A Role for Bivalent Genes in Epithelial to Mesenchymal Transition

Francisco Sadras (B.Sc, Hns)

Thesis submitted for the degree of

Doctor of Philosophy

Faculty of Sciences

School of Biological Sciences

Department of Molecular & Cellular Biology

Affiliated with

Gene Regulation Laboratory

Centre for Cancer Biology

IMVS, SA Pathology

April 2017

SA PATHOLOGY

THE UNIVERSITY of ADELAIDE

Centre for Cancer Biology

Contents

Abstract

Epithelial to Mesenchymal Transition (EMT) is an important and complex cellular process in embryonic development, wound healing and tumour progression. EMT is often triggered or facilitated through the action of master EMT transcription factors including ZEB1 and TWIST. It has been proposed that prior to malignant progression a subset of tumour cells undergo an EMT which facilitates the development of key malignant properties. In recent years, a clear link between developmental and cancer associated EMT has triggered an increased interest in the role of developmental EMT genes in a cancer setting. Many key developmental genes have a bivalent or poised promoter signature which changes to active during differentiation; this is believed to elicit a faster response time in comparison to an exclusively repressed promoter.

Owing to the relevance of EMT in development and pathologies my thesis aimed to answer the core question of whether bivalent genes are relevant in malignant EMT. To answer this question I undertook four aims:

1. To uncover novel bivalent genes that were activated in an EMT.
2. To characterise the expression and role of ADM2, PLEKHO1 and RASA3 in EMT.
3. To characterise ZEB1 isoform expression during EMT.
4. To identify novel ZEB1 target genes.

Aim 1: We utilised a common model of human EMT, whereby human mammary epithelial cells (HMLE) undergo EMT in response to TGFβ to become mesenchymal (mesHMLE). We performed ChIP-seq against histone3 lysine4 tri-methylation (H3K4me3) and histone3 lysine27 tri-methylation (H3K27me3) alongside RNA-seq to identify genes that changed

from a bivalent to an active epigenetic signature with concomitant changes to RNA levels. From this data set 429 genes that exhibited this epigenetic change including the well-known EMT factors ZEB1 and TWIST1. From this list four genes that were not previously associated with a bivalent signature were studied in detail. Three of these, ADM2, PLEKHO1 and RASA3, had not previously been associated with EMT but had EMT associated properties, while one, ZEB1 was a well-established master EMT transcription factor.

Aim 2: Chromatin immunoprecipitation, ChIP-reChIP was used to confirm the change in epigenetic marks for ADM2, PLEKHO1 and RASA3 promoters alongside a combination of molecular and bioinformatics analyses to determine expression levels in epithelial and mesenchymal cell lines. Cellular migration assays where levels of these genes were manipulated showed that ADM2 and PLEKHO1 have both an individual and a synergistic effect on migration while RASA3 did not affect migration.

Aim 3: ZEB1 isoform expression during EMT was analysed and it was determined that there was no significant change in relative expression over this process.

Aim 4: ENCODE ZEB1 ChIP-seq was analysed to obtain insights into ZEB1 binding and to identify novel potential targets of importance to EMT. Established ZEB1 target genes such as *CDH1* and *CRB3* were identified and 26 novel genes with known or potential roles in EMT were chosen for further study. Of these, *F11R* and *INADL* were found to be ZEB1 responsive. Direct ZEB1 binding was confirmed through ChIP-qPCR. Interestingly, both of these genes are associated with tight-junctions as is the previously established ZEB1 target *CRB3.* This strongly implicates ZEB1 in mediating tight-junction regulation.

While bivalent genes have not been ignored in the field of EMT they have, so far, been understudied. My work addressed this issue and identified ADM2 and PLEKHO1 as novel EMT associated genes that play an important role in migration. I also established ZEB1, a master regulator of EMT, as a bivalently regulated gene. These contributions help establish bivalently regulated genes as a valuable, underutilised resource for the identification of novel EMT genes.

Declaration

This work contains no material which has been accepted for the award of any other degree in any university or other tertiary institution to Francisco Sadras and, to the best of my knowledge and belief, contains no material previously published or written by another person except were due reference has been made in the text.

I give consent to this copy of my thesis, when deposited in the University Library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968.

I also give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library catalogue, and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

Francisco Sadras

Date:

Preface

Work presented in this thesis was generated through collaboration and assistance from the following:

Dr Joanna Attema, Andrew Bert, Dr David Lawrence and Dr Katherine Pillman produced the HMLE-mesHMLE ChIP and RNA-seq data sets.

Dr Phillip Gregory and Suraya Roslan provided the expression data for ZEB1, CDH1 and VIM in Table 15.

Dr Katherine Pillman provided guidance in the analysis of the ENCODE ZEB1 ChIP-seq data set in Chapter 5.

Andrew Bert provided the qRT-PCR data for Figure 18. He also provided chromatin from the mesHMLE cell line for the experiment in Figure 6 and the qRT-PCR data for the HMLE cells.

Acknowledgments

I would like to thank, first and foremost my primary supervisor, Professor Gregory Goodall. You provided the guidance and support that helped mould both me and my thesis. In addition to the helpful and sharp insights provided with the research you provided a strong motivating force and a very understanding and patient ear when times were difficult. For every thesis it is safe to say that it would not be completed without the supervisor, in this case that is true many times over.

I am grateful to my co-supervisor Dr. Joanne Attema who helped me settle into the lab and develop my own projects while fostering my laboratory skills. Your attention to detail and punctilious ways are aspects which I wish I had incorporated into my personal work sooner.

I am thankful to my co-supervisor Associate Professor Simon Conn who had an innovative technique or approach to solve any problem and for stimulating conversations.

Thanks go to my co-supervisor Dr. Phillip Gregory for valuable insights into the HMLE-mesHMLE cell line model and always lending a willing ear.

I would like to thank my co-supervisor Dr. Murray Whitelaw for helpful conversations and reagents.

I would also like to express my gratitude to all members of the Goodall Laboratory. Thanks to Matthew Anderson and Suraya Roslan for providing helpful laboratory advice paired with a strong sense of humour. Thanks to Andrew Bert for his help with ChIP and qRT-PCR experiments and general laboratory technique, your insights and technical skills were

invaluable. Thanks to James Conway for a lot of laughs and a friendship that began with our first year summer scholarships. Thanks to Dr. Katherine Pillman for help with analysing the ZEB1 ChIP-seq data and for fostering my interest in programming. Thanks to my fellow PhD student Victoria Arnet for her help with the Incucyte. Thanks to Catherine Phillips and Kaitlin Scherer for their technical assistance and good humour. Thanks to Rosemary Sladic for always being cheerful and keeping the lab organised. And thank you to all other members of the lab for creating a great work environment.

I would like to thank the Bioinformatics team, especially Dr. David Lawrence, Dr. Katherine Pillman, Dr. Paul Wang and Dr. John Toubia for encouraging me to attend bioinformatics workshops and expand my skills, and for the pizza, beer and board games nights.

Thanks to Dr Xiaochun Li for her help with the Incucyte and for being a great conference partner.

I would like to thank my family, my parents, Victor and Ana, for providing me with the support and opportunities necessary to start a PhD and for all their help throughout. I would like to thank my sister, Teresa, for providing helpful advice and for providing a path for me to follow even though our journeys will soon branch out.

Finally, I would like to thank my partner, Jiyoon Choi. Without your endless love and encouragement throughout this journey I would have struggled to find the inspiration and motivation needed to complete this thesis.

List of Figures

List of Tables

Abbreviations

3'UTR	3' untranslated region
ADM2	Adrenomedullin 2
BLAT	Basic Local Alignment Tool
Bp	base pairs
BPE	Bovine Pituitary Extract
BSA	Bovine Serum Albumin
cDNA	complementary DNA
ChIP	Chromatin cross linking immunoprecipitation
ChIP-seq	Chromatin cross linking immunoprecipitation sequencing
cm	Centimetre
CO2	carbon dioxide
Co-IP	co-Immunoprecipitation
Ct	Cycle threshold
DAPI	4'-6-Diamidino-2-phenylindole
DMEM	Dulbecco's Modified Eagle Medium
DNA	Deoxyribonucleic acid
EDTA	Ethylenediaminetetraacetic acid
EMT	Epithelial to mesenchymal transition
FCS	Foetal calf serum
FL	Firefly luciferase
F11R	F11 Receptor
GFP	Green fluorescent protein
hrs	Hours
HEK293	Human embryonic kidney 293
Hg18	Human genome version 18
Hg19	Human genome version 19
INADL	InaD-like
kb	Kilobase
M	Molar
MDCK	Madin Darby canine kidney
mg	milligram
min	minute
miR	microRNA
miRNA	microRNA
ml	millilitre
mM	millimolar
mRNA	messenger RNA
NaCl	sodium chloride
ng	nanogram
nM	nanomolar
nt	Nucleotide
ORF	Open reading frame

PAGE	Polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PLEKHO1	Pleckstrin Homology Domain Containing, Family O Member 1
pM	pico Molar
PTEN	Phosphate and tensin homolog
qPCR	quantitative polymerase chain reaction
qRT-PCR	quantitative reverse transcription polymerase chain reaction
RASA3	RAS P21 Protein Activator 3
rcf	Relative centrifugal force
RL	renilla luciferase
RNA	ribonucleic acid
rpm	revolutions per minute
rRNA	ribosomal RNA
RT	reverse transcriptase
sec	seconds
SDS	sodium dodecyl sulphate
siRNA	small interfering RNA
TBE	tris buffered saline
TE	Tris EDTA
U	Uracil
µg	Microgram
µl	Microliter
µm	micromolar
UTR	untranslated region
ZEB	Zinc finger E-Box binding homeobox

Chapter 1 Introduction

1.1 Preamble

The aim of this thesis was

1. To uncover novel bivalent genes that were activated in an EMT
2. To characterise the expression and role of ADM2, PLEKHO1 and RASA3 in EMT
3. To characterise ZEB1 isoform expression during EMT
4. To identify novel ZEB1 target genes

Bivalent genes have a combination of activating and repressive epigenetic marks, Histone3 lysine 4 trimethylation (H3K4me3) and Histone 3 lysine 27 trimethylation (H3K27me3). These marks can often be found cohabiting promoter regions, particularly during development [1]. A bivalent promoter is in an intermediate state, where it is open enough for transcription machinery to bind, and even for some low level transcription to occur, but not sufficiently open for high levels of transcription. This is thought to leave genes in a "poised" state where they can rapidly be switched on by the removal of the repressive marks, if needed. As such, genes with this bivalent signature are typically associated with rapid cellular changes, such as those that occur during lineage commitment, tissue specification and wound healing and, we posit, EMT. An additional inference that can be made from the speed at which these bivalent genes can be activated is that, genes with this signature are important for early response or to help trigger these changes.

The cancer epigenome typically undergoes dramatic global changes to DNA methylation and histone modification profiles as well as mutations or altered expression of key chromatin

modifying enzymes [2]. As one would expect, these changes result in global dysregulation of gene expression profiles some of which assist in tumour progression [3]. These epigenetic changes can lead to both the silencing of tumour suppressor genes as well as the activation of oncogenes [4]. Since epigenetic mutations and changes are heritable across cellular division, advantageous mutations are rapidly selected for in a growing tumour [2]. These factors led to my interest in investigating the role of bivalent genes in EMT.

This literature review covers a basic introduction into cancer and breast cancer, EMT, ZEB1 and epigenetics.

1.2 Cancer

1.2.1 Cancer is a global issue

Cancer is the current leading cause of death worldwide with 8.2 million cancer deaths in 2012 [5]. The vast majority of cancer related deaths are attributed to secondary metastases rather than the primary tumour [6], metastasis is explained in more detail in 1.13 and 1.1.4. Locally, cancer is a terrible affliction which affects ~2% of the Australian population. Beyond this number, a great many people are affected by cancer through its devastating effects on family and friends. Cancer is the disease of the modern age with incidence increasing over past decades as life spans increase and eating habits deteriorate [7]. For these reasons cancer is the focus of much of modern medical research. Cancer research received the largest proportion of NHMRC grants between 2011-2015 with an average allocation of ~15% of total funds per year [8].

So what is cancer? At its core, cancer is a disease of uncontrolled growth. Cancer cells initiate and maintain this growth through a combination of ten well recognised hallmarks, defined in the landmark reviews of Hanahan and Weinberg (Figure 1) [9, 10]. These hallmarks include aberrant proliferation, apoptosis resistance and a capacity to invade and metastasise. My research focussed on the hallmarks of invasion and metastasis with a small foray into proliferation.

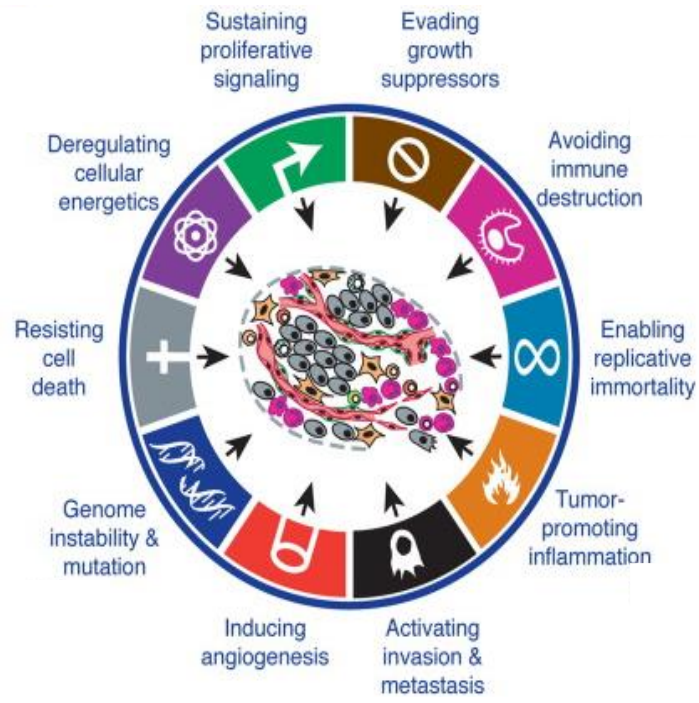


Figure 1 The hallmarks of cancer including the six original, and four additional, hallmarks.

Adapted from Hanahan and Weinberg [1].

These hallmarks were traditionally associated with genetic mutations [11, 12]. However, research is increasingly suggesting that alongside genetic mutations, epigenetic mutations play a key role in cancer development [2]; this is further discussed in section 1.3.3.

Tumours can arise from any cell in any tissue of the body resulting in a large variety of cancers with different mutations and properties. A large part of what makes cancer so

difficult to treat is the large variety of tumours that arise as a result of this initial variation. Compounding on this difficulty, tumours arising from the same tissue may also be very heterogeneous. However, some mutations and expression changes are commonly seen in solid tumours or in specific subtypes of tumours and these mutations are often chosen as drug targets, such as epidermal growth factor receptor [13, 14]. The primary goal of my thesis was to serve as a pilot study for whether bivalent epigenetic regulation could serve as a marker for genes relevant in malignant EMT. In addition, the discovery of novel genes could potentially lead to the development of novel therapies, although that was beyond the scope of my study.

1.2.2 Breast cancer

Breast cancer is the most common cancer in women and the second most common cancer overall with 1.7 million new cases in 2012 [5]. It is a major health issue both in the developed and the developing world, though it has a higher rate of incidence in the developed world [15]. Breast cancer is a highly heterogeneous disease that can be divided into multiple subtypes. These subtypes are primarily classified based on their genetic composition and tissue of origin. The most commonly used molecular classification is based on the presence of certain extracellular receptors, Progesterone Receptor positive (PR), Estrogen Receptor (ER) positive or human epidermal growth factor type 2 receptor positive (HER2) [16, 17]. In the first two cases the tumour is termed hormone sensitive, as the tumours depend on estrogen or progesterone for continued growth. As a consequence, hormone treatment can greatly improve prognosis. For HER2 positive tumours this receptor can be targeted using drugs such as trastuzumab and pertuzumab. Triple negative breast cancers lack these three receptors, consequently they are resistant to these forms of therapy and are much more aggressive. This combination of factors result in poorer prognosis [18]. In all cases, once

tumours are metastatic and malignant, treatment becomes significantly more difficult and prognosis deteriorates.

1.2.3 Benign and malignant tumours

Solid tumours, including breast, can be separated into two main categories, benign and malignant. Benign tumours are generally not life threatening and can generally be effectively treated by surgery. In contrast, malignant tumours are responsible for the vast majority of cancer related deaths and surgery is not an effective treatment [6]. The key difference between these two tumour types is that benign tumours are surrounded by a fibrous capsule and cannot spread throughout the body while malignant tumours can through a process known as metastasis. Consequently, increasing our understanding of how tumours become metastatic is crucial to combat this disease.

1.2.4 EMT and MET in development and cancer metastasis

To develop from benign to malignant tumours must overcome numerous barriers [19]. As the first step, some cells must dissociate from the primary tumour mass and enter the circulatory or lymphatic system. Secondly they must survive in this foreign and hostile environment, withstanding anoikis. Thirdly they must extravasate and invade into a new niche. Finally they must colonise this new niche and regain their growth capacity before they are clinically relevant metastases. Only a fraction of cells pass through each subsequent stage with just ~1% of tumour micro-metastases undergoing the final step to regain growth potential [20]. This multi-step process, through which primary tumour cells become metastatic, is thought to require a conversion from an epithelial to a mesenchymal phenotype.

Epithelial cells are the most common cells in our bodies; they form the basic structures of our tissues, and they are also the primary component of solid tumours. They are characterised by basal-apical polarity, very limited mobility, are differentiated and require attachment to other cells to survive. Mesenchymal cells are much rarer, in adults these exist almost exclusively as stem cells and fibroblasts, apart from during wound healing. In contrast to epithelial cells, mesenchymal cells are undifferentiated, have indistinct polarity, can withstand anoikis and can freely migrate. All of these properties are thought to be required for tumours to become metastatic potentially implicating a transition between these states as key for cancer progression.

It was first observed that cells could transition from an epithelial to a mesenchymal morphology in 1896 [21]. Our modern terminology derives from a 1968 paper by Hay [22] where the transition was referred to as an Epithelial Mesenchymal Transformation, eventually becoming Epithelial to Mesenchymal Transition (EMT). The reverse process, a Mesenchymal Epithelial Transition (MET) reverses the changes induced by an EMT reverting cells to an immotile epithelial proliferative phenotype. These transitions were initially described in their physiological roles during development and wound healing [21]. Interestingly, wound healing and cancer have been associated for almost a century. One of the earliest written recognitions of this similarity was published by Montrose T. Burrows in 1924 [23]. Throughout the following decades research on wound healing and cancer continued [24] and in 1986 Dvorak coined the idea that a tumour is a "wound that does not heal" [25]. Today the focus has migrated towards the process of EMT in association with metastasis. An EMT followed by an MET is thought to explain the process of metastasis as follows.

As EMT occurs in an epithelial based tumour, some of the tumour cells will gain mesenchymal properties and will dissociate from the primary tumour and potentially enter the circulatory system. Many are killed by the immune system, while others fail to locate a hospitable niche to set up secondary metastases. However, in a minority of cases these metastasising cells will extravasate and find a niche where they can survive and form micro metastases. The majority of these will remain dormant and not pathological, however a minority will undergo an MET and trigger the formation of new tumours which can restart this process [20]. Once tumours are metastatic, treatment becomes much more difficult as surgery is rarely an option and the tumours will diverge genetically making chemo or genetic therapy less likely to be successful.

1.2.5 Transcription factors in cancer and EMT

EMT can be triggered by a large variety of signalling molecules including TGFβ and WNT [26, 27]. The primary downstream effectors of these signalling pathways in EMT are a shared group of transcription factors known as master EMT transcription factors [28]. These include, SNAIL, SLUG, TWIST, ZEB1 and ZEB2. These transcription factors have some shared targets, in particular E-cadherin – a key epithelial cell adhesion gene. However, each also has their own individual targets. These transcription factors are of different importance, and are activated in a different order, depending on the system. An analysis on four microarray studies looking at these master EMT transcription factors by Aigner *et al* [29] showed that ZEB1 resulted in the strongest repression of epithelial genes.

Human ZEB1 operates in a double negative feedback loop with the epithelial-promoting miR-200 family, as described by our laboratory, and others [30-32]. This dynamic between two key players in the EMT is thought to be critical for determining cell state. High levels of

miR-200 promote epithelial morphology and transcription profiles, while high levels of ZEB1 promote mesenchymal morphology and transcription profiles. This combination of properties, in addition to some recent data, discussed in Chapter 3, made ZEB1 of interest to my research.

1.2.6 ZEB1

ZEB1 (δEF1/AREB6/zfhx1a/TCF8) is a vertebrate zinc finger E-box binding homeobox transcription factor with important roles in development, posterior polymorphous corneal dystrophy (PPCD), EMT and tumour metastasis and progression [33]. ZEB1 has C and N terminal zinc finger domains flanking a central homeodomain and C-terminal binding protein Interaction Domains (CID) [33]. ZEB1 has a homologue, ZEB2 which has a highly similar protein structure. While the overall sequence similarity between ZEB1 and ZEB2 is 60%, there is a 92% conservation ratio within the zinc finger domains and the homeodomain. This implies that these regions are functionally important and that the ZEB proteins likely share binding partners, targets and function. I will focus on describing ZEB1 for the reasons previously mentioned.

ZEB1 was first discovered in 1991 and was described as a negative repressor of delta 1-crystallin and interleukin 2 [34, 35]. Subsequently, ZEB1 has been found to repress important epithelial genes including E-cadherin and the two miR-200 clusters. Research has focussed on its role in EMT and cancer metastasis where it is considered a key regulator and driver of tumour progression [29, 30, 32, 36-39]. Consistent with this, knockdown of ZEB1 in metastatic cancer cell lines decreased tumour cell proliferation and impeded the formation of secondary metastases [40-43] while ZEB1 up-regulation was associated with high grade tumours and increased metastasis [44-50].

