquero, E. C., Castells, A., … Postigo, A. (2010). ZEB1 represses E-cadherin and induces an EMT by recruiting the

SWI/SNF chromatin-remodeling protein BRG1. *Oncogene*, *29*(24), 3490–3500. https://doi.org/10.1038/onc.2010.102

Schardt, J. (2020). [The use of immune checkpoint inhibitors in routine oncology]. *Zeitschrift Fur Rheumatologie*, *79*(8), 809–817. https://doi.org/10.1007/s00393-020-00876-2

Scheel, C., & Weinberg, R. A. (2012). Cancer stem cells and epithelial–mesenchymal transition: Concepts and molecular links. *Seminars in Cancer Biology*, *22*(5–6), 396–403. https://doi.org/10.1016/j.semcancer.2012.04.001

Schmidt, J. M., Panzilius, E., Bartsch, H. S., Irmler, M., Beckers, J., Kari, V., … Scheel, C. H. (2015). Stem-cell-like properties and epithelial plasticity arise as stable traits after transient twist1 activation. *Cell Reports*, *10*(2), 131–139. https://doi.org/10.1016/j.celrep.2014.12.032

Shaulian, E., & Karin, M. (2002). AP-1 as a regulator of cell life and death. *Nature Cell Biology*, *4*(5), E131-6. https://doi.org/10.1038/ncb0502-e131

Shi, J., Wang, Y., Zeng, L., Wu, Y., Deng, J., Zhang, Q., … Zhou, B. P. (2014). Disrupting the interaction of BRD4 with diacetylated Twist suppresses tumorigenesis in basal-like breast cancer. *Cancer Cell*, *25*(2), 210–225. https://doi.org/10.1016/j.ccr.2014.01.028

Sikandar, S. S., Kuo, A. H., Kalisky, T., Cai, S., Zabala, M., Hsieh, R. W., … Clarke, M. F. (2017). Role of epithelial to mesenchymal transition associated genes in mammary gland regeneration and breast tumorigenesis. *Nature Communications*, *8*(1), 1669. https://doi.org/10.1038/s41467-017-01666-2

Sinha, V. C., & Piwnica-Worms, H. (2018). Intratumoral Heterogeneity in Ductal Carcinoma In Situ: Chaos and Consequence. *Journal of Mammary Gland Biology and Neoplasia*, *23*(4), 191–205. https://doi.org/10.1007/s10911-018-9410-6

Sørlie, T., Perou, C. M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., … Børresen-Dale, A. L. (2001). Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(19), 10869–10874. https://doi.org/10.1073/pnas.191367098

Sørlie, T., Tibshirani, R., Parker, J., Hastie, T., Marron, J. S., Nobel, A., … Botstein, D. (2003). Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proceedings of the National Academy of Sciences*, *100*(14), 8418–8423. https://doi.org/10.1073/pnas.0932692100

Spaderna, S., Schmalhofer, O., Wahlbuhl, M., Dimmler, A., Bauer, K., Sultan, A., … Brabletz, T. (2008). The transcriptional repressor ZEB1 promotes metastasis and loss of cell polarity in cancer. *Cancer Research*, *68*(2), 537–544. https://doi.org/10.1158/0008-5472.CAN-07-5682

Sporn, M. B., & Lippman, S. M. (2003). Agents for Chemoprevention and Their Mechanism of Action. In D. Kufe, R. Pollock, & R. Weichselbaum (Eds.), *Holland-Frei Cancer Medicine.*

6th edition. Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK12522/

Thiery, J. P., Acloque, H., Huang, R. Y. J., & Nieto, M. A. (2009). Epithelial-mesenchymal transitions in development and disease. *Cell*, *139*(5), 871–890. https://doi.org/10.1016/j.cell.2009.11.007

Thomssen, C., & Wand, D. (2012). Hereditärer Brustkrebs. *Der Onkologe*, *18*(3), 216–223. https://doi.org/10.1007/s00761-011-2095-8

Trelstad, R. L., Hay, E. D., & Revel, J. D. (1967). Cell contact during early morphogenesis in the chick embryo. *Developmental Biology*, *16*(1), 78–106. https://doi.org/10.1016/0012-1606(67)90018-8

Tsai, J. H., Donaher, J. L., Murphy, D. A., Chau, S., & Yang, J. (2012). Spatiotemporal regulation of epithelial-mesenchymal transition is essential for squamous cell carcinoma metastasis. *Cancer Cell*, *22*(6), 725–736. https://doi.org/10.1016/j.ccr.2012.09.022

Vandewalle, C., Van Roy, F., & Berx, G. (2009). The role of the ZEB family of transcription factors in development and disease. *Cellular and Molecular Life Sciences : CMLS*, *66*(5), 773–787. https://doi.org/10.1007/s00018-008-8465-8

Vannier, C., Mock, K., Brabletz, T., & Driever, W. (2013). Zeb1 regulates E-cadherin and Epcam (epithelial cell adhesion molecule) expression to control cell behavior in early zebrafish development. *The Journal of Biological Chemistry*, *288*(26), 18643–18659. https://doi.org/10.1074/jbc.M113.467787