ZEB1 expression is broad but tightly regulated. During mouse development ZEB1 is first detected in the neuroepithelium and lateral plate mesoderm at E8.5, in neural crest derivatives and limb bud mesenchyme at E9.5, skeletal precursors at E11.5 [51] and cornea by E13.5 [52]. In adult humans ZEB1 is widely expressed, but most abundantly in the bladder, uterus, skeletal muscle, aorta, and thymus [53].

In accordance with these observations, ZEB1 knockout mice show severe defects during development in thymus, skeletal, corneal and neural tissues and exhibit ectopic expression of epithelial genes. As a consequence of these defects homozygous knockouts die perinatally due to cyanosis, while heterozygous knockouts are viable, [51, 52, 54]. These defects are thought to arise due to a combination of a lack of EMT and ectopic epithelial gene expression during key stages of development, reinforcing the importance of ZEB1 in EMT as a negative regulator of epithelial genes [51].

Homeodomains and zinc fingers can be important for both DNA and protein interactions; however, an early study showed that the ZEB1 homeodomain did not possess DNA binding activity [55]. Accordingly, while the canonical homeodomain DNA binding sequence is WFxNx, in ZEB1 this is WFEKM. This is likely to impair or completely ablate its capacity to bind DNA [56]. In contrast, both sets of zinc fingers have DNA binding activity. The N terminal fingers were described as preferential activators with a GTCACCTGT consensus sequence while the C terminal fingers were originally described as preferential repressors with a TACCTG consensus sequence [55, 57]. These motifs have been redefined as Z-box consensus sequence. Despite this early evidence for ZEB1 as an activator, most subsequent work has focussed on the role of ZEB1 as a repressor [30].

Many functional Z-box consensus sequences are evolutionarily conserved as mutation of this motif inhibits ZEB1 binding [30, 31]. There is some discrepancy in the literature regarding whether ZEB1 requires a paired binding site or if a single site is sufficient. An early study showed that a single Z-box motif repressed a reporter construct by ~40% while two motifs resulted in ~80% repression [58]. This suggests that a single site is necessary and sufficient for ZEB1 occupancy while more sites augment repression. Most subsequent studies claimed that ZEB1 required two binding sites for repression with a spacer region of a flexible distance with a functional range of at least 14-44 bp [30, 31, 57]. Therefore, two models exist: one states that ZEB1 can bind and repress target genes with a single binding site while the other claims a minimum of two are necessary. One explanation for this discrepancy is that, as a double binding site yields a stronger effect, these target genes are easier to detect and validate. I investigate this in Chapter 5.

ZEB1 is a key regulator of EMT and cancer metastasis. Its up-regulation can induce an EMT and is strongly associated with high tumour grade while the opposite is true for its down-regulation. ZEB1 is an effector for many oncogenic pathways including TGF-β, Wnt, NF-κB and FGF2 signalling. ZEB1 is primarily reported as a transcriptional repressor of several important epithelial genes, in some cases targeting transcription factors or miRNAs (Table 1), thus indirectly affecting the expression of many genes.

Despite extensive research on ZEB1, we only have a partial understanding of various facets of its function and regulation. These include ZEB1's post-translational modifications, binding partners, target genes and their regulation. While the interaction and effects on target genes of some of ZEB1's binding partners, like CtBP1, are well known, others remain poorly

understood and it is likely that several remain to be identified and characterised. Likewise there is fragmented information for its target genes. I investigate this area in Chapter 5.

Table 1 Summary of ZEB1 target genes and their role in cancer and maintenance of the epithelial phenotype.
Arrows indicate up or down-regulation by ZEB, a tick indicates that the gene has a known role as tumour suppressor, oncogene or contributes to epithelial phenotype

Gene	Regulation	Role			References
		Tumour suppressor	Oncogene	Epithelial phenotype	
miR-200	↓	✓		✓	[30-32, 36]
miR-203	↓	✓		✓	[59]
miR-34	↓	✓		✓	[60]
CRB3	↓	✓		✓	[29]
HUGL2	↓	✓		✓	[29]
PATJ	↓	✓		✓	[29]
uPA	↓	✓		✓	[61]
P73	↓	✓		✓	[62]
CDH1	↓	✓		✓	[63] [57] [64]
EPCAM	↓	✓	✓	✓	[65] [66]
PKP3	↓	✓		✓	[67]
ERα	↓	✓	✓	✓	[68]
PAI-1	↑		✓		[61]
mir-199a-3p/214	↑		✓		[69]
VDR	↑	✓			[70]
MMP2	↑		✓		[71]

1.3 Epigenetics

1.3.1 An overview

Epigenetics refers to "the structural adaptation of chromosomal regions so as to register, signal or perpetuate altered activity states" [72]. This typically involves histone modifications and DNA methylation. I will focus on the role of histone modifications in cancer and EMT.

Cellular DNA is condensed and packaged as chromatin. This is a dynamic structure that, at the base level, consists of units of 147 bases of DNA wrapped around nucleosomes [73]. Nucleosomes are primarily composed of a histone octamer that consists of pairs of the four core histone proteins (H2A, H2B, H3, H4) [74]. These histone proteins have an overall positive charge which interacts with the negative DNA charge facilitating binding and efficient compaction. The octamer of core histone proteins are held together by a linker histone protein, H1. Core histone proteins are largely globular but possess a histone tail with a high concentration of lysine and arginine residues which can undergo a variety of covalent post-translational modifications. These modifications act in concert to determine the chromatin state. There are dozens of possible histone modifications, their combination further enhances the complexity and diversity of epigenetic effects, this complexity is known as the histone code [75].

Very broadly, there are three main types of histone modifications, activating, repressive and recruiting. Activating modifications help promote gene transcription, typically by opening or relaxing chromatin. This allows for greater access to the DNA by RNA polymerase and other transcription factors. Activating acetylation marks typically act by neutralising positive charges on the histone tails decreasing the attraction between histones and DNA,

consequently, hyper acetylated promoters are associated with increased transcriptional activity [75, 76].

In contrast to this change to histone charge and a direct link between histone and DNA affinity produced by histone acetylation, histone methylation can be either repressive or activating depending on the residue modified and the number of methyl groups added. The methyl groups tend to act as scaffolding for transcription factors and other epigenetic modifiers with different modifications attracting specific factors [77, 78]. For example, H3K4me3 is associated with active promoters while H3K27me3 is associated with inactive promoters [2, 75].

Repressive modifications primarily function by decreasing transcription machinery access to DNA. H3K4me3 in particular is associated with facultative heterochromatin [79]. This is a repressed chromatin state which is maintained throughout cell division and can recruit further repressive epigenetic marks such as DNA methylation [80]. Prior to transcription these repressive marks must be removed. As I previously mentioned, a key aspect of how histones function is the interactions and cumulative effects that arise from neighbouring marks. My work has focussed on the study of the combination of one specific activating and one specific repressive epigenetic mark, H3K4me3 and H3K27me3. Promoters possessing this combination are known as bivalent genes.

1.3.2 Bivalent genes in development and EMT

Bivalent marks are often present near promoter regions, particularly during development [1]. A promoter with this combination is in an intermediate state between activated and repressed, where it is sufficiently open for transcription machinery to bind, and even for some

low level transcription to occur, but not sufficiently open for high levels of transcription. These characteristics are thought to leave genes in a "poised" state where they can rapidly be switched on by the removal of the repressive marks [1, 81, 82]. As such, genes with this bivalent signature are typically associated with rapid cellular changes, such as those that occur during lineage commitment, tissue specification and EMT [1, 83]. An additional inference that can be made from the speed at which these bivalent genes are activated is that they are important for early response or to help trigger these changes. Additional evidence for this occurs during development where for developmental regulators, such as the *Pax6*, *MyoD* and *Hox* genes, are enriched for bivalent chromatin marks [82].

1.3.3 Epigenetic dysregulation in cancer

The cancer epigenome often undergoes dramatic global changes to DNA methylation and histone modification profiles as well as mutations, or altered expression, of key chromatin modifying enzymes [2]. As one would expect, these changes result in global dysregulation of gene expression profiles some of which assist in tumour progression [3]. These epigenetic changes can lead to both the silencing of tumour suppressor genes as well as the activation of oncogenes [4]. Since epigenetic mutations and changes are heritable across cellular division, these changes can be rapidly selected for in a growing tumour [2]. Some well-studied examples of this include the propensity for tumours to undergo an increase in DNA methylation. For example, a recent study in breast cancer showed that p53 and CDH1 had strong aberrant DNA methylation in the PAM50 breast cancer subtype [84]. As these genes clearly code for very strong tumour and metastasis supressing genes, this reinforces the importance of epigenetic regulation in cancer.

Interestingly, despite my observations in the previous section on bivalent genes in EMT and cancer, the Polycomb Repressor Complex (PRC2), and the H3K27me3 marks that it deposits are often upregulated in cancer cell lines [85]. This suggests that PRC2 could be acting as both a tumour suppressor and an oncogene, depending on the situation as follows. In epithelial cells and benign tumours, specific oncogenes are maintained in a bivalent state or in facultative heterochromatin; however, after an EMT or the transition to a malignant tumour, these oncogenes lose the repressive mark and become expressed while tumour suppressors lost their active mark and become repressed. This suggests that there is additional regulation of the PRC2 complex and demethylases occurring in a cell type specific manner.

1.4 Thesis structure

In Chapter 2 I outline the Materials and Methods used throughout my project, including molecular, cellular and biometric techniques.

In Chapter 3 I identified genes that underwent a bivalent to active epigenetic shift during HMLE EMT. I chose three genes to study in detail, ADM2, PLEKHO1 and RASA3 and confirmed that they had genuine bivalent promoters and expected expression patterns in a variety of epithelial and mesenchymal lines using *in vitro* and *in silico* methods. Subsequently, I investigated the effect of modulating their expression on proliferation and migration and determined that ADM2 and PLEKHO1 expression positively correlated with migration and had a small effect on proliferation while RASA3 had no effect on these properties. Finally I determined that the PLEKHO1-mediated changes were reliant on EGF signalling and that ADM2 and PLEKHO1 act in a co-operative manner.

In Chapter 4 I investigated whether there was a ZEB1 isoform switch during HMLE-mesHMLE EMT. I induced an EMT in the HMLE cells using TGFβ and collected RNA at several time points. I determined the relative levels of each isoform using isoform specific qRT-PCR primers and found that these did not vary significantly across EMT.

In Chapter 5 I characterised a range of ZEB1 binding properties, confirmed the Z-box as the primary ZEB1 motif, identified the number of ZEB1 sites required for binding and the linker distance between these, and finally, identified and characterised novel ZEB1 targets. To achieve these goals I analysed publicly available ChIP-seq data. This provided insights for ZEB1 motif and binding patterns as well as providing a list of potential ZEB1 target genes. I tested 26 potential target genes using qRT-PCR, luciferase and ChIP-qPCR. Following this battery of tests I identified F11R and INADL as novel, *bona fide* ZEB1 targets.

In Chapter 6 I summarise my findings and describe how they are well positioned to contribute to the landscape of current research into epigenetic regulation of EMT.

Chapter 2 Materials and methods

In this section I outline the Materials and Methods used throughout my project, including molecular, cellular and bioinformatic techniques. Further, where specific methods are required to address particular questions, details are presented in subsequent chapters.

2.1 General cell culture

2.1.1 Cell culture

All cell lines were cultured in growth as media outlined in Table 2. Cells were grown to ~80% confluence prior to splitting using either TryplE or trypsin (Table 2). HMLE cells were passaged a maximum of 20 times before returning to fresh thaws as they can be quite divergent after this point (personal laboratory experience).

Table 2 Media, supplements and disassociation reagents used for cell culture.

Cell line	Media composition	Disassociation reagent
MDA-MB-231	10% FCS 90% DMEM	Trypsin
HEK293T	10% FCS 90% DMEM	Trypsin
MCF10A	95% DMEM/F12 5% HS 20ng/ml EGF 0.5µg/ml hydrocortisone 100ng/ml cholera toxin 10µg/ml insulin	Trypsin
MCF10A – starved	100% DMEM/F12	Trypsin
HMLE	HUMEC Supplement BPE	TrypLE
mesHMLE (Weinberg medium)	90% DMEM/F12 10% FCS 20ng/ml EGF 0.5µg/ml hydrocortisone 10µg/ml insulin	TrypLE

2.1.2 Freezing cells

Cells were

grown to 80% confluence trypsinised and counted then pelleted. Cells were resuspended at $2x10^6$ cells/ml in DMEM + 20% FCS. An equal volume of DMEM + 20% FCS + 20% DMSO was slowly added to the cells with gentle mixing. 1ml of cells were transferred to cryogenic tubes and stored in a Mr Frosty at -80ºC O/N. Cells were then stored in liquid nitrogen for up to 10 years.

2.1.3 Thawing cells

Cells were thawed in a 37ºC water bath then moved to a 10ml Falcon tube. DMEM + 10% FCS was added dropwise then cells were pelleted, resuspended and seeded in a T25 flask.

2.1.4 Plasmid and siRNA transfection

Cells were plated to be 70% confluent by next day in P/S free media and transfected as shown in Table 3.

Table 3 Composition of DNA and siRNA transfection for a 6-well plate.

DNA	**siRNA**
Dilute 6µl of LF2000 in 250ul OptiMEM	7ul of RNAiMAX in 250µl OptiMEM
Dilute 2.5µg of DNA in 250µl OptiMEM	30pmol of siRNA in 250µl OptiMEM

These transfection solutions were prepared in Eppendorf tubes and left for 5mins then combined and added dropwise to the wells. Cells were harvested 48-72h post transfection. Transfection for different size vessels was scaled accordingly.

2.1.5 Virus production

All virus work was carried out according to OHS and OGTR regulations. 293T cells were plated in a T25 flask and transfected with three viral vectors containing GAG/Pol (1.25µg),

REV4 (1.05µg) or VSV-G (1.125µg) and the lentiviral expression vector (2.1µg). The media was changed after 24hs and the viral supernatant (media) was harvested and combined after 48 and 72 hours and stored at either 4^{o}C O/N or -80^{o}C indefinitely.

2.1.6 Viral transduction

Cells were plated in a T25 flask and allowed to reach 75% confluence. Cells were then transduced with 1ml of viral supernatant and 4µg/ml of polybrene. The media was changed 48 hours later and antibiotics were added at the concentration determined from the kill curve.

2.1.7 Kill curve and stable selection

HMLE cells were plated at either 1 or 5x10^{6} in a 6 well plate while MCF10As were plated at 5x10^{6} in a 6 well plate and allowed to adhere overnight. Puromycin was added in a concentration range from 0-5µg/ml while blasticidin was tested at 0-20µg/ml with 6 concentrations for each antibiotic.

2.1.8 Incucyte proliferation assay

Cells were plated at 3x10^{3} in a 96 well plate in 100µl of media then placed in the IncuCyte for monitoring. Subsequent to the cells approaching 70% confluence the plates were removed and the images were analysed as described in the following section.

2.1.9 Incucyte analysis – proliferation assay

An image collection was made by adding four representative images at different time points from the same well. Using this I made a processing definition which acts to distinguish between cells and background thus identifying the cells and determining how the confluence

changed over time. Importantly, the same processing definition was used for each replicate to maintain reproducibility. Confluence was used to assess proliferation rate. Once exported to Excel this allows for a simple determination of rate of proliferation using the rate function. In addition to this, movies were created from representative wells to obtain insight into whether there were intrinsic differences in cell size contributing to the observed proliferation rate.

2.1.10 Incucyte scratch wound assay

Cells were diluted to $3x10^5$/ml and 100µl/well were plated in a 96 well plate and allowed to grow to confluence overnight. The following day the WoundMaker was used to scratch the plate according to the manufacturer's instructions. The cells were washed 1x in PBS before replacing media and placing into the IncuCyte for imaging.

2.1.11 Incucyte analysis – scratch wound assay

An image collection was made as described in 2.1.9. This allows for measurement of how the wound-width changes over time. Importantly, the same processing definition was used for each replicate to maintain reproducibility. Prior to analysis, images were visually assessed individually to determine whether the scratch was adequate for my needs (clean, consistent size) and those that met the criteria were analysed using the processing definition. The metric I chose to use was Wound Width. Once exported to Excel this allows for a simple determination of rate of wound closure using the rate function. In addition to this, movies were created from representative wells to obtain insight into changes in migration behaviour that would not be observed from the raw data.

2.2 Molecular techniques

2.2.1 Lysate preparation

Lysates were extracted from cell lines by first dissociating the cells using trypsin or TRYPLE followed by centrifugation. The cell pellet was resuspended in substrate trapping buffer + 1x Protease Inhibitor Cocktail (PIC) and sonicated in six 30 second boosts in an ice cooled waterbath. Subsequently this was centrifuged at 6000rpm for 5min at 4^{o}C and supernatant was obtained. A fraction of this (5µl) was removed for Bradford analysis. The remaining supernatant was diluted 5:1 with Laemelli buffer and heated at 95^{o}C for 5 minutes. 5x loading buffer was added and the lysate was stored at -20^{o}C or used immediately.

2.2.2 Bradford and Western blot

The previously separated lysate was diluted 1:40 with MQ-H2O and 20µl was pipetted in duplicate into a 96 well plate. A standard curve of 0-.175µg of BSA protein in .025µg increments was used for normalisation. 200µl of Bradford reagent diluted 1:4 with MQ-H2O was added to each well and the concentration was determined using a plate reader at 595nm.

Gels were poured as a separating followed by a stacking gel. The separating gel consisted of 4ml 4x lower gel buffer, 4ml of H2O, 2ml of polyacrylamide, 20µl of TEMED and 10µl of APS. This was poured in between two glass plates and 1ml of water was pipetted on top. The gel was allowed to set for ~20minutes then the water was poured off. Subsequently the stacking gel, consisting of 4ml of H2O, 2ml of upper gel buffer 20µl of TEMED and 40µl of APS, was poured on top and a 10well separator was wedged in. This gel was allowed to set for ~25 minutes and either kept at 4^{o}C O/N or used immediately. 25-50µg of sample or 10µl of Kaleidoscope ladder (Bio-Rad) were loaded in each well and separated at 120V for 1.5

hours or until the front had run off. The gel was transferred onto a nitrocellulose membrane at 250mA for 2 hours at 4^{o}C. The membrane was blocked in 5%SMP in TNT O/N at 4^{o}C for ZEB1 probing or for 1 hour at RT for all other antibodies. The membrane was probed with antibodies diluted as shown in Table 4 then washed 3x in TNT and probed with appropriate secondary antibodies as shown in Table 4 then washed 3x in TNT. The membrane was visualised using either the LAS-4000, Odyssey or ChemiDoc.

Table 4 Antibody details for Western Blot and ChIP where applicable

Antibody (Company, Catalogue number)	Species	Dilution for WB	Dilution for ChIP
αZEB1 H-102 (Santa Cruz Biotech, Sc25388)	Rabbit	1:200	1:200
αZEB1 E20x (Santa Cruz Biotech, Sc25388x)	Goat	1:1000	1:2000
αTubulin (Abcam, ab-7291)	Mouse	1:5000	
αFlag (Sigma-Aldrich, F1804)	Mouse	1:5000	
αV5 (Invitrogen, R96025)	Mouse	1:5000	
αH3 (Abcam, ab1791)	Rabbit		1:400
αH3K4me3 (Abcam, ab8580)	Rabbit		1:833
αH3K27me3 (Millipore, 07-449)	Rabbit		1:1000
αMouse HRP (Thermo Scientific 31430)	Goat	1:10000	
αRabbit HRP (Thermo Scientific 31460)	Goat	1:10000	
αGoat HRP (Sigma-Aldrich SAB3700244)	Rabbit	1:10000	

2.2.3 RNA isolation, cDNA synthesis and qRT-PCR

Total RNA was extracted from cell lines using Trizol (Invitrogen), according to the manufacturer's instructions. For mRNA analysis, complementary DNA (cDNA) was synthesized from 1.0μg of total RNA using the QuantiTect reverse transcription kit (Qiagen) according to the manufacturer's instructions with random primers. qRT-PCR was subsequently performed in triplicate with a 1:4 dilution of cDNA using the QuantiTect SyBr green PCR system (Qiagen) on a Rotorgene 6000 series PCR machine (Corbett Research). Data collected were analysed using the Rotorgene software. Relative expression levels were

determined using the Rotorgene quantitation software. Real time primers used are shown in Table 5.

Table 5 Primer sequences for qRT-PCR

Primer name	Sequence
ZEB1 isoform 6 F	GCTGGGAGGATGACATAAAAGA
ZEB1 isoform 6 R	GTCCTCTTCAGGTGCCTCAG
ZEB1 isoform 8 F	GAGGAGGTGACTCGAGCATT
ZEB1 isoform 8 R	GCCCTTCCTTTCCGTTATTG
ZEB1 isoform 2 F	CGGCGCAATAACGTTACAA
ZEB1 isoform 2 R	CTGTCATCCTCCCAGCAGTT
ANKS1A F	GCTTCAGGATCCAGGAGGAG
ANKS1A R	TTGCCAACTCTGCACTGGGG
AP3B1 F	GCCTCTTTAGCAGCGATTTG
AP3B1 R	ATCCGCTTCATAGCATCCAG
BAD F	GGATGAGTGACGAGTTTGTGG
BAD R	GATGTGGAGCGAAGGTCACT
BRCA1 F	CACCCAATTGTGGTTGTGC
BRCA1 R	CACTGTCCAACACCCACTCT
CRB3 F	GGGGCAAATACAGACCACTT
CRB3 R	GAGGGAGAAGACCACGATGA
CTNNB1 F	GCTTGGAATGAGACTGCTGA
CTNNB1 R	GCCATATCCACCAGAGTGAAA
DFFB F	CTCCATCAACCCCTACAGTAACA
DFFB R	AAATACTCCCAGTCCACTTCTCTT
EPCAM F	TGGAATTGTTGTGCTGGTTA
EPCAM R	GTGTCCATTTGCTATTTCCCTTC
ETV7 F	GGGGCAGAAACTCCTGTTC
ETV7 R	TGTGGGTCCTCCAACTTTGT
F11R F	TATAGCCGAGGCCACTTTGA
F11R R	TGCAGATGATAGGCGGTGAG
FLNA F	GTAGGCCAGAAGAGCAGCTT
FLNA R	ACGTGCTTCACCAGGATCTC
GAB2 F	GTTGATTATCTGGCCCTGGA
GAB2 R	TGGGTCTTCTCCTTGTCCAC
GSK3A F	GCCATTCTCATCCCTCCTC
GSK3A R	GTGAGGGTAGGTGTGGCATC
HIF1A F	TGGAAACGTGTAAAAGGATGC
HIF1A R	TGGTCAGCTGTGGTAATCCA
INADL F	CTGCTGAACACCATCCAGAA
INADL R	AACACCTTCCAGGGTCTCG
KRAS F	TATGGAATTCCTTTTATTGAAACATC
KRAS R	TACACACTTTGTCTTTGACTTCTTTTT
LLGL2 F	AACTGGCGTTCACATCGAG
LLGL2 R	CCCGAAAGGCCATCAGTAGT
MAK F	AATAAACCCCCACACTTGGA

MAK R	CCCGACCAGTTTTTGTGTTC
POLR1D F	GCAAAACGTGGGAAAACTAGA
POLR1D R	GCCCTCTTTTTGTTCATGGT
PTEN F	AGCGTGCAGATAATGACAAGG
PTEN R	GATTTGACGGCTCCTCTACTG
SPINT1 F	ACTTTGAGGAAGAGCAGCAG
SPINT1 R	CCACCACCACAATGCAGAT
SYK F	GCAGGGTGTCCAAGAGAGAT
SYK R	AGCGGTTAGTTCACCACGTC
TAF3 F	GAGATGAGTGGGGCAATCAG
TAF3 R	TGGTACCAGTCATCGCAGTC
WTAP F	GCAGAGTTGGCTTTACAGAAGAA
WTAP R	AACTGCTGGCGTGTCTCC
GGA1 F	CTCAGCAATCACCCAGGTC
GGA1 R	ACATCCCCCATCTCGTTGTA

2.2.4 PCR

PCRs were initially performed using Phusion enzyme according to the manufactures instructions with a TM two degrees lower than that provided by Primer3. If this resulted in no product a lower temperature was tested. If it resulted in too many products the temperature was raised. PCR primers were primarily used for cloning (Table 6). This, depending on the purpose, required the addition of: restriction enzyme sites, gateway recombination sites and epitope tags.

Table 6 Primer sequences used for luciferase cloning.

Primer Name	Sequence
CRB3 F	ATGGTACCATCTCGGCTCACTGCAATCT
CRB3 R	ATAAAGCTTGCCCACAGGAGAGACAACTG
F11R F	ATGGTACCAAAAAGGTGGCTTAGTCAGCA
F11R R	ATAAAGCTTGACTACAGCGAGGGGACTGA
INADL F	ATGGTACCAAAATGCCTTGCCTGTGTTC
INADL R	ATAAAGCTTATCTGGGCGCAAAATAGGAT

2.2.5 Restriction enzyme based cloning

Plasmid DNA was digested using the following reaction mixture (Table 7) for both cloning and diagnostic purposes and incubated at 37°C for 60 mins.

Table 7 Reaction mixture for restriction digest.

Component	Quantity
DNA Plasmid	1 µg
10x buffer	2 µL
BSA (If needed)	2 µL
Enzyme (each)	.5 µL (5 units)
MQ H2O if BSA used	14.5 µL
MQ H2O if no BSA used	12.5 µL
Total	20 µL

If more than one enzyme was used the total H20 added was adjusted accordingly.

2.2.6 Ligation

Ligation reactions were performed using 4µl of DNA insert (from digest), 0.5µl of DNA vector and appropriate ligation buffer and T4 DNA ligase enzyme.

Reaction mixtures contain a ratio of approximately 3:1 insert to vector and 1 unit of T4 DNA ligase. Vectors were prepared by digestion with appropriate restriction enzymes. Reaction was either left overnight at 4°C or for 4 hours at room temperature. Negative control ligations were used for each different vector in the ligation, replacing the DNA insert volume with H_2O.

2.2.7 DNA Gel Electrophoresis

DNA gel electrophoresis was carried out on a 1-2% agarose gel containing EtBr or GelRed. The DNA samples were prepared with 4xDNA loading buffer and run at 100V in TBE. Gel was visualised using a UV trans-illuminator.

2.2.8 Gel Extraction

Following DNA gel electrophoresis fragments were excised and purified using the Qiagen Gel Extraction kit following the manufacturer's instructions

2.2.9 Gateway cloning

All gateway cloning reactions were carried out using the following reagents (Table 8 and Table 9).

Table 8 BP reaction components

Component	Quantity
PCR product	5µl
Entry vector	150ng
1x TE buffer	8µl - vectors
Enzyme	2µl
Total	10µl

Table 9 LR reaction components.

Component	Quantity
Entry vector	150ng
Destination vector	150ng
Buffer	8µl- vectors
Enzyme	2 µl
Total	10µl

Enzyme was added last after thawing on ice for 30 seconds. Reaction mixtures were incubated at RT for 1 hour, 1µL of Proteinase K was added and incubated at 37ºC for 10 mins. Reaction mixtures were transformed and subsequently plated on LB agar plates containing the appropriate antibiotic selection.

2.2.10 Bacterial transformation

DH5α competent cells were thawed on ice. ~10ng of purified plasmid, or 5µl of a ligation reaction, was added to 50 µL of competent cells and incubated on ice for 20 mins, heat-shocked at 42ºC for 2 mins and cooled on ice for 10 mins. 950 µL of LB was added and the bacteria were incubated at 37ºC with shaking for 45 mins. This was then centrifuged at 1,000rcf for 5m then media was removed, leaving ~150µl. Bacteria were resuspended in this smaller volume and plated on LB Agar plates containing the appropriate antibiotic selection and incubated overnight at 37ºC.

2.2.11 Luciferase Assay

Cells were plated to reach 70% confluence the next day in 24-well plates. Firefly and RL vectors were co-transfected alongside the experimental variable (siRNA or plasmid transfection). After 48 hours the cells were harvested

2.2.12 DNA crosslinking

Filter tips were used throughout to avoid contamination.

Cells were grown in multiple T75 flasks (typically eight) and processed in groups of four. Cells were dissociated and centrifuged then adjusted to $1x10^6$ cells/ml in DMEM (high glucose) + 10% FCS and 6mls were aliquoted to 10ml tubes. 162µl of 37% formaldehyde (1% final concentration) was added and incubated at room temperature for 5m with gentle inversion then centrifuged at 4ºC for 5m at 300rcf. The remaining steps were performed on ice. Media was removed leaving a small quantity (~50µl) behind and resuspended through

gentle inversion in 6ml of HBSS + PIC. The cells were centrifuged as before, and HBSS removed, then snap frozen in liquid nitrogen and stored at -80^{o}C.

2.2.13 ChIP-qPCR assays

Filter tips were used throughout to avoid contamination. 1x10^{6} crosslinked cells were used in each ChIP assay. Sonicator and centrifuge were cooled prior to thawing the cells.

Previously frozen aliquots of cross linked cells were thawed on ice. 300µl of lysis buffer and 3µl of 100x PIC was added to each tube, inverted, and incubated on ice for 10 min. 900µl of HBBS + 9µl PIC was added to each tube and flick-mixed and 215µl was added to low bind 1.5ml tubes (Axygen). Cells were sonicated for 30secs at a variable amount of cycles as empirically determined to yield chromatin fragments ranging from 200-500bp typically with ice renewed every 5 cycles. Cell debris was pelleted by centrifuging at 13,000rcf for 5min in a cooled centrifuge. Supernatant was combined in a 15ml tube and diluted two fold with 2x RIPA buffer and 100x PIC. 1/10 volume (40µl) was separated for input. The remaining volume was aliquoted into 400µl aliquots on ice.

Antibodies were added at appropriate amount (Table 4). Protein G Dynabeads (Invitrogen) were washed twice with 200µl of RIPA buffer and resuspended in the original volume. Immuniprecipitations were performed for 2h at 4oC with rotation. The antibody:protein:DNA complexes were then collected with empirically determined amounts of beads for 2h of rotation at 4^{o}C. The beads were washed at 4oC thrice using 200µl of RIPA buffer and once with TE buffer, then incubated with 200µl of elution buffer + Proteinase K for 2h in a thermomixer (1300rpm, 68oC to reverse the protein:DNA crosslinks. After incubation the eluates were collected. Genomic DNA was recovered by using phenol chloroform extraction

and ethanol precipitation. Pellets were washed in 70% ethanol, briefly air dried and resuspended in T10E buffer. Quantitation of ChIP DNA was determined through duplicate qRT-PCRs with gene specific primers (Table 10). Enrichment was expressed as % input normalised to control promoter regions. % input was calculated using the formula

$$\% \left(\frac{ChIP}{Input}\right) = 2^{\,Ct(ChIP)-Ct(Input)\ x\ Input\ dilution\ Factor} x\ 100\%$$

Described at http://sabiosciences.com/chipqpcr to account for chromatin sample preparation.

Table 10 Primer sequences for ChIP-qRT-PCR

Primer name	Sequence
F11R F	GGC GGT CCT ATG ACG GTA GT
F11R R	CTC CTC CTC CGG ACA ACC T
INADL F	ACCCTGGGCCTCTCACTTC
INADL R	TTTACCTGGTGCGCTGGAG
HPRT1 F	TGAGAGTTCAAGTTGAGTTTGGA
HPRT1 R	TGATAATTTTACTGGCGATGTCA
INADL -327bp from ZEB1 summit	GTTAAGCGGATTGACCCAACG
INADL -327bp from ZEB1 summit	GTGTAGAGCGAGACCCAAGAT
INADL -655bp from ZEB1 summit	AGTGGAAGAAGGCTATCGGGAC
INADL -655bp from ZEB1 summit	AACACAGGCAAGGCATTTTGAC
INADL -926bp from ZEB1 summit	CCCCGCCTGAAAGTCTTTATGT
INADL -926bp from ZEB1 summit	AAAATGGTATCCCTGTCCCCAG
INADL +254bp from ZEB1 summit	GGCGAGGATGGGGCG
INADL +254bp from ZEB1 summit	CCGACGCAGGAGGACG
INADL +636bp from ZEB1 summit	CTGGTCATCCCGTAGCGGAG
INADL +636bp from ZEB1 summit	ACTTTCCGACAGAACCTCCC
INADL +975bp from ZEB1 summit	GACTGCATTTAGGGTCTGGCT
INADL +975bp from ZEB1 summit	CATCCTCAGTGGAACCGACAT
F11R +272bp from ZEB1 summit	TCATATTGGCGATCCTGTTGTG
F11R +272bp from ZEB1 summit	CAACCTCTGACCCGACCAAC
F11R +608bp from ZEB1 summit	CCACACTCATCTACCCAGCG
F11R +608bp from ZEB1 summit	CGGGGTTCGTGGATGCTG
F11R+910bp from ZEB1 summit	AGGTTCCAGCGAAGTTCCTTT
F11R+910bp from ZEB1 summit	TGGGAAGAGTCACTCAATCAACA
F11R -364bp from ZEB1 summit	GCTCATCTCTCAAAGCCGGT
F11R -364bp from ZEB1 summit	GGGAGAAGACCCGGAGAGAA
F11R -605bp from ZEB1 summit	ACGTAATTCTGCTCTTCCCAA
F11R -605bp from ZEB1 summit	CCACTGAACTACAGCCTGGG
F11R -894bp from ZEB1 summit	TAGCAAAGGTGTTGAGCATGAAT
F11R -894bp from ZEB1 summit	TGTGCTGACTAAGCCACCTTTT

2.2.14 Bioinformatic analysis

I performed a variety of bioinformatic analyses using different tools throughout my PhD, these were, ChIP-seq analysis, motif discovery, NCI-60 expression analysis, sequence searches, primer design, gene ontology and diagnostic restriction digests and plasmid maps.

The ChIP-seq analysis was performed with the help of Katherine Pillman. In brief, we obtained ZEB1 ChIP-seq files for GM12878 and HEPG2 as FASTA files from ENCODE (http://genome.ucsc.edu/cgi-bin/hgFileUi?db=hg19&g=wgEncodeHaibTfbs). We used FastQC to look at quality and read numbers and to trim adapters and remove reads that did not meet the quality or length requirements. We then mapped the remaining reads using fastq. As the final QC, we removed the duplicates using Picard (https://github.com/broadinstitute/picard), this is done to remove potential PCR bias. Most of the runs had <10% duplicates while both GM12878 ZEB1 ChIP-seq runs had ~40% duplication. We used MACS [86] to identify transcription factor binding sites and used Python to link the peaks to gene sites with the prospective peaks falling within -2.5 to .5kb from the TSS.

The NCI-60 expression analysis relies on the use of publically available microarray expression data across 60 different cell lines, this is explained in more detail in Section 3.2.1. Sequence searches were carried out using ENSEMBL (link) to identify the location of exon-exon junction in ZEB1 for the design of isoform specific primers and to locate the promoter region of potential ZEB1 target genes for luciferase cloning. Primer design in the first two years of my PhD involved a combination of Primer 3 to design primers, Oligo Calc to exclude primers with potential hairpins and cross binding and nBLAST to ensure that the primers would specifically amplify my target. In the second two years of my PhD this was simplified to using the NCBI primer design toolset which combines the features of the other three applications. Motif discovery was carried out on the analysed ChIP-seq data using MEME-ChIP. The gene ontology analysis used the Gene Ontology Project software [87, 88] to determine what pathways were enriched in bivalent genes becoming activated in my EMT

model. Diagnostic restriction enzyme tests were chosen using A Plasmid Editor (APE) as were plasmid maps for post Gateway recombination.

2.2.15 Statistical Analysis

For tests between two conditions unpaired student t-tests were used. For tests with multiple variables significance was first determined through ANOVA and Fisher's PLSD was used as the post-hoc test. Significance was determined at the 0.05 level.

Chapter 3 Identification and characterisation of novel bivalent genes in an *in vitro* model of EMT

3.1.1 Chapter outline

In this Chapter I describe my work on identifying and characterising novel bivalent genes in the HMLE-mesHMLE EMT model. As previously discussed (Chapter 1), I was interested in exploring the potential role for bivalent genes in a malignant EMT model due to their importance in developmental EMT and the potential for using these gene lists to identify novel genes associated with this process. In this Chapter, I first explain the HMLE-mesHMLE EMT model, and the ChIP-seq experiments performed on these cell lines. I then outline what genes I chose for future investigation, why, and their effect on EMT markers. Finally, I assess their biological significance by quantifying their effect on proliferation and migration.

3.1.2 HMLE cell line model

As mentioned in Chapter 1, I wanted to examine the relationship between bivalent genes and EMT due to their importance in development and preliminary evidence on their role as regulators of EMT. I used a HMLE EMT model to evaluate this relationship. This is an easily manipulated model which has been used extensively in Prof. Greg Goodall's laboratory [89, 90]. HMLE cells are human mammary epithelial cells which have been immortalised through the addition of hTERT and large T antigen [91]. These cells are largely epithelial in nature and cannot form tumours in nude mice or form colonies in soft agar [91]. Treatment with TGFβ induces an early crisis stage where many of the cells die. However, this is followed by an EMT, and the resulting cells are termed mesenchymal HMLE (mesHMLE). This EMT is

associated with large changes in mRNA and protein expression and morphology (Figure 2). My goal was to use this model to identify genes that went from a bivalent epigenetic state to an active state and had a concomitant change in RNA expression (1.3.2). To do this, Dr Joanne Attema (Goodall Laboratory) interrogated HMLE and mesHMLE cells using ChIP-seq against H3K4me3 and H3K27me3 alongside RNA-seq. David Lawrence (Goodall Laboratory) performed the bioinfomatic analysis to assign the ChIP peaks and call genes. Using this information I searched for, and identified, hundreds of genes that underwent a bivalent to active epigenetic transformation in the HMLE EMT. I discuss the data set in the following section and the role of selected bivalent genes throughout the rest of the thesis.

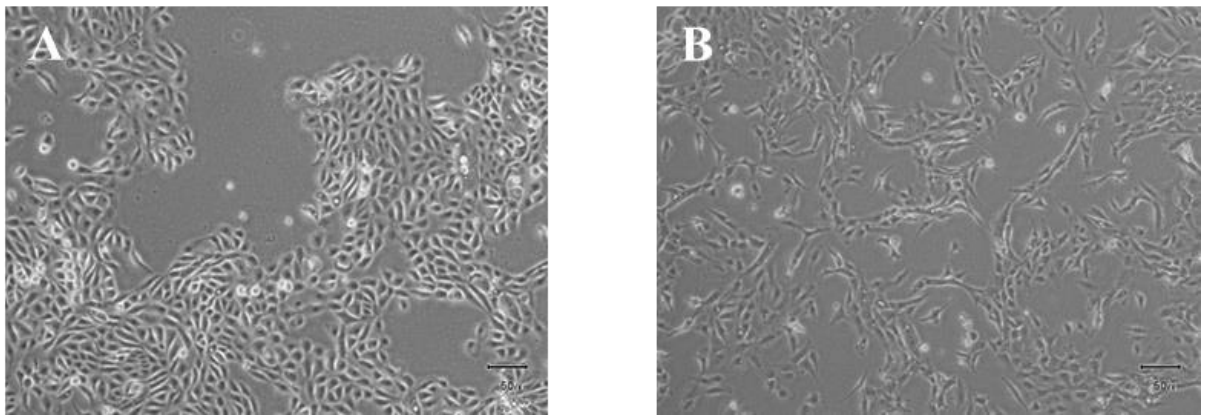


Figure 2 Morphological changes in HMLE versus mesHMLE cells

(A) HMLE cells were grown in HUMEC media and photographed at 40x magnification. (B) mesHMLE cells were grown in Weinberg media and photographed at 40x magnification.

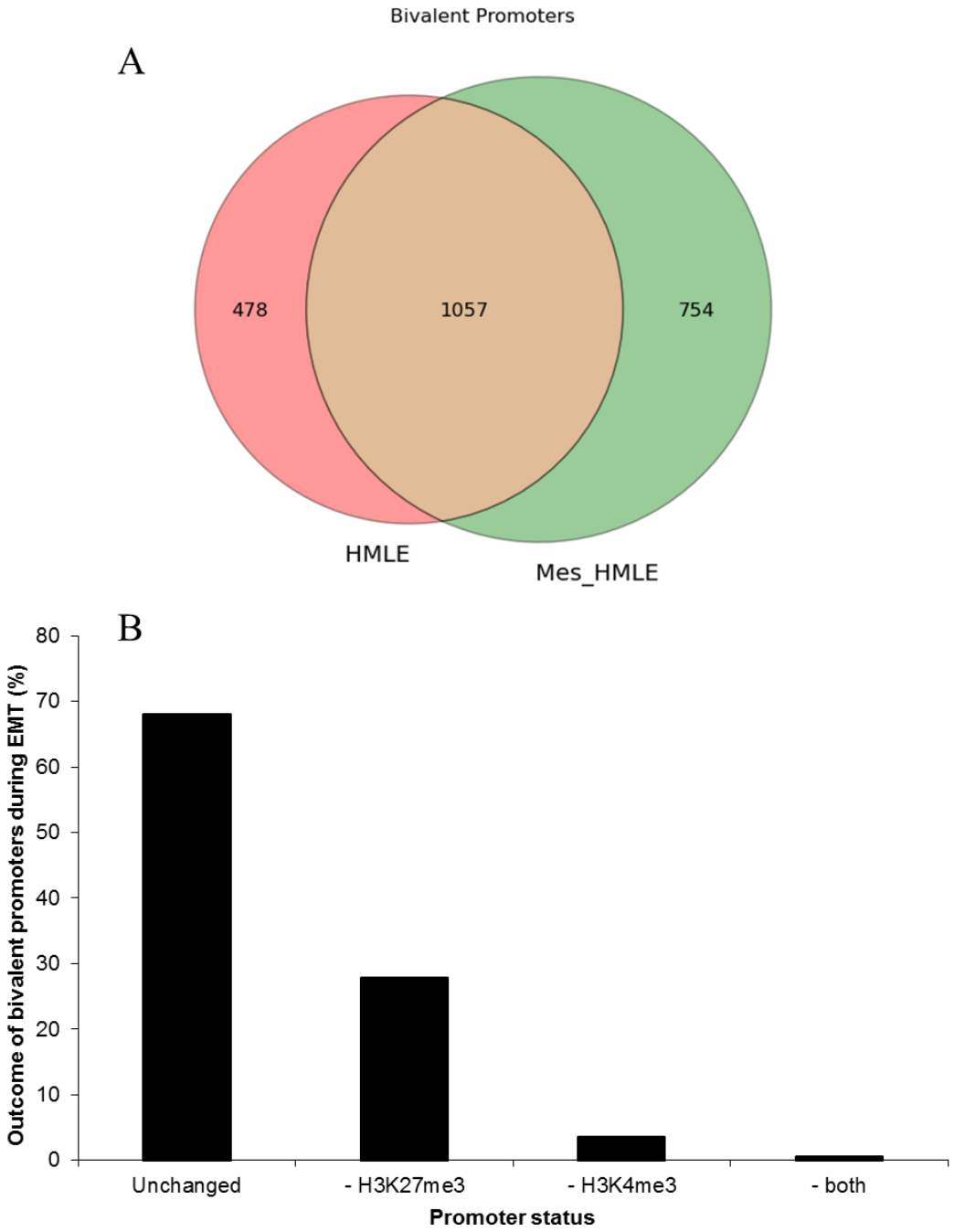


Figure 3 Epigenetic changes in bivalent promoters after EMT in HMLE cells treated with TGFβ for 18 days.

(A) Venn diagram of bivalent promoters in HMLE and mesHMLE cells. (B) Outcome of bivalent promoters during HMLE-mesHMLE EMT.