Wang, G., Pan, J., Zhang, L., & Wang, C. (2019). Overexpression of grainyhead-like transcription factor 2 is associated with poor prognosis in human pancreatic carcinoma. *Oncology Letters*, *17*(2), 1491–1496. https://doi.org/10.3892/ol.2018.9741

Watanabe, K., Villarreal-Ponce, A., Sun, P., Salmans, M. L., Fallahi, M., Andersen, B., & Dai, X. (2014). Mammary morphogenesis and regeneration require the inhibition of EMT at terminal end buds by Ovol2 transcriptional repressor. *Developmental Cell*, *29*(1), 59–74. https://doi.org/10.1016/j.devcel.2014.03.006

Wen, H. Y., & Brogi, E. (2018). Lobular Carcinoma In Situ. *Surgical Pathology Clinics*, *11*(1), 123–145. https://doi.org/10.1016/j.path.2017.09.009

Werner, S., Frey, S., Riethdorf, S., Schulze, C., Alawi, M., Kling, L., … Assmann, V. (2013). Dual roles of the transcription factor grainyhead-like 2 (GRHL2) in breast cancer. *The Journal of Biological Chemistry*, *288*(32), 22993–23008. https://doi.org/10.1074/jbc.M113.456293

Wickham, H. (2016). *ggplot2*. https://doi.org/10.1007/978-3-319-24277-4

Williams, E. D., Gao, D., Redfern, A., & Thompson, E. W. (2019). Controversies around epithelial-mesenchymal plasticity in cancer metastasis. *Nature Reviews. Cancer*, *19*(12), 716–732. https://doi.org/10.1038/s41568-019-0213-x

Wu, Y., & Zhou, B. P. (2010). Snail: More than EMT. *Cell Adhesion & Migration*, *4*(2), 199–

203. https://doi.org/10.4161/cam.4.2.10943

Xiang, X., Deng, Z., Zhuang, X., Ju, S., Mu, J., Jiang, H., … Zhang, H.-G. (2012). Grhl2 determines the epithelial phenotype of breast cancers and promotes tumor progression. *PloS One*, *7*(12), e50781. https://doi.org/10.1371/journal.pone.0050781

Xu, J., Lamouille, S., & Derynck, R. (2009). TGF-beta-induced epithelial to mesenchymal transition. *Cell Research*, *19*(2), 156–172. https://doi.org/10.1038/cr.2009.5

Yang, J., Mani, S. a, Donaher, J. L., Ramaswamy, S., Itzykson, R. a, Come, C., … Weinberg, R. a. (2004). Twist, a master regulator of morphogenesis, plays an essential role in tumor metastasis. *Cell*, *117*(7), 927–939. https://doi.org/10.1016/j.cell.2004.06.006

Yang, J., & Weinberg, R. A. (2008). Epithelial-mesenchymal transition: at the crossroads of development and tumor metastasis. *Developmental Cell*, *14*(6), 818–829. https://doi.org/10.1016/j.devcel.2008.05.009

Yu, M., Bardia, A., Wittner, B. S., Stott, S. L., Smas, M. E., Ting, D. T., … Maheswaran, S. (2013). Circulating breast tumor cells exhibit dynamic changes in epithelial and mesenchymal composition. *Science (New York, N.Y.)*, *339*(6119), 580–584. https://doi.org/10.1126/science.1228522

Zanconato, F., Forcato, M., Battilana, G., Azzolin, L., Quaranta, E., Bodega, B., … Piccolo, S. (2015). Genome-wide association between YAP/TAZ/TEAD and AP-1 at enhancers drives oncogenic growth. *Nature Cell Biology*, *17*(9), 1218–1227. https://doi.org/10.1038/ncb3216

Zhang, B., Zhang, Z., Xia, S., Xing, C., Ci, X., Li, X., … Dong, J.-T. (2013). KLF5 activates microRNA 200 transcription to maintain epithelial characteristics and prevent induced epithelial-mesenchymal transition in epithelial cells. *Molecular and Cellular Biology*, *33*(24), 4919–4935. https://doi.org/10.1128/MCB.00787-13

Zheng, X., Carstens, J. L., Kim, J., Scheible, M., Kaye, J., Sugimoto, H., … Kalluri, R. (2015). Epithelial-to-mesenchymal transition is dispensable for metastasis but induces chemoresistance in pancreatic cancer. *Nature*, *527*(7579), 525–530. https://doi.org/10.1038/nature16064

Götz, head of the Institute of Stem Cell Research, for giving me the opportunity to perform my doctoral research at her institute.

I am also very grateful for all of my friends who always had a sympathetic ear and supported me during my time as a doctorate researcher.

Last but not least, I would like to express my deepest gratitude to my mum, my dad, my two sisters, and my partner. Thank you for your love, your constant support, your understanding, your motivating words, and your belief in me.