3.1.3 Analysis of HMLE-mesHMLE ChIP-seq data

From the ChIP-seq results we found 1535 bivalent genes in HMLE cells and 1811 bivalent genes in mesHMLE cells (Figure 3 A, Supplementary Table 1). Of the bivalent HMLE genes, 1057 (68%) remained unchanged, 429 (28%) lost H3K27me3 and 49 (4%) lost either H3K4me3 or both marks (Figure 3 B). I was interested in the 429 genes that lost only the H3K27me3 mark as they represent the genes that went from a bivalent to active state.

Before examining the list of 429 genes I had to determine whether the change in epigenetic marks was accompanied by a change in mRNA expression. This is particularly important as it has been shown that loss of H3K27me3 is often associated with an increase in DNA methylation. This DNA methylation can lead to the maintenance of repression independently of the histone epigenetic marks [81]. To confirm that genes changing from a bivalent to active chromatin state were also being actively transcribed I cross referenced the ChIP-seq data with our RNA-seq data. I found that 300 (~70%) were strongly upregulated at the RNA level (Supp. Table 1). This shows that, in most cases, the change from bivalent to active epigenetic marks was accompanied by an upregulation in mRNA as expected. Due to the lack of transcriptional activation for the remaining 129 genes I focussed on the group of 300 upregulated genes as they are more likely to play a functional role. This group included the established EMT genes TGFβR3, TWIST and ZEB1, but not the other master EMT TFs SNAIL, SLUG or ZEB2. This implicates the two transcription factors, ZEB1 and TWIST, as early responders in our EMT model (Table 11). Based on this, and previous interest in ZEB1 due to its interaction with the miR-200 family, I decided to look at ZEB1 in more detail and to identify novel ZEB1 targets (Chapters 4 and 5). As well as this group of genes that went from bivalent to active during EMT I observed that 394 genes underwent the opposite epigenetic transformation including CDH1 and ESRP1, two key epithelial genes (Table 11).

Following from the previous reasoning that bivalent genes are important for the rapid transformation of cells, this suggests that, as CDH1 and ESRP1 become bivalent genes, they are important for future transition of cells in our *in vitro* model. It would be interesting to identify whether they undergo similar epigenetic regulation *in vivo*. Since it is known that CDH1 and ESRP1 are essential epithelial genes [92, 93], the transition in question is likely to be an MET.

Table 11 Summary of gene expression and epigenetic state of three candidate genes, ADM2, PLEKHO1 and RASA3, two mesenchymal associated genes, ZEB1 and TWIST1, and two epithelial associated genes, CDH1 and ESRP1.

Expression was measured as area under the peak. For fold change in expression or epigenetic marks, green represents a positive, and red represents a negative, fold change between HMLE and mesHMLE cells.

Gene	HMLE expression	MesHMLE expression	Fold change in expression (log2)	q_value	H3K4me3	H3K27me3	HMLE bivalent status	MesHMLE bivalent status
ADM2	0.1	1.8	4.09	1	0.62	-1.14	bivalent	active only
PLEKHO1	3.4	10.1	1.57	0.35	1.09	-3.74	bivalent	active only
RASA3	1.5	6.0	2.00	0.08	1.85	-4.38	bivalent	active only
ZEB1	0.2	3.0	3.79	1	0.61	-4.39	bivalent	active only
TWIST1	1.9	24.6	3.68	0.0001	0.95	-4.45	bivalent	active only
CDH1	63.4	0.1	-9.82	0.0001	-3.60	4.03	active only	bivalent
ESRP1	28.5	0.0	-13.00	0.000263	-4.68	4.52	active only	bivalent

To determine whether there were any over represented networks in the group of bivalent-active genes I used gene ontology analysis. Gene ontology uses evidence-supported annotations to describe the biological roles of individual genomic products (e.g. mRNAs, proteins, ncRNAs and complexes) by classifying them according to molecular function, location and biological processes [94]. This is primarily used to interrogate large data sets to identify whether any particular aspect is enriched. I carried out gene ontology analysis using the Gene Ontology Project software [87, 88]. From this I identified that, for the bivalent to active gene list, the five most enriched biological processes included cell migration and cell morphogenesis (Table 12). In contrast, for bivalent genes that were not activated, the majority of enriched biological processes were associated with organ development (Table 12). Also, while cell migration was in the top five biological processes for bivalent to active genes, it was not present in the 200 most enriched processes for all bivalent genes. This highlights two things. Firstly, as expected, many bivalent genes were associated with developmental functions. Secondly, and most importantly, a subset of these bivalent genes, with EMT properties, were selectively activated during EMT in the HMLE-mesHMLE EMT model. In addition to this, I observed the very interesting inclusion of MET as the second most over represented biological process when looking at all bivalent genes in HMLE cells. This suggests that after an MET, genes that are important to this process will not maintain expression but will rather return to a bivalently repressed state.

Table 12 Top 5 biological processes, as determined by Gene Ontology analysis, for bivalent genes identified in HMLE cells separated based on their transition to an active or inactive state.

GO biological process	Homo sapiens REFLIST (20814)	Number of hits	Total genes expected	Fold Enrichment	P-value
All bivalent genes (1331 genes)					
renal vesicle morphogenesis (GO:0072077)	13	8	0.83	9.62	2.14E-02
mesenchymal to epithelial transition (GO:0060231)	15	9	0.96	9.38	6.31E-03
renal vesicle development (GO:0072087)	14	8	0.9	8.94	3.66E-02
glomerulus development (GO:0032835)	49	20	3.13	6.38	1.27E-06
metanephros morphogenesis (GO:0003338)	28	11	1.79	6.14	2.32E-02
Bivalent to not active (990 genes)					
type B pancreatic cell differentiation (GO:0003309)	18	8	0.86	9.34	2.64E-02
enteroendocrine cell differentiation (GO:0035883)	19	8	0.9	8.85	3.90E-02
endochondral bone morphogenesis (GO:0060350)	46	14	2.19	6.4	6.52E-04
glandular epithelial cell differentiation (GO:0002067)	40	12	1.9	6.31	6.27E-03
glomerulus development (GO:0032835)	49	14	2.33	6.01	1.39E-03
Bivalent to Active (341 genes)					
regulation of cell morphogenesis involved in differentiation (GO:0010769)	337	20	5.52	3.62	9.13E-03
embryonic organ development (GO:0048568)	410	21	6.72	3.13	4.75E-02
negative regulation of cell differentiation (GO:0045596)	609	29	9.98	2.91	3.42E-03
regulation of cell morphogenesis (GO:0022604)	529	25	8.67	2.88	1.52E-02
cell migration (GO:0016477)	759	32	12.43	2.57	1.15E-02

To investigate the potential for novel specific role players among the 300 bivalently regulated genes, I chose three genes, ADM2, PLEKHO1 and RASA3 (Table 11), based on two criteria: genes with known roles in EMT-related properties, such as cytoskeleton regulation, migration and proliferation and no previous association with EMT. These criteria were chosen to maximise the possibility of identifying novel pathways or role players associated with established EMT-mechanisms. In the next section I provide a brief characterisation of these three genes that further justifies their study in the context of this thesis.

3.1.4 ADM2, PLEKHO1 and RASA3

ADM2 is a secreted hormone and a member of the intermedin family [95]. I chose to study ADM2 due to its links with motility and invasion in trophoblasts during pregnancy [96-98]. While this system is far removed from my focus on breast cancer EMT, it features two key EMT properties of migration and invasion. ADM2 also has multiple links with angiogenesis, primarily in relation to cardiovascular disease through activating ERK and AKT signalling [99-102]. Angiogenesis is another key property of tumours as they require additional blood flow to support continued growth. While studying its role in angiogenesis and these signalling pathways was outside the scope of my research, ADM2's role in EGF signalling presents an interesting link with PLEKHO1 discussed in 3.2.8 and 3.2.9. During my project, additional studies on ADM2 were published outlining its role as a prognostic biomarker in breast cancer [103] and pancreatic adenocarcinoma [104] with high levels of ADM2 in plasma or tissue correlating with poor survival. This highlighted the potential role for ADM2 in cancer progression and increases the significance of my results.

PLEKHO1 is a scaffolding protein with roles in EGF, insulin, TGFβ and MCSF signalling. It has multiple modes of action depending on the particular stimuli and system. As a

consequence of this it has a controversial role in cancer where it can act as either an oncogene or a tumour suppressor. Resolving this dichotomy for my system was part of my interest in PLEKHO1. During my research, a paper that showed that PLEKHO1 plays a role in prostate cancer metastasis downstream of EGF signalling was published [105]. In short, they showed that: EGF activation results in PIP3 formation, PLEKHO1 binds to PIP3 and brings CK2α to the membrane. CK2α phosphorylates and activates PAK1. PAK1 is responsible for many of the migration and proliferation outputs of EGF signalling [105]. I explore the potential for this mechanism in my system in 3.2.7.

RASA3 is associated with RAS signalling, which is a key process that is often deregulated in cancer and EMT [106, 107]. There are three human *RAS* genes, Kirsten rat sarcoma (KRAS), neuroblastoma rat sarcoma (NRAS) and Harvey sarcoma (HRAS) [108]. These three are highly similar and are generally referred to collectively as RAS signalling. Despite this, there are differences between the three with KRAS mutations observed at a higher rate in cancer. RAS proteins act as intermediate switches in several signalling pathways including EGF signalling, which, as previously mentioned, is of particular interest to my research.

RAS can exist in active and inactive forms. It alternates between these forms depending on whether it is binding GDP or GTP. When binding GDP it is in an inactive state and this is the default conformation. Upon stimulation, the inactivating GDP is exchanged for an activating GTP group; this process is mediated by GEFs [109]. Active RAS is naturally inactivated through hydrolysis of GTP into GDP; this process is accelerated through GAPs, including RASA3. RAS is commonly mutated in tumours to a constitutively active form in cancer. This implicates RAS GAPs as tumour suppressors [110].

Despite belonging to this group of tumour suppressors, RASA3 underwent the same epigenetic and expression changes as ZEB1, ADM2 and PLEKHO1, becoming actively transcribed after an EMT. This apparently contradictory group of characteristics, and its potential involvement with PI3K signalling, made RASA3 of interest as a potential brake on EMT which I briefly investigated.

These three genes have different hypothesised functions. The common thread between them is their potential role in EMT. As I performed the same, or similar, assays for each gene I primarily discuss them as a group with a focus on their relevance on EMT in the following sections.

3.2 Results

3.2.1 ADM2, PLEKHO1 and RASA3 are predominantly expressed in mesenchymal cell lines

The expression data in 3.1.3 showed that ADM2, PLEKHO1 and RASA3 were upregulated at the RNA level in the HMLE-mesHMLE EMT model. To further understand their expression pattern in a variety of cell lines I used two resources: *in silico* analysis of the NCI-60 dataset and *in vitro* analysis of 9 breast cell lines.

The *in silico* analysis used the NCI-60 dataset which consists of microarray data across 60 cancer cell lines [111]; the data was normalised using Z-scores defined as

$$Zscore = \frac{\text{expression in cell line} - \text{mean expression in all cell lines}}{\text{standard deviation of expression in all cell lines}}$$

This normalised for the variability in expression levels across different genes. The Z-score represents how many standard deviations away from the mean a given cell line is. Henceforth, a positive Z-score indicates a gene was over-expressed in a given cell line in relation to the population of lines, and a negative Z-score indicates relative under-expression. I interrogated the data set for CDH1, ZEB1 and VIM expression levels and defined three co-occurring criteria to segregate the cell lines into epithelial or mesenchymal subsets: positive Z-score for CDH1 and negative for ZEB1 and VIM were taken to indicate epithelial and the reverse for mesenchymal cells. Out of the 60 cell lines, 13 met this criteria for an epithelial cell line and 23 met these criteria for a mesenchymal cell line (Table 13). The remaining 24 cell lines did not meet these biologically strong criteria. Using the subset in Table 13, I examined the expression of ADM2, PLEKHO1 and RASA3 in epithelial and mesenchymal cell lines.

Table 13. Z-scores for the epithelial marker CDH1 and the two mesenchymal markers, VIM and ZEB1.

Cells were classified as epithelial if Z-scores were positive for CDH1 and negative for VIM and ZEB1 and the reverse to be classified as mesenchymal. Z-scores are shown for ADM2, PLEKHO1 and RASA3. This table shows the 36 cell lines from the NCI-60 data set that met these criteria. Red indicates a positive Z-score and green represents a negative Z-score.

Classification	Cell Line	Marker genes			Target genes		
		CDH1	VIM	ZEB1	ADM2	PLEKHO1	RASA3
Epithelial	BR:MCF7	2.12	-1.88	-1.25	0	0.42	-0.62
	BR:T47D	2.12	-1.97	-1.29	0.55	-1.07	-1
	CO:COLO205	1.20	-1.45	-0.93	2.36	-1.36	-0.21
	CO:HCC_2998	1.46	-2.01	-1.15	-0.19	-1.66	-0.91
	CO:HCT_116	0.44	-1.69	-0.37	0.29	-0.53	-0.32
	CO:HCT_15	1.89	-2.07	-1.34	0.46	-0.49	-1.64
	CO:HT29	1.65	-2.01	-1.24	0.05	-1.69	0.45
	CO:KM12	1.09	-1.20	-0.98	2.02	-0.99	-0.61
	CO:SW_620	0.81	-0.49	-0.01	1.95	-0.05	1.41
	LC:EKVX	0.98	-0.26	-0.05	-0.2	-0.5	-0.7
	LC:NCI_H322M	1.90	-1.87	-1.17	-0.04	-1.4	-1.23
	OV:OVCAR_3	0.67	-1.61	-1.33	0.35	-0.14	-1.21
	OV:OVCAR_4	1.16	-1.1	-1.27	0.99	0.87	-1.12
Mesenchymal	BR:MDA_MB_231	-0.64	0.43	1.08	0.27	0.32	0.81
	BR:HS578T	-0.49	1.08	1.91	2.47	1.01	0.4
	BR:BT_549	-0.48	0.67	1.39	0.92	0.98	0.74
	CNS:SF_268	-0.66	0.66	0.25	-0.48	0.97	1.73
	CNS:SF_295	-0.67	0.61	0.58	-0.88	0.33	0.29
	CNS:SF_539	-0.85	0.63	1.51	n/a	2.2	1.39

CNS:SNB_75	-0.54	0.60	0.35	-0.57	0.47	1.1
LE:SR	-0.82	0.30	1.07	-1.6	-0.86	-1.61
ME:LOXIMVI	-0.73	0.65	0.83	-1.27	-0.11	0.65
LC:HOP_62	-0.41	0.41	0.13	0.36	0.8	0.33
LC:HOP_92	-0.64	0.92	1.10	-0.49	0.19	0.61
LC:NCI_H226	-0.6	0.68	0.19	0.38	0.03	0.02
LC:NCI_H23	-0.62	0.29	0.55	-0.55	-0.65	-0.43
LC:NCI_H460	-0.37	0.06	0.23	0.11	-0.63	-0.29
OV:OVCAR_8	-0.52	0.35	0.34	-0.2	-0.77	0.28
OV:SK_OV_3	-0.06	0.51	0.81	-0.87	0.32	-0.64
PR:PC_3	-0.11	0.39	0.29	-0.01	-0.36	0.14
RE:786_0	-0.75	0.93	0.94	-0.68	0.27	-0.84
RE:A498	-0.53	0.56	0.09	1.01	-0.92	-0.03
RE:RXF_393	-0.72	1.00	1.11	1.2	1.15	-0.17
RE:SN12C	-0.64	0.25	0.96	-0.81	-0.25	0.03
RE:TK_10	-0.69	1.12	0.04	0.72	0.06	-0.56
RE:UO_31	-0.65	0.46	0.39	-1.15	-0.18	0.29

ADM2 shifted from a positive Z-score in 77% of epithelial lines to 40% of mesenchymal lines with a change in average Z-score from 0.67 to -0.09 (Table 13). PLEKHO1 shifted from being expressed in 15% of epithelial lines to 60% of mesenchymal lines with an average Z-score shift from -0.67 in epithelial lines to 0.21 in mesenchymal lines (Table 13). RASA3 shifted from being expressed in 15% of epithelial lines to 65% of mesenchymal lines with an average Z-score shift from -0.60 to 0.18. To provide some reference for these Z-score changes, CDH1 went from an average Z-score of 1.34 in epithelial lines to -0.57 in mesenchymal lines while ZEB1 went from an average Z-score of -0.95 in epithelial lines to 0.70 in mesenchymal lines. PLEKHO1 and RASA3 undergo similar, although more modest, changes to Z-values compared to the benchmarks of CDH1 and ZEB1. While ADM2, unexpectedly, displays the opposite changes although in a more modest fashion.

To obtain additional insights I then explored the statistical associations between marker genes, and between marker and target genes (Table 14 and Figure 4). Figure 4 A and Table 14 show that the relations between CDH1 and VIM were very close to the y = -x line (i.e. the fitted slope close to 1). This means a close quantitative correspondence whereby the reduction (or increase) in the expression of CDH1 is perfectly matched by the increase (or reduction) in the expression of VIM. The relationship between the expression of ZEB1 and VIM was close to the y = x line, although scattered, suggesting a commensurate up- and down-regulation of these two genes (Figure 4 B). PLEKHO1 and RASA3 had a positive relationship with both ZEB1 and VIM and a negative relationship with CDH1 (Figure 4 C and Table 14). ADM2 did not display a significant relationship with any of the marker genes (Figure 4 D and Table 14)

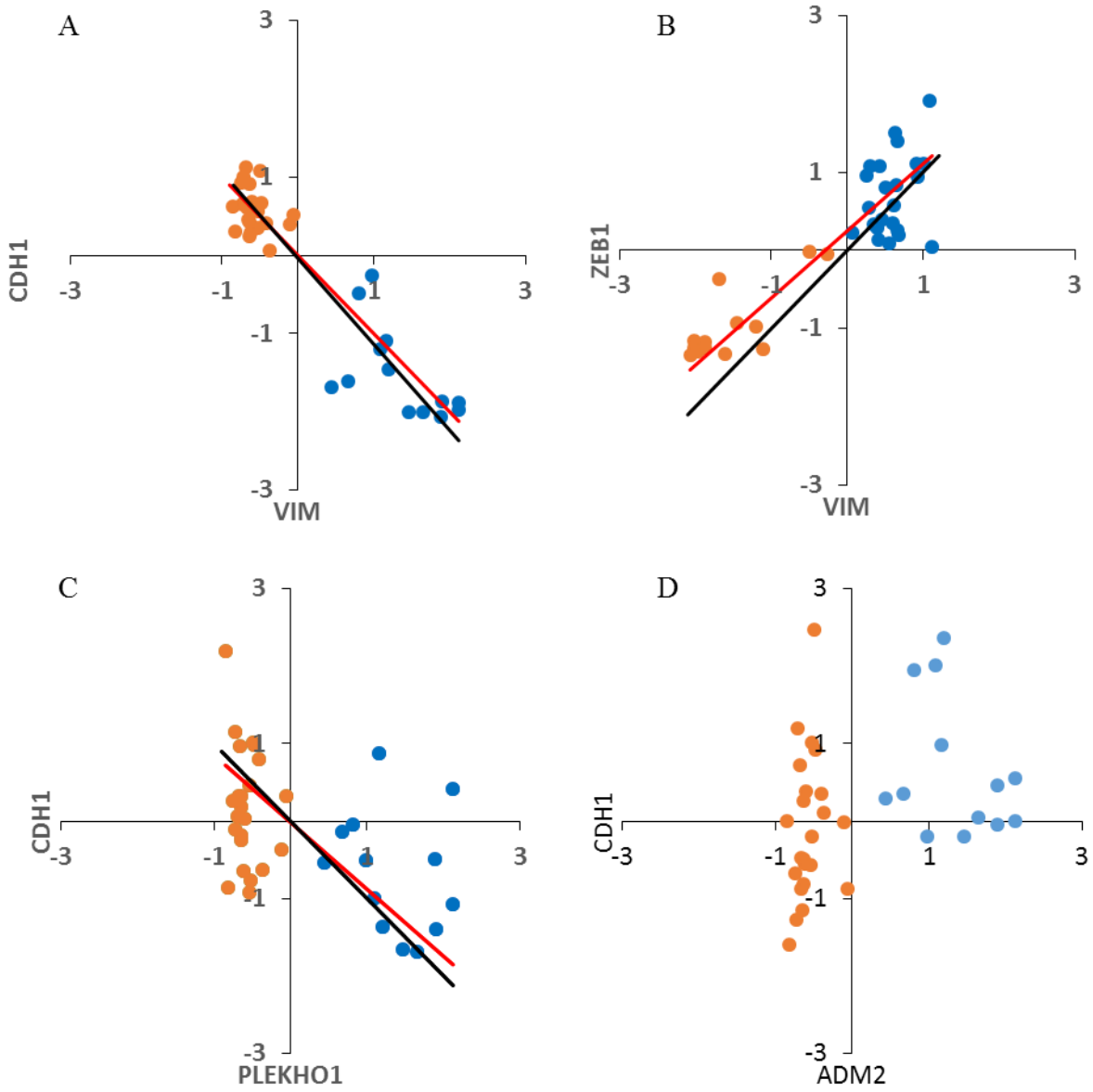


Figure 4. Selected pairwise comparisons between marker and candidate genes.

Values are Z-scores, which are a relative indication of over (positive) or under (negative) gene expression. The data set was split into epithelial (orange) and mesenchymal (blue) groups. The red lines are the reference y=-x (A and C) and y=x (B). The black lines are the fitted reduced maximum axis regression when significant ($P<0.0005$).

Table 14 Pair-wise associations between the three control and three candidate bivalent genes based on a subset of the NCI-60.
Values are slopes of reduced maximum axis regressions. All slopes given are significant at p<0.005.

	CDH1	VIM	ZEB1	ADM2	PLEKHO1	RASA3
CDH1						
VIM	-1.10					
ZEB1	-0.94	0.86				
ADM2	N.S.	N.S.	N.S.			
PLEKHO1	-0.87	0.79	0.92	N.S.		
RASA3	-0.84	0.77	0.89	N.S.	0.97	

The *in vitro* analysis used a collection of nine breast cell lines that consisted of five epithelial and four mesenchymal lines. I harvested RNA from these lines and used qRT-PCR to determine the expression level of the three candidate and control genes (Figure 5). As a method of evaluating the data I used the ratio of the average expression in mesenchymal to epithelial cell lines (Table 15).

Table 15 Summary of expression levels from the nine cell line panel assays expressed as the average epithelial or mesenchymal expression and the ratio.

Gene	Average epithelial expression	Average mesenchymal expression	Ratio (M/E)
ZEB1	0.39772	24.40815	61.37016
VIM	0.690588	52.65554	76.2474
CDH1	0.67142	0.002016	0.003003
ADM2	1.040104	6.917008	6.650306
PLEKHO1	1.22027	9.973599	8.173273
RASA3	0.883939	7.216412	8.163928

This largely mirrored what I observed in the NCI-60 data set with the mesenchymal cell lines having substantially higher expression levels of PLEKHO1, RASA3, VIM and ZEB1 compared to the epithelial cell lines and the reverse for CDH1. Interestingly, unlike in the NCI-60 data set, ADM2 displayed a strong mesenchymal bias in these nine breast cell lines. To reconcile this difference with the NCI-60 data set I examined ADM2 expression in the five NCI-60 breast cell lines. This revealed that in these ADM2 is preferentially expressed in the mesenchymal lines with an average Z-score of 0.275 in epithelial lines versus 1.22 in

mesenchymal lines. This indicates that the role of ADM2 may be more cell type specific than for PLEKHO1 and RASA3. Overall the *in vitro* data supports the previous evidence that ADM2, PLEKHO1 and RASA3 are predominantly expressed in mesenchymal cell lines. For PLEKHO1 and RASA3 this relationship was consistent across multiple cell types while for ADM2 this was predominantly in breast cells. Of note, for the three markers, ZEB1, CDH1 and VIM, there was a much more pronounced difference in expression level compared to the three candidate genes, ADM2, PLEKHO1 and RASA3.

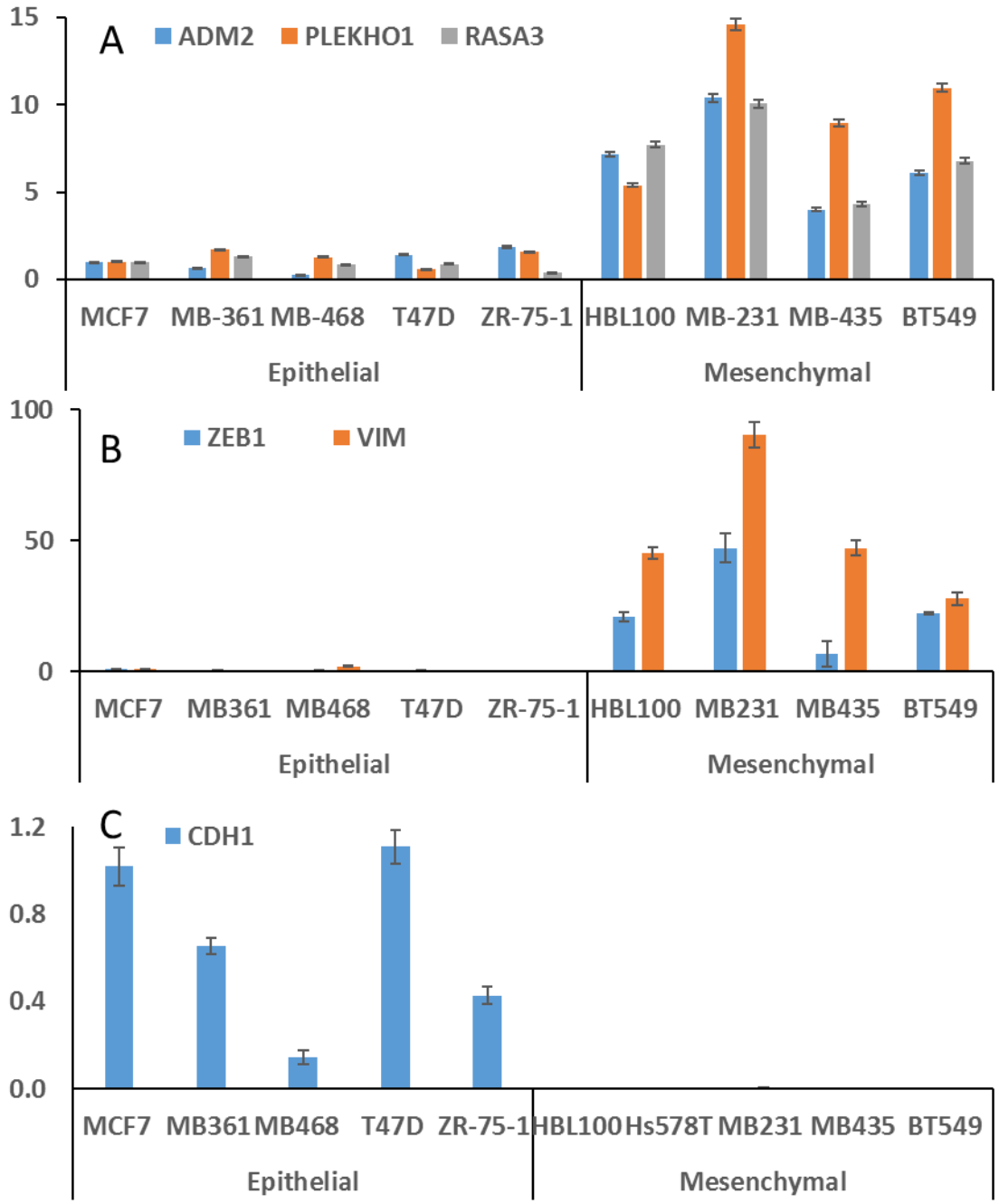


Figure 5 Expression levels of ADM2, PLEKHO1, RASA3 and the three EMT markers in nine breast cell lines.

qRT- PCR of ADM2, PLEKHO1 and RASA3 (A), ZEB1 and VIM (B) and CDH1 (C) in nine breast cell lines relative to MCF7 and normalised to GAPDH. Error bars represent error propagation of three technical replicates.

3.2.2 ADM2, PLEKHO1 and RASA3 have validated bivalent promoters

To determine whether the shortlisted bivalent genes were common to other systems, I compared the bivalent ChIP-seq data (Section 3.1.1) to a study that analysed bivalent promoters in colorectal cancer versus matched mucosal tissue [81]. There was a 50% overlap in bivalent genes between the HMLE data set and the mucosal tissue (Supplementary Table 1). However, when comparing genes that went from bivalent to active in HMLE-mesHMLE and mucosal-cancerous the overlap was only 10%. This suggests that while there might be many shared bivalent genes, different ones will be switched on in different systems that undergo varied processes. Importantly, none of the genes I chose to study underwent a bivalent to active transformation in the colorectal cancer data set and only ADM2 was present as a bivalent gene in that data set. Interestingly, while in our data set the master EMT transcription factors ZEB1 and TWIST underwent a bivalent to active transformation, SLUG filled this role in the CRC system [81]. This difference highlights the growing consensus that different master EMT transcription factors and pathways are important in different systems [112-114]. Supporting this, ZEB1 was bivalently regulated and acted as a key driver of EMT in a breast cancer study [115].

Other potential sources for differences between data sets are relative enrichment levels and cut-offs used for determining whether a peak is positive for a given epigenetic mark. For example, Malouf *et al* [116] evaluated the presence of bivalent genes in HMLE cells undergoing a TWIST-induced EMT. Surprisingly, they found only 33% the amount of bivalent genes as we did in a very similar model. Unfortunately, as their raw data are not available I cannot determine whether there was a significant overlap between our data sets.

To classify a gene promoter as bivalent it is important to verify that the activating H3K4me3 modification and the repressive H3K27me3 modification are present simultaneously on the same allele, rather than each being present alone on different alleles, or in different cells in the population. A possible complication with identifying bivalent promoters from two bivalent marks from ChIP-seq experiments is that it is difficult to ensure that both of these marks were present on the same chromosome from the sequencing data alone. To confirm that both epigenetic marks are present at the same allele, the typical experiment is a sequential ChIP against one epigenetic mark followed by the other. This ChIP-reChIP technique ensures that only DNA with the first epigenetic mark is available for the second ChIP, maximising the likelihood of co-localisation at the same allele. Andrew Bert performed these Chip-reChIP experiments in HMLE cells to confirm that both marks were present within the same region of chromatin. This was attempted for both H3K27me3/H3K4me3 and the reverse, however enrichment was only observed for H3K27me3/H3K4me3 ChIP-reChIP. The data was expressed as a percentage of input, this is an indication of how much of the total DNA encompassing this region contains the epigenetic marks chipped for. This percent input was then normalised to the negative control average of ACTNB and GAPDH promoters and we showed that ADM2, PLEKHO1 and RASA3 had 2-5 fold greater enrichment compared to the control in HMLE cells and this bivalent mark was almost completely lost in mesHMLE cells (Figure 6). With these confirmations that ADM2, PLEKHO1 and RASA3 were predominantly mesenchymally-expressed and had *bona fide* bivalent promoters that were activated during EMT I subsequently studied their effects on the EMT properties of proliferation and migration. To do this it was necessary to modulate their expression. I did this through overexpression and knockdown systems for each gene.

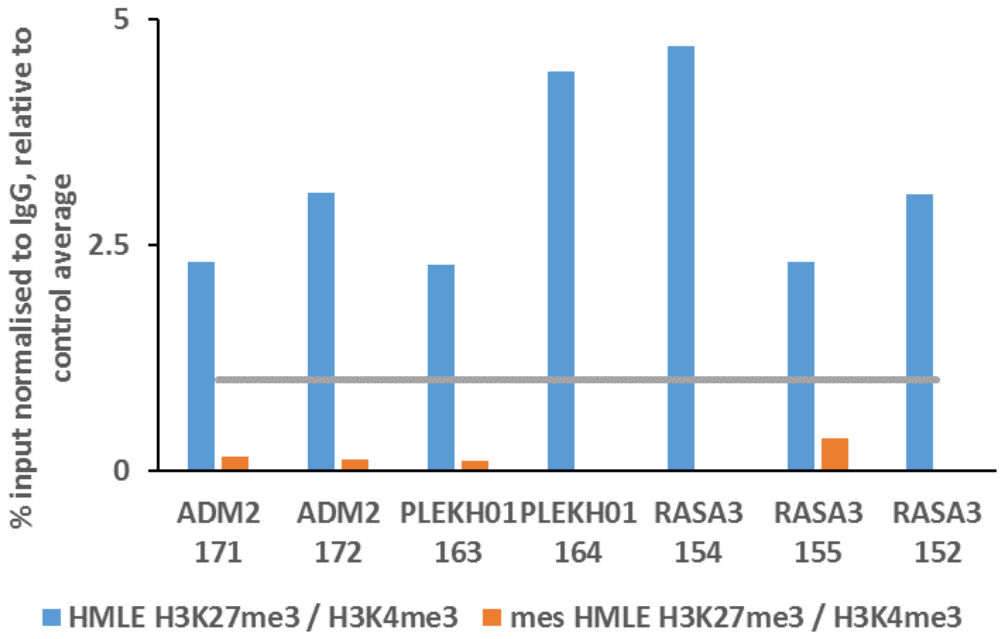


Figure 6 Enrichment levels of bivalent epigenetic marks for PLEKHO1, ADM2 and RASA3.

Enrichment levels of epigenetic marks were measured using ChIP-reChIP. $1x10^6$ HMLE (blue) or mesHMLE (orange) cells were used for each experiment. Comparison of input levels of three bivalent genes measured by qRT-PCR and normalised to H3K27me2/IgG control. Expression values are relative fold change normalised to the average for GAPDH, ACTNB, PAX5 and MYOD. The grey line denotes no change. SD between two technical replicates was <1% of CT values.

3.2.3 Construction of ADM2, RASA3 and PLEKHO1 expression vectors and selection of stable overexpression cell lines

I created lentivirus constructs to transduce both HMLE and MCF10A cells with constitutive overexpression vectors coding for ADM2, PLEKHO1 or RASA3. Each construct was under a strong CMV promoter. I then attempted to expand these as clonal pools using antibiotic selection. Through multiple attempts the HMLE cells did not grow successfully while the MCF10A cells proved more efficient yielding the requisite four clonal pools, MCF10A_EV, MCF10A_ADM2, MCF10A_PLEKHO1 and MCF10A_RASA3. I used qRT-PCR to confirm

that ADM2, PLEKHO1 and RASA3 were significantly overexpressed (Figure 7). Importantly, overexpressing any one of these did not affect expression of the others. I attempted immunoblotting with two antibodies for ADM2 and PLEKHO1 and three for RASA3 however, I was unable to obtain clear signal from these.

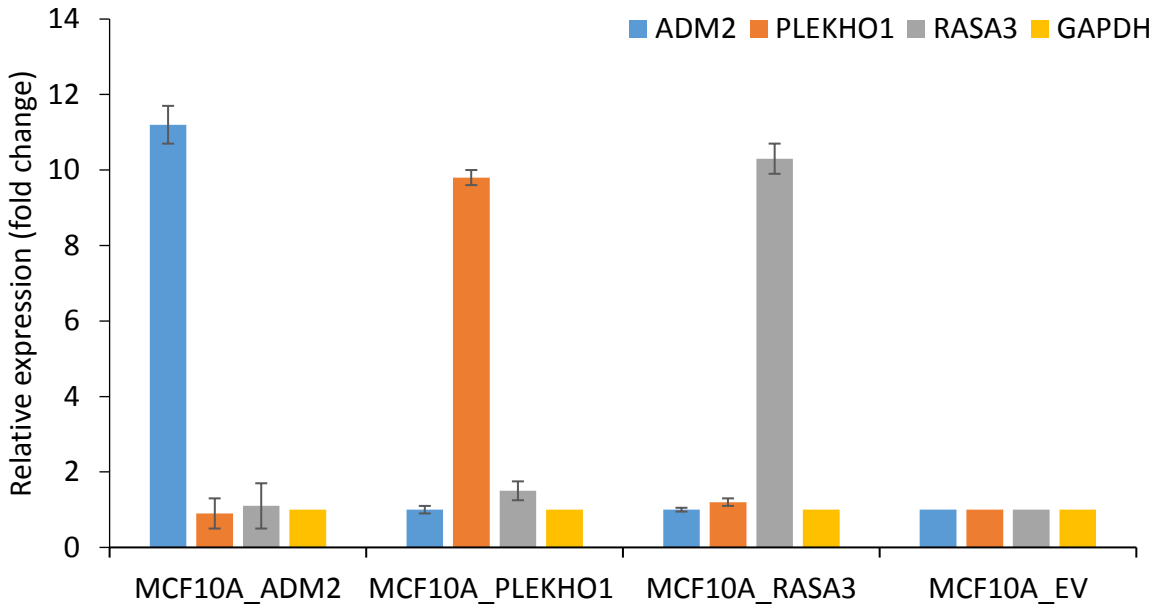


Figure 7 Expression level of ADM2, PLEKHO1 and RASA3 in MCF10A stable cell lines

qRT-PCR of ADM2, PLEKHO1 and RASA3 from MCF10A stable cell lines relative to the EV control and normalised to GAPDH. Error bars represent ±SD of three biological replicates.

I was primarily interested in identifying a putative role for ADM2, PLEKHO1 or RASA3 in the EMT associated property of migration. A key consideration when performing migration assays is to ensure that either the proliferation rate is largely similar across the treatments or that the migration assay is brief to minimise potentially confounding variables. Despite my assay being brief, under 24h, I still determined the proliferation rate for each of the cell lines.

3.2.4 ADM2, PLEKHO1 and RASA3 have no consistent effect on proliferation

I assayed the effect of ADM2, PLEKHO1 and RASA3 on proliferation rate using the IncuCyte. The Incucyte consists of a microscope gantry that resides in a cell incubator. This allows for photographs of hundreds of wells to be taken at regular time intervals. A networked external controller hard drive gathers and processes the image data. These images can be used to obtain insights into multiple aspects of cell morphology, physiology and behaviour as well as creating time lapse videos using the still images to obtain qualitative behaviour data. Specifically, for proliferation assays, I used the cell confluence data provided by the IncuCyte. To perform the assay MCF10A_ADM2, MCF10A_PLEKHO1 and MCF10A_RASA3 were seeded at $3x10^3$ cells/well and photographed over a 48h time course. This was performed in triplicate and a small but statistically significant increase in proliferation was detected for MCF10A_RASA3 but not for MCF10A_ADM2 or PLEKHO1 (Figure 8).

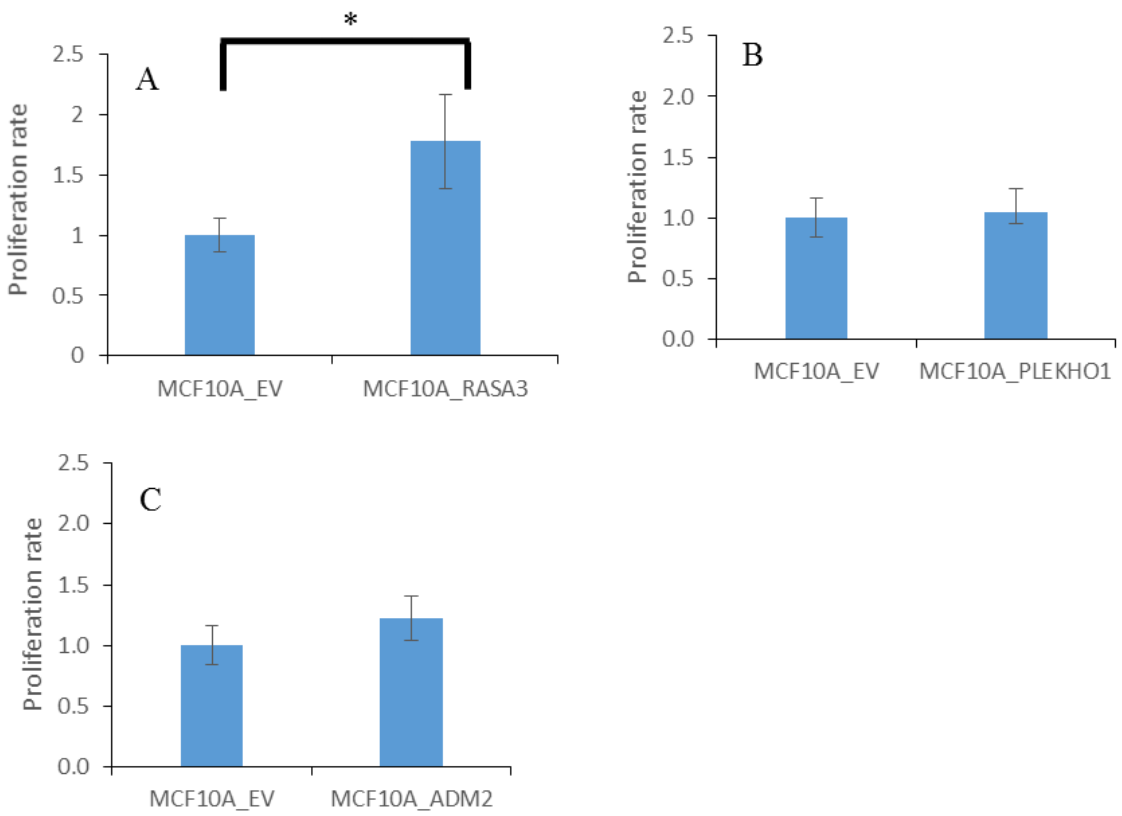


Figure 8 Proliferation rate of MCF10A stable cell lines overexpressing RASA3, PLEKHO1 or ADM2.

Stable MCF10A cell lines overexpressing RASA3 (A), PLEKHO1 (B) or ADM2 (C) were seeded in a 96-well plate with 6 technical replicates. Proliferation rate was measured as confluence using the IncuCyte and normalised to the empty vector control. Error bars represent ±SD of three biological replicates. Asterisk indicates a statistically significant difference ($p<0.05$) compared to the empty vector.

As a complement to these experiments I performed transient siRNA knockdown experiments in MDA-MB-231 and mesHMLE cells. I purchased three different siRNAs against each gene from the Sigma Silencer-Select range. I transiently transfected these into MDA-MB-231 and mesHMLE cells individually and as a pool to determine what produced the strongest knockdown. While the pool did, in some cases, produce slightly stronger knockdown effects, the problem with a pool is that it enhances off target effects. Consequently, as the difference was small, I chose to continue with a single siRNA for each gene (Figure 9 B).

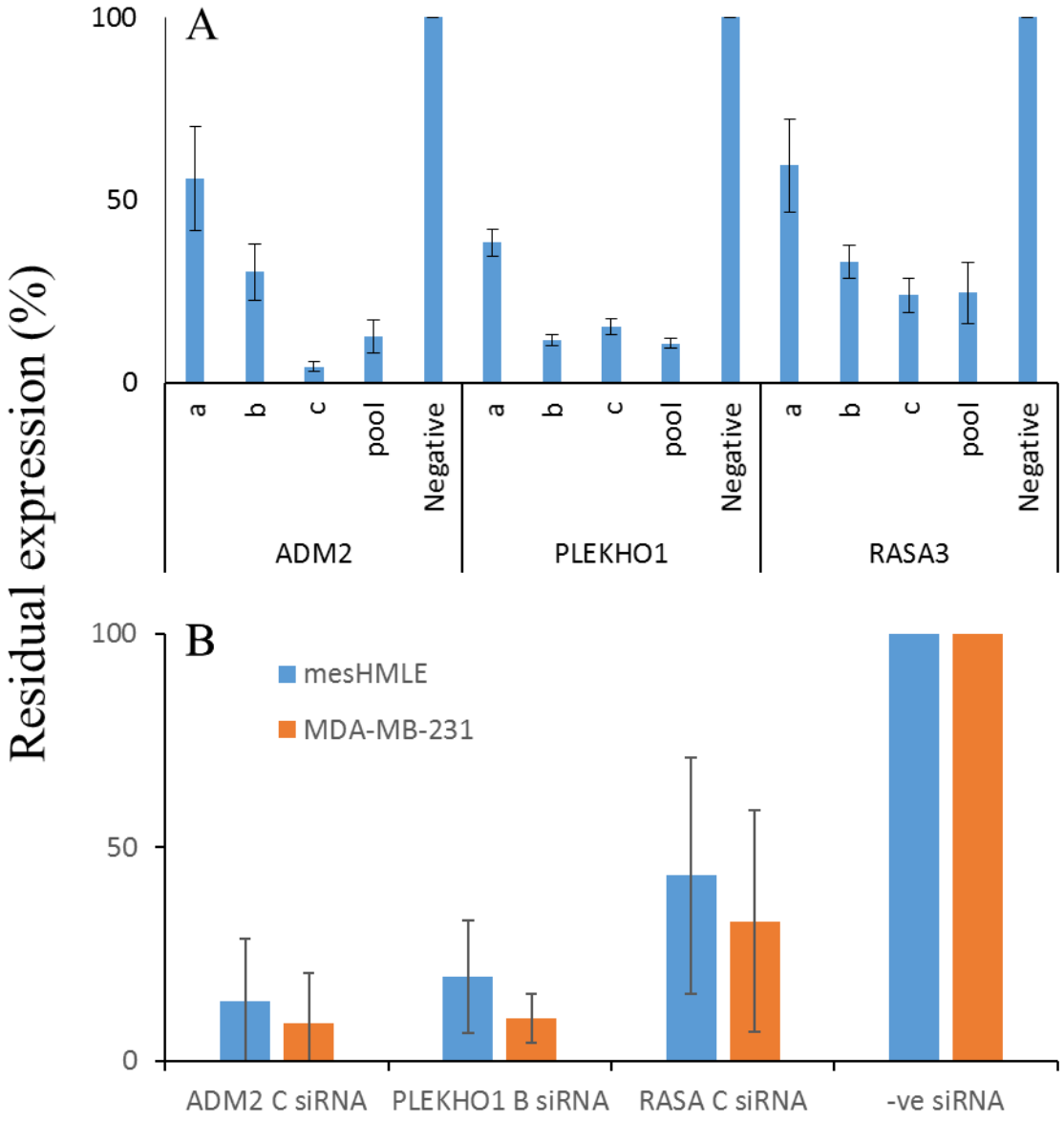


Figure 9 Expression level of ADM2, PLEKHO1 and RASA3 in mesHMLE and or MDA-MB-231 cells after transient siRNA knockdown.

MDA-MB-231 and mesHMLE cells were transiently transfected with three different siRNA and a pool (A) or the strongest siRNA of these (B). Values are relative to the negative siRNA control and normalised to GAPDH. Error bars represent the error propagation of three technical replicates (A), ±SD of three biological replicates ((B), MDA-MB-231) or ±SD of two biological replicates ((B), mesHMLE).

I carried out the proliferation assays for mesHMLE and MDA-MB-231 cells as described for the MCF10A cells except that the cells were first seeded and transfected in a 6-well plate and then plated into a 96-well plate for the proliferation assay. The results largely mirrored those observed with the MCF10A cells with ADM2 knockdown resulting in a modest decrease in proliferation while this was only observed in mesHMLE cells for PLEKHO1 knockdown and in neither cell line for RASA3 (Figure 10).

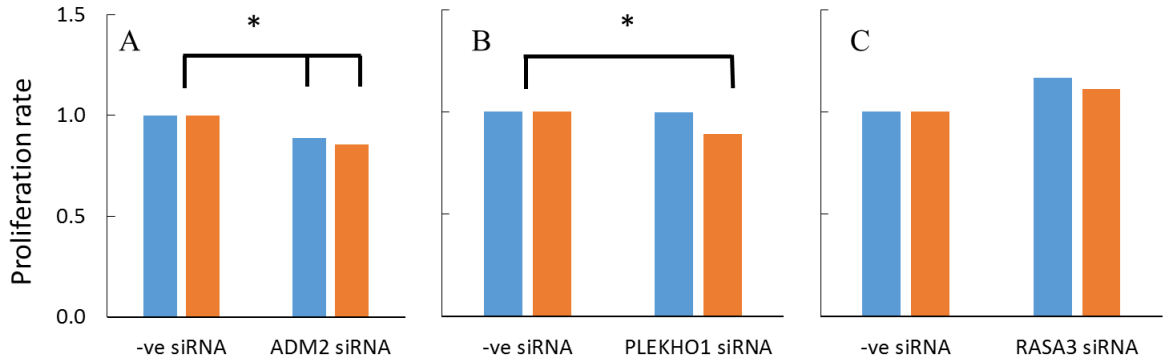


Figure 10 Effect on proliferation of transient siRNA transfection of RASA3, PLEKHO1 or ADM2 in MDA-MB-231 and mesHMLE cells

mesHMLE (orange) and MDA-MB-231 (blue) cells were transiently transfected with RASA3 (A), PLEKHO1 (B) or ADM2 (C) and seeded in a 96well plate with 6 technical replicates. Proliferation rate was measured as confluence using the IncuCyte and normalised to the empty vector control. Error bars represent ±SD of three biological replicates. Asterisk indicates a statistically significant difference ($p<0.05$) compared to the –ve siRNA transfection control.

These three sets of experiments showed that these three genes have no consistent effect on proliferation. This facilitates the use of both overexpression and knockdown systems in understanding the role of these genes in migration.

3.2.5 PLEKHO1 and ADM2 overexpression increases cell migration in MCF10A cells

I was interested in identifying whether ADM2, PLEKHO1 or RASA3 overexpression affected migration. Migration was of interest due to the links between migration and cancer metastasis outlined in 1.2.4. To assess this I performed wound healing assays using the Incucyte technology described in the previous section. Briefly, in a wound healing assay cells are plated in a 96-well plate and then scratched using the Wound Maker, which ideally results in a single clean and clear scratch through the middle of the well. The rate of wound closure can be used as an approximation of migration. This rate can be derived in two main ways, both of which involve microscopy. The first is to simply photograph the cells at the time they are scratched and again 24h later to determine how much area has been covered. The second involves taking photos of the cells at regular time intervals, which allows for two things. Firstly, one can determine the rate of closure over time in greater detail, consequently it is possible to determine whether this changes over time. Secondly, one can use the still images to make a time lapse video. This is useful to visually determine whether there is a different migration mechanism between different samples.

I seeded the MCF10A stable overexpression cells as described (2.1.10), using 12 replicates per cell line and they were allowed to reach confluence overnight. I then scratched these using the WoundMaker and placed them in the Incucyte. After 24hours I identified which wells had been cleanly scratched and selected these, typically resulting in 3-6 wells per cell line. This was repeated in triplicate and I found that there was a 1.5-2x statistically significant increase in rate of wound closure for MCF10A cells with stable overexpression of either ADM2 or PLEKHO1 but not RASA3 (Figure 11).

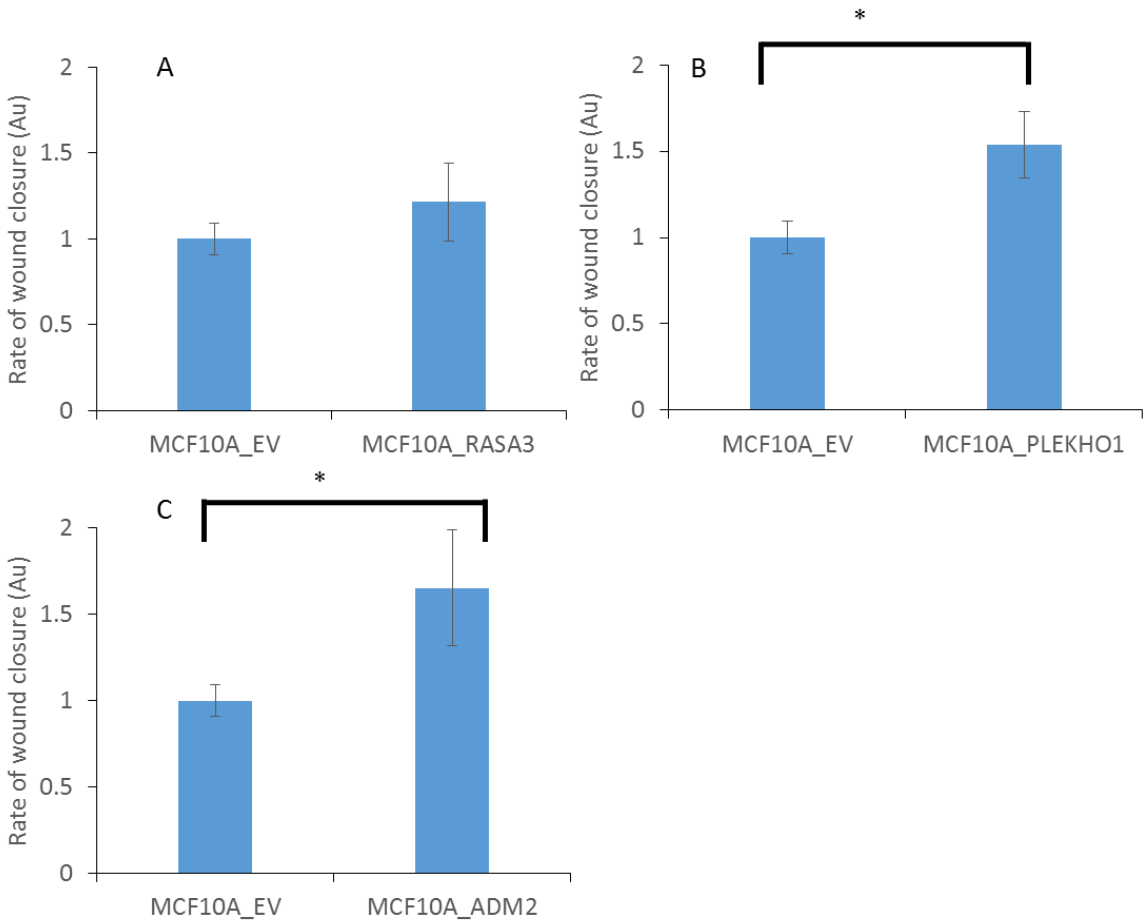


Figure 11 Effect on migration of constitutive overexpression of RASA3, PLEKHO1 and ADM2 in MCF10A cells

MCF10A stable overexpression cell lines, RASA3 (A), PLEKHO1 (B), and ADM2 (C), were seeded at $4x10^4$ in a 96well plate with 12 technical replicates then scratched using WoundMaker. Migration rate was normalised to the empty vector control. Error bars represent ±SD of three biological replicates. Asterisks indicate statistically significant difference ($p<0.05$) compared to the empty vector

3.2.6 PLEKHO1 and ADM2 knockdown decreases migration in MDA-MB-231 and mesHMLE cells

To further verify that ADM2 and PLEKHO1 contribute to cell migration, I examined the effect of their knockdown in two mesenchymal cell lines, (MDA-MB-231 and mesHMLE Figure 12). The results from the siRNA transfections mirrored my observations with the

stable over expression with a 30-40% reduction in migration in ADM2 and PLEKHO1 knockdown cells compared to cells transfected with a negative siRNA control and no significant change for RASA3 knockdown cells.

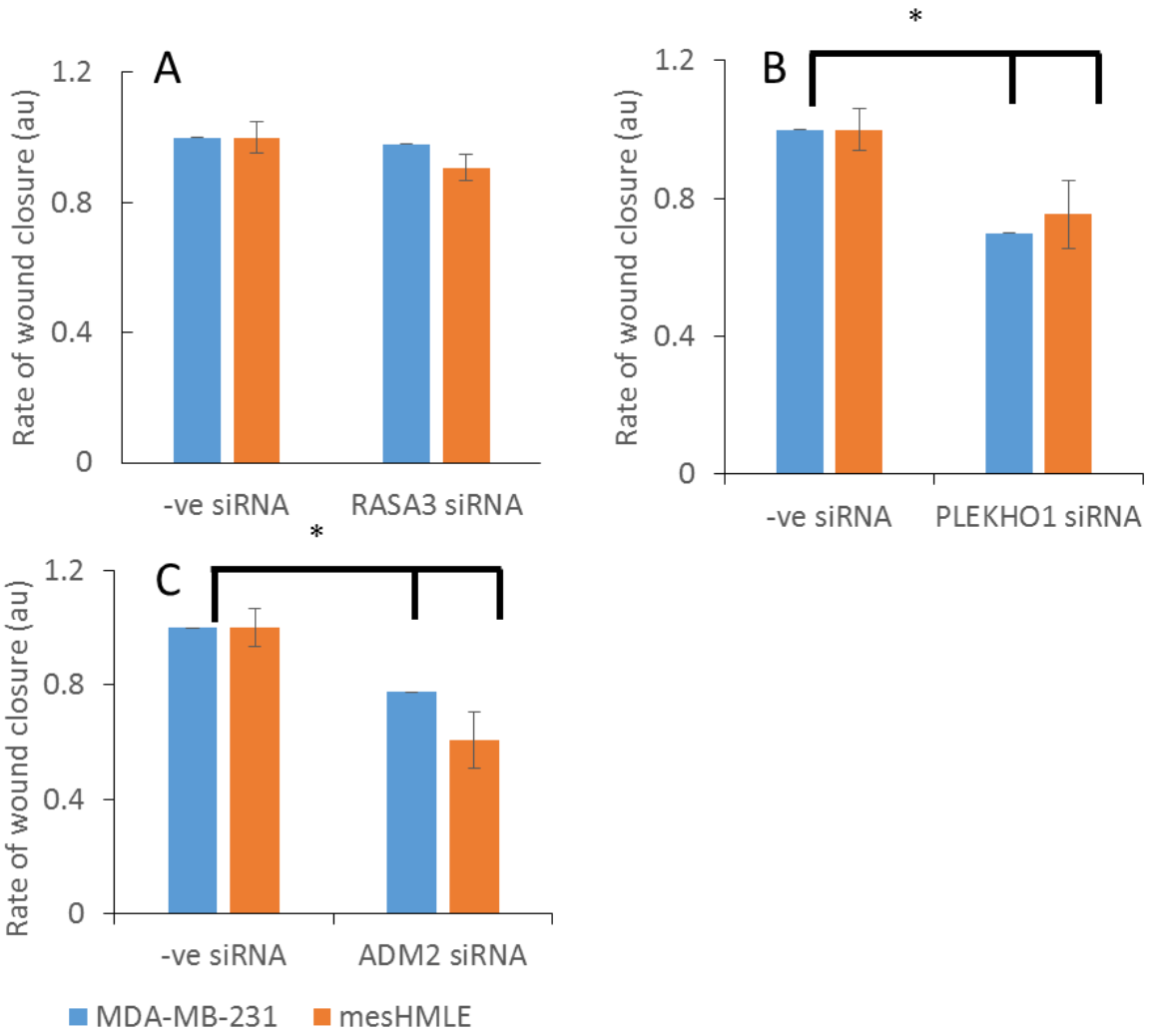


Figure 12 Effect of transient siRNA knockdown of RASA3, PLEKHO1 and ADM2 on migration in MDA-MB-231 and mesHMLE cells

siRNA transfected mesHMLE and MDA-MB-231 cells were seeded at $4x10^4$ in a 96well plate with 12 technical replicates then scratched using WoundMaker. Cells were transiently transfected with siRNA against RASA3 (A), PLEKHO1 (B) or ADM2 (C) alongside a negative siRNA control. Migration rate was normalised to the negative siRNA control. Error bars represent ±SD of three biological replicates. Asterisks indicate statistically significant difference ($p<0.05$) compared to the empty vector

3.2.7 Role of EGF signalling in PLEKHO1 mediated migration

After finding that PLEKHO1 and ADM2 expression correlated with migration I investigated what mechanism was responsible for this effect, principally for PLEKHO1. I based my hypothesis on a paper published in late 2015 by Kim *et al* [105]. They showed that in response to EGF mediated PIP3 production, PLEKHO1 shuttles CK2α to the cell membrane where it phosphorylates PAK1 a key downstream effector of EGF, and other, signalling pathways. The activation of PAK1 then increased migration. To test this, I chose to examine the effect of EGF on my MCF10A stable overexpression cell lines. As MCF10A cells typically grow in media that contains EGF, other growth factors and serum it was key to starve the cells prior to adding EGF.

I used a 6h starvation period to minimise EGF signalling without negatively affecting the cells. To resolve the issue of EGF signalling rapidly activating a negative feedback loop while I needed to measure migration over at least 12h I examined the effects of EGF at 25ng/ml or 50ng/ml for either 15m or as a continued dose with media replenishment after 12h compared to a treatment that remained starved throughout the experiment.

These results largely supported my hypothesis. As expected, both cell lines exhibited virtually no proliferation or migration when grown under starved conditions (Figure 13). All cell lines treated with EGF for 15m gave a modest increase in migration. The relative migration rate was similar to that for cells grown in full media (1.5-2 fold). However, when cell lines were treated with a constant EGF dose, the rate of migration increased for every cell line with the most dramatic increase occurring in MCF10A_PLEKHO1 cells. Their migration rate was statistically the same as when grown in full media and displayed a 3-fold increase in relative migration rate compared to MCF10A_EV.

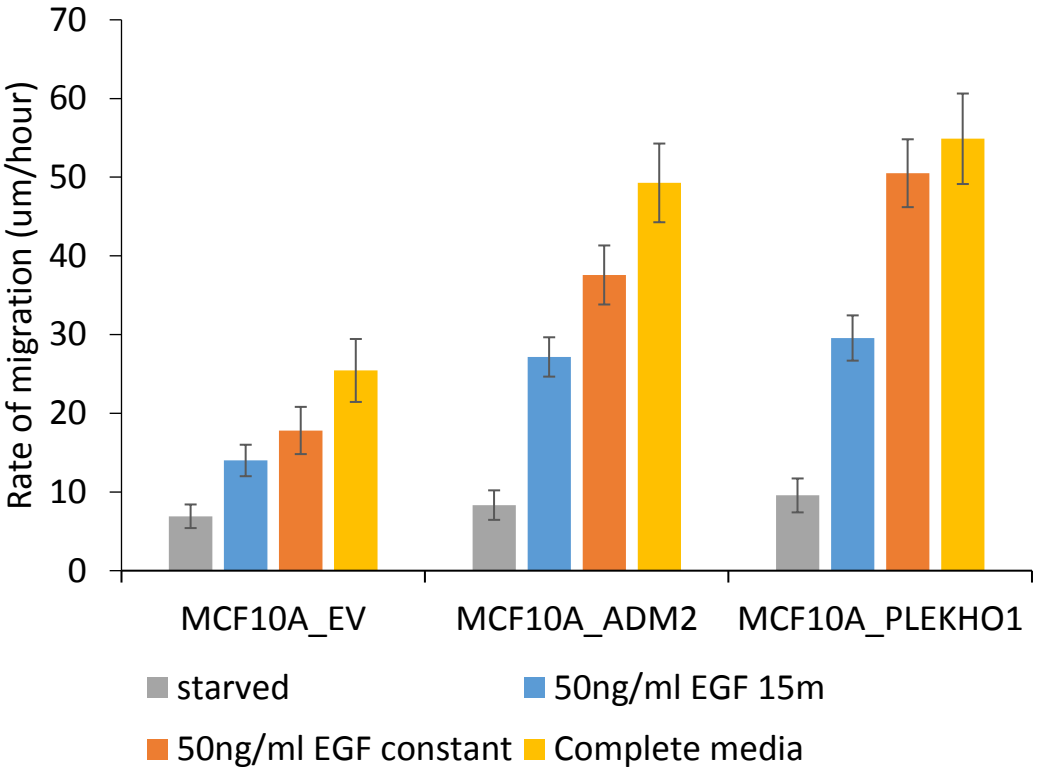


Figure 13 Effect of EGF induction on migration of MCF10A stable cell lines overexpressing ADM2 or PLEKHO1

MCF10A stable overexpression cell lines were seeded at $4x10^4$ in a 96well plate with 8 technical replicates and grown to confluence then starved for 6hs. After the scratch procedure cells were treated with different media, as described. Error bars represent ±SD of three technical replicates.

As mentioned, MCF10A_PLEKHO1 cells grown in constant EGF or full media displayed the same rate of migration while for the other cell lines there was a statistically significant difference between EGF and full media treatments. This suggests that the main influence in PLEKHO1-mediated migration was from EGF signalling and that the additional factors in full media have little effect on the migration rate of MCF10A_PLEKHO1 cells in the presence of EGF. This combination of observations suggests that there could be a maximum rate of migration for MCF10A cells determined by mechanical properties under the circumstances tested. For parental cells, this maximum is achieved under the influence of the

combination of growth factors and serum present in their normal growth media. The addition of either ADM2 or PLEKHO1 increases this rate to a new maximum. For MCF10A_ADM2 the hypothesised maximum requires the same, or similar, combination of factors. However, for MCF10A_PLEKHO1 cells this maximum is reached through just EGF. In summary

1. MCF10A_EV cells grown in their full media migrate at a rate determined by the growth factors and serum present.
2. MCF10A_PLEKHO1 cells grown in full media migrate at an increased rate compared to MCF10A_EV.
3. When either cell line is grown with no growth factors or media, there is virtually no migration highlighting the importance of growth factors in this process.
4. MCF10A_EV cells treated exclusively with EGF migrate at an intermediate rate, lower than that under full media.
5. MCF10A_PLEKHO1 grown in EGF mirrors the migration rate observed in full media.

This suggests that EGF accounts for some of the migration observed in parental MCF10A cells but is not responsible for all migration points #1 and #4. PLEKHO1 overexpression increases this migration rate but is dependent on growth factors in the media, primarily EGF (points #2, #3, #5). This does not mean that all of the migration observed in point #2 is due to PLEKHO1 and EGF but it does suggest that this combination is responsible for the majority of the migration observed, thus highlighting the importance of PLEKHO1 in EGF mediated migration in MCF10A cells.

3.2.8 Co-culturing ADM2 and PLEKHO1 stable cell lines results in a synergistic increase in migration

Interestingly, ADM2 has been shown to activate PI3K signalling in cardiac tissue [117]. Based on my previous data, of the interaction between EGF and PLEKHO1, this raised the question of whether ADM2 and PLEKHO1 function synergistically. As ADM2 is a secreted hormone [95] this did not require co-transfection assays. I first tested this through direct co-culturing of the cells and performing scratch wound assays as described (2.1.10).

For the co-culturing experiment I combined MCF10A_ADM2, MCF10A_PLEKHO1 and MCF10A_EV in factorial pairs and treated them with either full media, continuous EGF, 15m EGF or starvation. In all cases the combination of MCF10A_EV with either MCF10A_PLEKHO1 or MCF10A_ADM2 increased migration compared to the empty vector alone (Figure 14). For the 15m EGF treatment, MCF10A_PLEKHO1 combined with MCF10A_ADM2 had a greater rate of migration compared to either individually. This trend was not repeated in the continuous EGF or complete media treatment supporting my previous hypothesis that there is a maximal rate of migration in MCF10A_PLEKHO1 cells and that this is primarily reached through EGF. As I hypothesised that ADM2 was influencing PLEKHO1 through EGF signalling, I would not expect a further increase as a result of co-culturing when EGF is sufficient.

The most exciting result from these experiments was the effect on migration under starved conditions. As previously described, starving cells grown in isolation displayed virtually no migratory capacity. This observation was repeated for all of the combinations apart from MCF10A_ADM2 + MCF10A_PLEKHO1 where there was a significant increase in rate of migration approaching 1/3rd the migration rate of MCF10A_PLEKHO1 in full media. This

partial rescue of migration through the addition of two novel EMT genes is strong evidence for their importance in EMT.

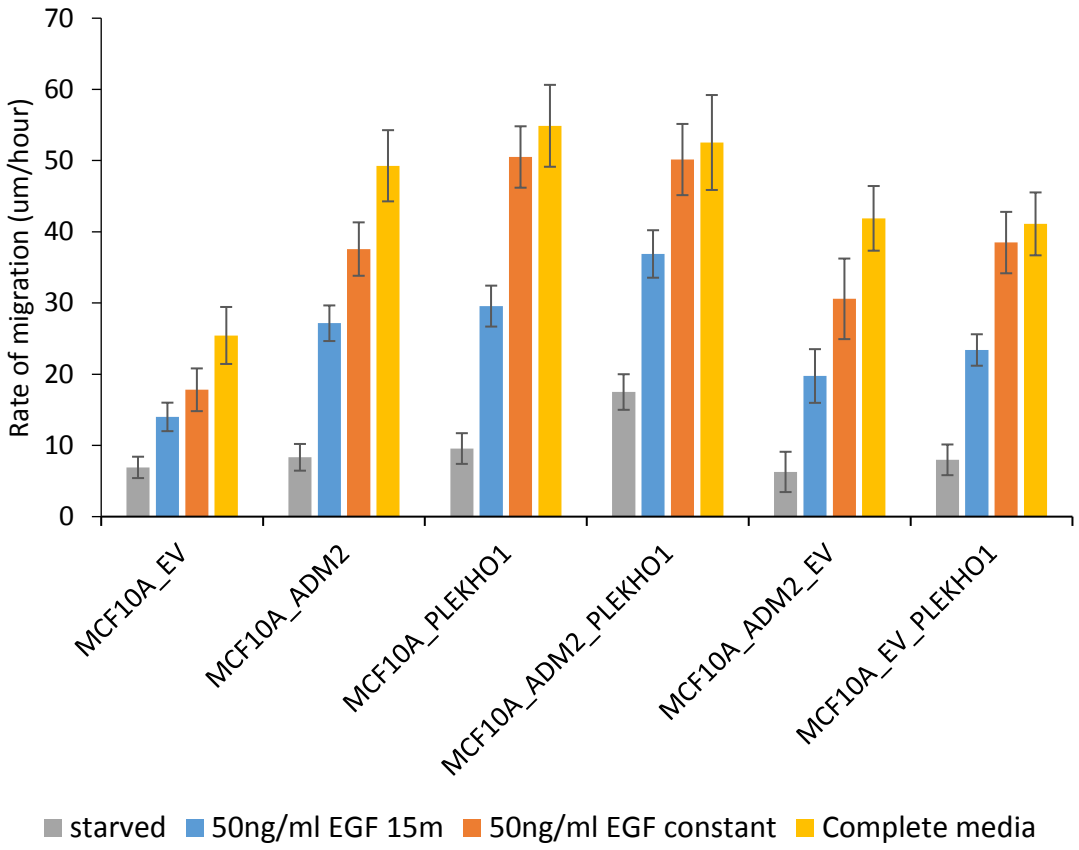


Figure 14 Effect of co-culturing MCF10A_EV, MCF10A_ADM2 and MCF10A_PLEKHO1 under different media conditions

MCF10A stable overexpression cell lines were seeded at $2x10^4$ of each cell line in a 96well plate with 8 technical replicates, grown to confluence then starved for 6hs. After the scratch procedure cells were grown in different media, as described. Error bars represent ±SD of three technical replicates

To confirm that the observed synergistic effect on migration was due to the secreted nature of ADM2 I examined the effect of MCF10A_ADM2 conditioned medium on MCF10A_PLEKHO1 migration using the same factorial pairs.

3.2.9 Treating MCF10A_PLEKHO1 cells with MCF10A_ADM2-conditioned media increases migration rate

For the majority of cell line combinations tested there did not appear to be a difference in migration between ADM2 conditioned media or serum-free media. However, PLEKHO1 cells treated with ADM2 media exhibited a dramatic increase in migration rate (Figure 15 A). As this experiment was only performed once I do not have statistical evidence to support this, however the results appear quite strong and are in keeping with my previous observations and the hypothesised pathway. An interesting observation is that MCF10A_PLEKHO1 cells treated with ADM2 media had an elevated migration rate compared to either MCF10A_PLEKHO1_ADM2 or MCF10A_PLEKHO1_EV. A likely explanation for this is that MCF10A_PLEKHO1 is a pure population of cells that is acted on by ADM2 while the MCF10A_PLEKHO1_EV/ADM2 are combinations of the two cell lines where only half of the population responds to ADM2. An experiment to test this would be to co-culture different ratios of cells. If this hypothesis is correct, then I would expect the migration rate to largely correlate with the proportion of MCF10A_PLEKHO1 until there were insufficient MCF10A_ADM2 cells to activate the MCF10A_PLEKHO1 cells.

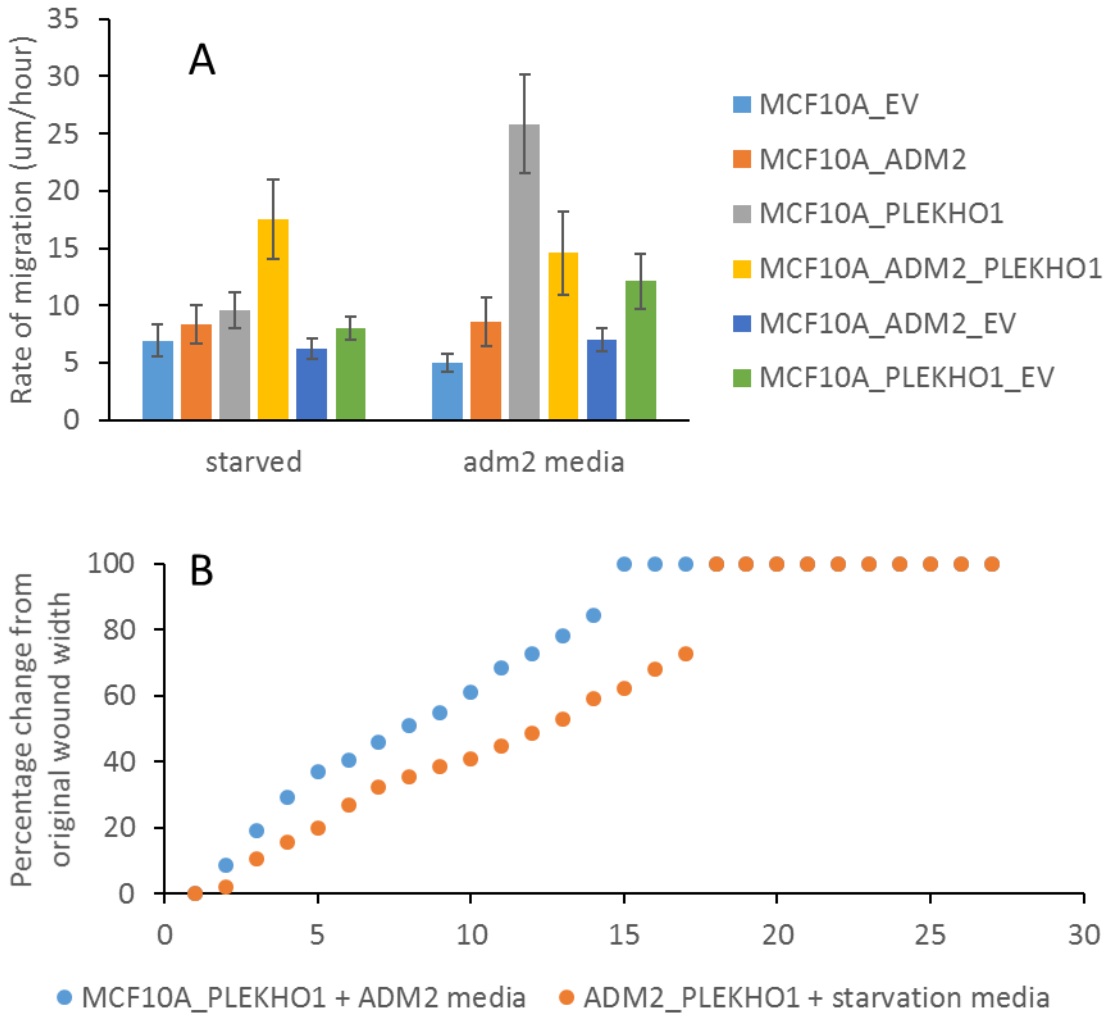


Figure 15 Effect of ADM2 media on individually or co-cultured MCF10A_EV, MCF10A_ADM2 and MCF10A_PLEKHO1.

MCF10A stable overexpression cell lines were seeded at a total of $4x10^4$ cells, with a 50:50 ratio where applicable, in a 96well plate with 8 technical replicates, grown to confluence then starved for 6hs. After the scratch procedure cells were grown in either starved or ADM2 conditioned media as described. (A) Rate of migration expressed as um/hour from the Incucyte data. (B) Percentage change in wound-width relative to the original wound-width for MCF10A_PLEKHO1 cells grown in ADM2 media (blue) and co-cultured MCF10A_ADM2 and MCF10A_PLEKHO1 cells grown under starvation conditions (orange). Error bars represent ±SD of 3-5 technical replicates.

Importantly, there was a greater increase in migration for MCF10A_PLEKHO1 cells treated with ADM2 media than for MCF10A_PLEKHO1 cells co-cultured with MCF10A_ADM2

cells under starvation conditions (Figure 15 A). A possible reason for this difference in migration rate is that when MCF10A_PLEKHO1 and MCF10A_ADM2 cells are co-cultured it takes time for ADM2 to be produced, secreted and taken up by the MCF10A_PLEKHO1 cells while when ADM2 media is added directly to the MCF10A_PLEKHO1 cells the only required step is uptake. Consequently, when the stable cell lines are co-cultured in clean media there is a lag period before enough ADM2 is secreted to affect MCF10A_PLEKHO1 cells. This can be observed by looking at the cumulative percentage change in wound-width relative to the original wound-width (Figure 15 B). This combination of factors strongly support my original hypothesis that ADM2 acts upstream of PLEKHO1 and can mediate PLEKHO1 migration.

To summarise:

- MCF10A_PLEKHO1 cells exhibit statistically greater migration than MCF10A_EV cells in complete media (Figure 11).
- MCF10A_PLEKHO1 cells do not have statistically greater migration than MCF10A_EV cells under starvation conditions but do after EGF stimulation (Figure 13).
- ADM2 media appears to have no effect on MCF10A_EV migration (Figure 15).
- ADM2 media appears to upregulate migration in MCF10A_PLEKHO1 cells (Figure 15).

This indicates that in MCF10A cells PLEKHO1 mediated migration is dependent on EGF and the other MCF10A media components, however, this can be partially replaced by ADM2.

3.3 Discussion

3.3.1 Overview

Through my research, described in this Chapter, I contributed to the growing evidence that bivalent genes are important to EMT [81, 115, 116]. I first identified a number of established EMT factors that possessed a bivalent epigenetic signature and became activated during HMLE EMT (3.1.3). I also identified the opposite, that is, key epithelial genes that are bivalently regulated after EMT. This implicates them as key genes for MET.

Furthermore, I interrogated the list of bivalent genes that were activated during EMT to identify potential novel EMT associated genes. From this list I chose three genes ADM2, PLEKHO1 and RASA3, that had not previously been associated with EMT. I tested their effects on EMT marker expression through qRT-PCR (3.2.1), and proliferation and migration in mesenchymal and epithelial mammary cells, primarily using the Incucyte, and showed that ADM2 and PLEKHO1 are novel EMT associated genes. The most impactful finding was the link between ADM2 and PLEKHO1 and their synergy to increase migration in MCF10A cells.

3.3.2 Role of bivalent genes in the HMLE EMT model

Our understanding of bivalent genes in EMT remains incomplete with only a handful of studies investigating their role [115, 116]. This is despite the importance of bivalent genes in the similar developmental EMT [82]. To reconcile this gap in the literature I used ChIP-seq on the HMLE EMT cell line model to determine whether bivalent genes were associated with EMT, and if so, what their role in the process was. We found 429 (28%) of bivalent genes in HMLE cells became active after an EMT (3.1.3). This list included prominent EMT genes

such as ZEB1 and TWIST. Interestingly, I found that there were also key epithelial genes, CDH1 and ESRP1 that underwent the opposite epigenetic changes, that is, they transitioned from an active epigenetic state to a bivalent state. Further evidence for the importance of bivalent genes in EMT comes from Gene Ontology studies where migration and morphogenesis were enriched in bivalent-active genes but not in the general bivalent list and MET being enriched in genes that went from active to bivalent. The first observation provides clear evidence that some aspects of cell migration are under bivalent regulation while the second observation strongly suggests that it is not just EMT but also MET that is bivalently regulated. The combination of these findings strongly suggests that the area of bivalent genes in EMT and MET is currently understudied (3.1.3). These findings also raise the possibility of interrogating bivalently regulated gene lists for novel EMT associated genes. I performed pilot experiments to determine whether this was a feasible approach by studying three bivalently regulated genes, ADM2, PLEKHO1 and RASA3.

3.3.3 Role of ADM2, PLEKHo1 and RASA3 on proliferation and migration

By manipulating their expression levels it was evident that none of these genes gave a strong consistent effect on proliferation (3.2.4). However, ADM2 and PLEKHO1 expression strongly correlated with migration (3.2.5, 3.2.6). While I was performing my studies another group published similar findings on a role for PLEKHO1 in migration [105] where they elegantly showed that, in response to EGF signalling, PLEKHO1 brought casein kinase 2 alpha (CK2α) to the plasma membrane where it activated P21 protein activated kinase 1 (PAK1). Without this activation, EGF induced migration decreased significantly in their colorectal cancer cell lines.

To determine whether PLEKHO1 was acting in a similar manner in my model, I tested the role of EGF mediated migration in MCF10A cells (3.2.7). My data strongly support their findings and goes further suggesting that MCF10A_PLEKHO1 + EGF results in the same migration rate as MCF10A_PLEKHO1 + full media.

Beyond this, a paper published during my PhD showed that ADM2 activated the PI3K mediated signalling [118]. As PI3K is the key molecule for PLEKHO1 activity, this raised interesting possibilities for my research as it suggested that ADM2 could act upstream of PLEKHO1 signalling.

I investigated this through two assays (3.2.8 and 3.2.9) and showed that either co-culturing MCF10A_ ADM2 and PLEKHO1 cells or treating MCF10A_PLEKHO1 cells with ADM2 conditioned media resulted in an a partial rescue of migration compared to starvation conditions. This finding strongly suggests a link between ADM2 and PLEKHO1 and also raises the possibility that bivalently regulated genes interact in a network in a similar way to the network hypothesis that has been extended to microRNAs and transcription factors. My findings suggest that epigenetic gene regulation could also fall under this network hypothesis possibly as an extension of transcription factors and chromatin remodelling factors.

3.3.4 Final perspective

There are two key findings from the research outlined in this Chapter. First, that ADM2 and PLEKHO1 are novel EMT associated genes that act synergistically and individually to increase migration rate in a range of breast cells. Second, flowing on from this, that mining bivalently regulated gene lists is a valid, and underused, method to identify novel EMT genes. There remain several hundred genes from my dataset alone that are unexplored. This

provides a resource to investigate their roles in cellular and developmental models of EMT to further illuminate this critical process.

Chapter 4 Dissecting the role of ZEB1 isoforms in EMT

4.1 Introduction

4.1.1 Splicing and isoforms

The one gene one enzyme hypothesis has been the central dogma of molecular biology for decades [119]. Over the past few years this simple model has evolved as our understanding of molecular biology has improved. A key problem with the central dogma is that it does not account for the great diversity of proteins and mRNAs that can arise from a single gene. This variation is largely a result of modifications to the mRNA during splicing and also protein post translational modifications. Of these, I will focus on alternative splicing, which occurs in the nucleus, largely during the processing stage where mRNA is spliced and capped [120].

Alternative splicing is regulated through the interaction of multiple RNA-binding proteins that associate with auxiliary *cis*-elements within pre-mRNA transcripts that can positively or negatively influence splicing of regulated exons and are designated as intronic or exonic splicing enhancers and silencers [121]. The interaction of RNA-binding proteins with these *cis*-elements determines whether an exon is included or not in the mature transcript. Combinatorial control of splicing involves contributions by the generally ubiquitous SR and hnRNP families of proteins in cooperation with tissue-specific regulators of splicing, such as QKI, ESRP1, MBNL, and CELF proteins [121, 122].

Splicing different exons together produces divergent transcripts from the same gene. This can change protein function in a dramatic or modest fashion as well as potentially affecting mRNA stability through changing which miRNA sites are present, amongst other mechanisms. It is estimated that ~95% of multi-exonic genes undergo alternative splicing and

that 85% of multi-exonic genes have a minor isoform that comprises at least 15% of total expression [123, 124]. This enhanced variety is fundamental for life as mutation of the splicing machinery results in embryonic lethal mutants in drosophila, chicken and mice [125-127].

Isoform switches frequently occur during development; they are also crucial for EMT. For example, knockdown of the splicing factor, ESRP1, triggers an EMT in PNT2 cells while the reverse occurs if it is overexpressed in MDA-MB-231 cells [93]. In addition, Prof. Greg Goodall's laboratory has recently examined the effects of another splicing regulator, QKI, on EMT. They found that it was upregulated during EMT and that its overexpression changed the isoform proportions of hundreds of coding and noncoding transcripts [128].

Alternative splicing is also frequently deregulated in cancer with mutations of intronic splice site of tumour suppressor genes often causing exon-skipping events that truncate proteins in a similar fashion to classical nonsense mutations [129, 130]. In addition, there are well documented examples of an isoform switch between two functional proteins with different roles. Loss of ESRP1 regulates CD44 isoform switch during EMT where the epithelial promoting CD44v changes to the mesenchymal promoting CD44s [131]. This combination of factors raised the questions of whether ZEB1 had isoforms and, if so, how they were regulated during EMT.

4.1.2 ZEB1 isoforms

Very little is known about ZEB1 isoforms; except for the study by Manavella *et al* [132]. They identified a full length (FLI) and a short isoform (SI) in rat cell lines and studied their effect on the ZEB1 promoter. They found that both isoforms could bind to the ZEB1

promoter. The FLI repressed transcription in a negative feedback loop; however, the SI had no direct effect on ZEB1 transcription. Instead, the SI acted as a dominant negative isoform inhibiting FLI activity. It remains unclear whether these isoforms are present in human as there are no isoforms matching their description in refseq or in our own RNA-seq data. There are, however, four human ZEB1 isoforms represented in refseq. Importantly, these have not previously been described in the literature. Based on the importance of isoform switching in many contexts, including cancer and EMT, I used the previously described HMLE-mesHMLE EMT model (3.1.2) to investigate whether ZEB1 underwent an isoform switch during EMT or if it was primarily regulated at the transcription level.

4.1.3 Chapter outline

In this Chapter I outline my research on ZEB1 isoforms during EMT. As previously discussed (Chapter 3), I was interested in ZEB1 because it is a bivalently regulated gene in my HMLE-mesHMLE EMT model and it has established links with EMT. To obtain insights into ZEB1 function, as well as for confidence in my choice of ORF for subsequent cloning work, I analysed ZEB1 isoform expression throughout EMT. My approach was to induce an EMT in HMLE cells using TGFβ and collect RNA at multiple time points throughout the transition. I used qRT-PCR to determine whether there was a ZEB1 isoform switch during EMT.

4.2 Results

4.2.1 ZEB1 isoforms - mRNA and protein differences

Using refseq I identified four ZEB1 isoforms. These are labelled Isoform 1 (I1), I2, I6 and I8. I2 is the canonical full length isoform typically used in the literature. I1 has a different, longer 5`UTR, but still contains a ribosome binding site and the elements required for translation. I6 excludes exon 2 and I8 excludes exon 3 (Figure 16). Importantly, none of these isoforms result in a frame shift and the excluded exons are all small. The 5` terminal zinc fingers are in exon 4 so we do not expect these splice variants to change ZEB1 DNA binding; however, this cannot be ruled out.

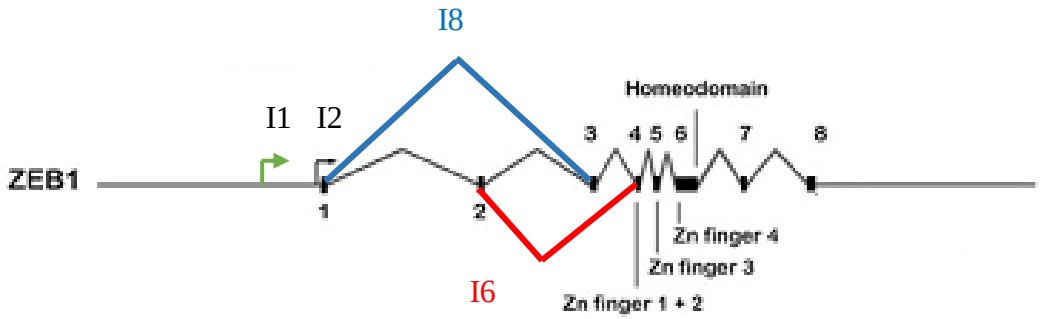


Figure 16 Schematic of the different ZEB1 isoforms.

The I1 TSS is indicated in green, the shared TSS is indicated in black. I2 includes all exons. Alternative splicing events for I6 (red) and I8 (blue) are specified.

A possible effect of the inclusion or exclusion of these exons was to affect pI which could affect ZEB1 properties. To determine whether pI differed across the isoforms I determined the overall charge of exons 2 and 3 and what effect they had on full length ZEB1 using ExPASy [133]. Exon 2 has a pI of 4.51, Exon 3 has a PI of 3.91 and full length ZEB1 has a pI of 4.87. The exclusion of Exon 2 or Exon 3 resulted in a change in pI of 0.03 or 0.17 respectively. A modest change was expected as both exons comprise only ~2% of full length ZEB1. Based on these observations I did not expect the change in isoform composition to

elicit a dramatic change in ZEB1 function, though the potential for differential expression remained of interest. If there was an isoform switch, despite the evidence suggesting that the isoforms are similar, this would imply that the alternatively spliced exons contained domains with unrecognised functions, or affected ZEB1 structure in a way that was important for its activity.

To determine whether there was an isoform switch I first evaluated the feasibility of using antibodies to differentiate between the ZEB1 isoforms. However, all commercial ZEB1 antibodies are panZEB1, as the isoforms are not prevalent in the literature. Since there is little difference in isoform MW, differentiating between isoforms through Western blot was unlikely to be feasible. Supporting this, I had not previously observed a multiplicity of bands in ZEB1 Western blots further suggesting that these antibodies could not differentiate between the isoforms. To circumvent this issue, I designed isoform specific primers for each isoform by having one of each primer pair span the isoform specific region. I then used qRT-PCR to look at the mRNA abundance across EMT.

4.2.2 ZEB1 isoforms are not differentially regulated during EMT

To determine whether the ZEB1 isoforms were differentially regulated during EMT I induced an EMT in the HMLE cells with Weinberg medium (2.1.1) and TGFβ. I then harvested RNA at D0, D8, D12 and D18. By D18 the cells have undergone a complete EMT as measured by cadherin switch and other markers (personal communication, Philip Gregory). I also used RNA from the mesenchymal breast cancer cell line MDA-MB-231 as an additional mesenchymal control. I made cDNA from these RNAs and performed qRT-PCR in triplicate in three independent experiments using isoform specific primers (Table 5) and GAPDH as a control gene.

I saw substantial upregulation of total ZEB1 across the EMT but found little change in relative isoform expression (Figure 17). The canonical full length isoform, I2, comprised at least 80% of total ZEB1 at every time point tested. As the isoforms were not differentially regulated during EMT, this suggests that expression of specific ZEB1 isoforms are not important in this EMT model. Consequently, I did not further investigate ZEB1 isoforms, although it remains possible that these isoforms are differentially regulated in a different EMT model or in an entirely different context. In addition, when looking at mRNA levels we cannot ignore the possibility that the protein levels are independent of the mRNA levels [134, 135].

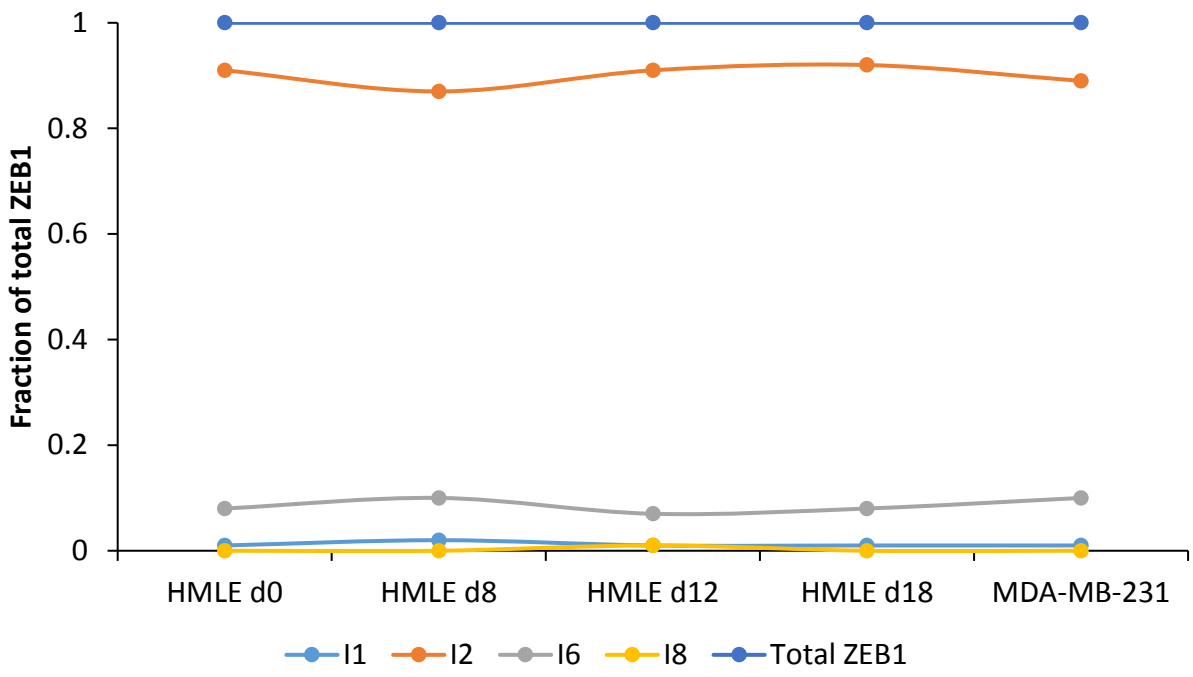


Figure 17 Relative expression of the four ZEB1 isoforms across an EMT time course in HMLE cells and in MDA-MB-231 cells.

qRT-PCRs were performed using isoform specific primers for each isoform. Isoform expression was normalised to GAPDH and is expressed as a fraction of total ZEB1 expression.

4.3 Conclusion: ZEB1 isoforms are not specifically regulated during

In this chapter I investigated the role of ZEB1 isoforms in EMT based on the strong evidence that isoform switching can promote either an EMT or an MET [131]. I used refseq to identify four ZEB1 isoforms (4.2.1) and characterised these in terms of potential effect on ZEB1 function. I then used the HMLE-mesHMLE EMT model and qRT-PCR to test whether they were differentially regulated during EMT (4.2.2). I did not detect any significant change in isoform expression during this EMT process. This suggests that an isoform switch does not contribute to the effect of ZEB1 on EMT. Rather, it is the overall increase in ZEB1 expression, and the negative feedback loop with the miR-200 family, that is crucial for ZEB1's effect on EMT [30, 31]. Nevertheless, there is a possibility that these isoforms are important in a different context, or indeed in this context, due to the high expression level of isoform 6. An important possibility to further explore ZEB1 isoform expression in EMT is to raise isoform specific antibodies. This would help conclusively determine the relative expression of each isoform across EMT. Before these experiments, it would be necessary to effectively optimise the antibodies to determine their binding efficiency.

Chapter 5 Identifying novel ZEB1 targets in EMT

5.1 Introduction

This Chapter outlines my research identifying and characterising novel ZEB1 targets. As discussed in Chapter 3, ZEB1 was of interest due to being identified as a bivalently regulated gene in the HMLE-mesHMLE EMT model (3.1.3). I first evaluated the epigenetic signature at the ZEB1 promoter using ChIP-reChIP. After confirming that ZEB1was bivalently regulated in EMT, I identified and characterised novel ZEB1 targets. My initial approach was to perform ZEB1-ChIP-seq but this proved unfeasible due to antibody limitations. To obtain a list of potential target genes I analysed publicly available data from the ENCODE project.

I chose 26 potential target genes based on a set of stringent criteria and tested the effect of modulating ZEB1 expression on their expression with a variety of cell lines and assays. From this list, I successfully identified two novel ZEB1 target genes.

5.2 Results

5.2.1 Validation of ZEB1 promoter bivalence

As with RASA3, PLEKHO1 and ADM2 in section 3.2.2, here I tested how the ZEB1 epigenetic signature changed in HMLE cells during EMT. I first used the ChIP-reChIP chromatin prepared in our laboratory by Andrew Bert to confirm that ZEB1 underwent a bivalent to active epigenetic change during HMLE EMT. These data were normalised to the same GAPDH, ACTBN, PAX5 and MYOD promoter average as previously (Figure 18 A). In addition, I analysed the prevalence of individual epigenetic marks (Figure 18 B and C).

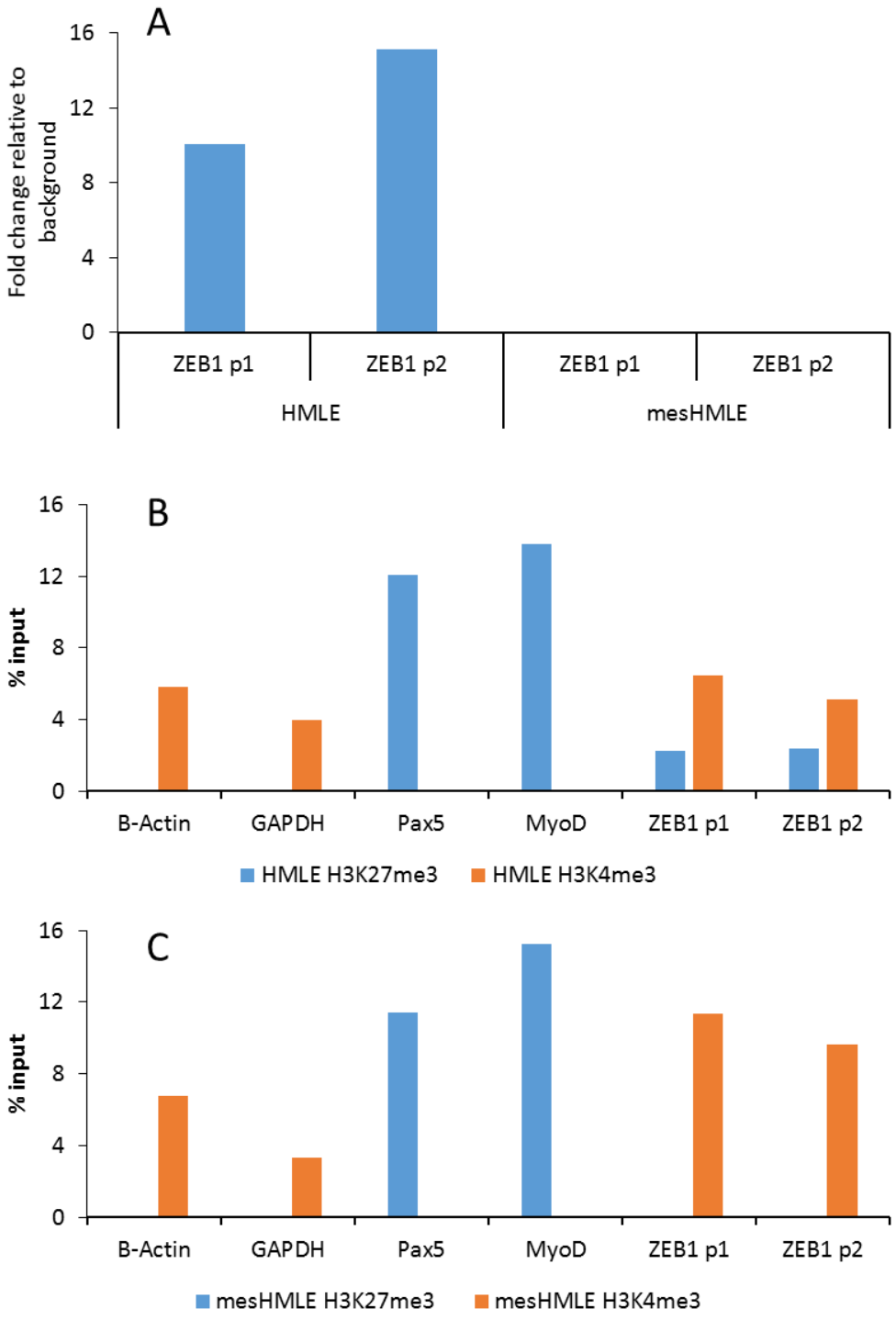


Figure 18 Enrichment levels of bivalent epigenetic marks for ZEB1 in HMLE and mesHMLE cells.

ChIP experiments were performed on $1x10^6$ cells for each experiment. All columns have values, for those without visible bars the enrichment was virtually zero. (A) Relative enrichment of H3K27me3/H3K4me3 ChIP-reChIP at two ZEB1 promoters relative to the average enrichment at B-Actin, GAPDH, PAX5 and MYOD promoters. (B, C) Input levels of H3K4me3 (orange) and H3K27me3 (blue) at two ZEB1 promoters and the four control promoters in αHMLE (B) and mesHMLE (C) cells.

In HMLE cells, ZEB1 had a strong bivalent signature 10-15 fold higher than the control average. In mesHMLE cells this was completely lost (Figure 18 A). This loss of bivalence was easily visualised through individual H3K4me3 and H3K27me3 ChIPs in HMLE (Figure 18 B) and mesHMLE (Figure 18 C) cells. Through the EMT transition the repressive H3K27me3 mark was lost and the activating H3K4me3 mark was enhanced. This combination of data strongly supports the initial ChIP-seq based observation that ZEB1 is a genuine bivalent gene.

5.2.2 ZEB1 ChIP-seq attempts

To identify new ZEB1 target genes I first attempted to optimise our ChIP protocols for ChIP-sequencing by changing crosslinking conditions, antibody and bead incubation time, sonication conditions and trialling multiple antibodies and concentrations. Despite this, I only achieved a maximum enrichment of ~0.4% of input and ~4-fold increase over background at the strongest ZEB1 sites and substantially lower at other validated targets (Figure 19). This level of enrichment is not sufficient for a ChIP-seq experiment [136]. To ensure that the ChIP procedure was working successfully I carried out a ChIP-qRT-PCR experiment against H3 alongside the ZEB1 ChIP (Figure 19). The high levels of enrichment with the H3 ChIP showed that the overall ChIP procedure worked successfully. Hence, it was likely that the primary cause for the low level of ZEB1 enrichment was linked to the antibody used.

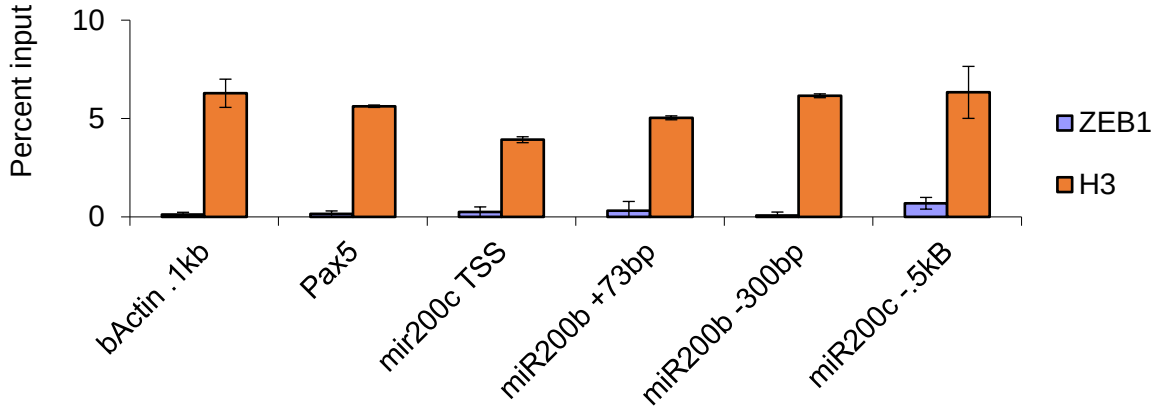


Figure 19 Comparison of percent input for H3 ChIP and ZEB1 ChIP.

qRT-PCR of ZEB1 (blue) and H3 (orange) ChIP across a range of positive and negative ZEB1 targets. Error bars represent ±SD of three technical replicates.

To circumvent the antibody problem I made a FLAG tagged ZEB1 construct. The goal was to make a stable inducible cell line in HMLE cells and to then ChIP against the epitope tag. As our lab had previously attempted constitutive overexpression of ZEB1 and found that this was rapidly supressed, I made an inducible vector.

I designed two sets of PCR primers that would amplify ZEB1 cDNA while, sequentially, mutating the stop codon, inserting a C-terminal Flag tag and adding Gateway attL recombination sites. I used these primers to amplify Flag tagged ZEB1 from MDA-MB-231 cDNA and clone this into a Gateway destination lentiviral doxycycline inducible vector with blasticidin resistance, plVTPTIP, kindly provided by the Whitelaw group (University of Adelaide).

Briefly, a doxycycline inducible system involves a tetracycline dependent promoter (P_{TRE3G}) and a reverse tetracycline-controlled transactivator (TRE3G). In the absence of doxycycline the transactivator will not bind to the promoter region resulting in a lack of transcription. The addition of doxycycline induces a conformational change in the transactivator leading to active transcription. These two components, P_{TRE3G} and TRE3G, are typically encoded in different vectors and co-transfected. A potential complication from this process is that it is unlikely that cells will be transfected at a 1:1 ratio. This causes different cells within a population to react differently to doxycycline induction. Incorporating both components in the same plasmid circumvents this issue. This increases the probability of a 1:1 relationship between these components in any given cell and eliminates the requirement for dual antibiotic selection. These properties significantly facilitate the production of single clones and increases the homogeneity of response in a mixed population.

I transfected 293T cells with plVTPTIP_ZEB1_FLAG and constructs coding for the viral proteins Gag, Pol and Rev for lentiviral production. I transduced HMLE cells to maintain continuity with the original ChIP-seq study where I identified ZEB1 as a bivalent gene (3.1.3). Prior to transducing the HMLE cells, I harvested protein from the transfected HEK293ts and used SDS-PAGE to determine level of ZEB1 overexpression using FLAG. I detected strong expression of a FLAG tagged protein at the correct size for ZEB1 in my ZEB1 transduced cells and not in the EV control validating my construct and transfection (Figure 20).

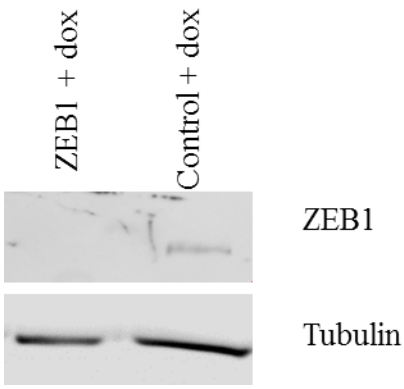


Figure 20 Western blot showing ZEB1 expression in 293t cells used to prepare virus compared to the EV control.

Cells were induced with 500ng/ml doxycycline for 24h. Tubulin was used as a loading control.

Prior to the transduction I carried out a kill curve experiment on HMLE cells using blasticidin concentrations ranging from 0 to 1µg/ml; 0.5µg/ml was lethal within four days (data not shown). This rate was used for blasticidin selection on the transduced HMLE cells.

There are two primary approaches to producing stables, pools and clones. A stable pool is a mixed population of antibiotic resistant cells while clones are derived from a single cell. Stable clones are derived by isolating a single cell from the population and expanding it. Individual clones can be isolated mechanically, e.g. using a cloning ring, or through limiting dilution where the population is diluted so that less than a single cell will be seeded per well. Both stable pools and individual clones have benefits and drawbacks. A pool is a mixed population; this decreases the probability of reporting on clone-specific effects. However, it can be harder to isolate an effect as different cells within the population are likely to respond differently to the same stimuli, obscuring the results. Individual stable clones have the opposite problems where it is often easier to observe an effect but it is not possible to determine whether this is a clone-specific effect. To avoid the limitations of these

approaches, I examined multiple stable clones. This allows one to assess the effects of the transduction in multiple independent populations greatly decreasing the probability that the observation is due to clone-specific effects.

After the blasticidin selection I used limiting dilution to seed a 96 well plate with an expected concentration of half a cell per well to reduce the probability of seeding multiple cells per well. From the 96 wells seeded, I expanded 29 successfully; of the remaining 65 wells, some had no cells or did not expand successfully. I induced ZEB1 expression in these 29 clones using 1µg/ml of doxycycline for 24h and tested their response through qRT-PCR using primers specific against ZEB1-FLAG (Figure 21). Of the 29 clones tested, clones 7, 22 and 29 responded strongly to doxycycline and had low background expression. Of note, none of these clones provided a completely clean system; low levels of ZEB1 were ectopically expressed in each clone without doxycycline induction. This low level of ZEB1 expression was sufficient to induce a partial EMT in the three clones. A consequence of this partial EMT was the expression of endogenous ZEB1 which led to further complications.

While the cells were grown in HMLE media, they formed very long neuron-like structures with limited cell-cell contacts (Figure 22). When the media was changed to Weinberg media, they underwent a dramatic and rapid morphological change to strongly resemble mesHMLE cells (refer to 3.1.2 for details on the HMLE EMT model). Importantly, these changes occurred within 24 hours excluding the possibility that the dramatic changes were due to the change in media. As a consequence of this partial EMT, exogenous ZEB1 was silenced in all cells (Figure 22) resulting in no net gain to my capacity to ChIP ZEB1. I also measured CDH1 expression for an indication of EMT status. CDH1 was lost after induction in all clones, representative data of this is shown for clone 22 (Figure 22).

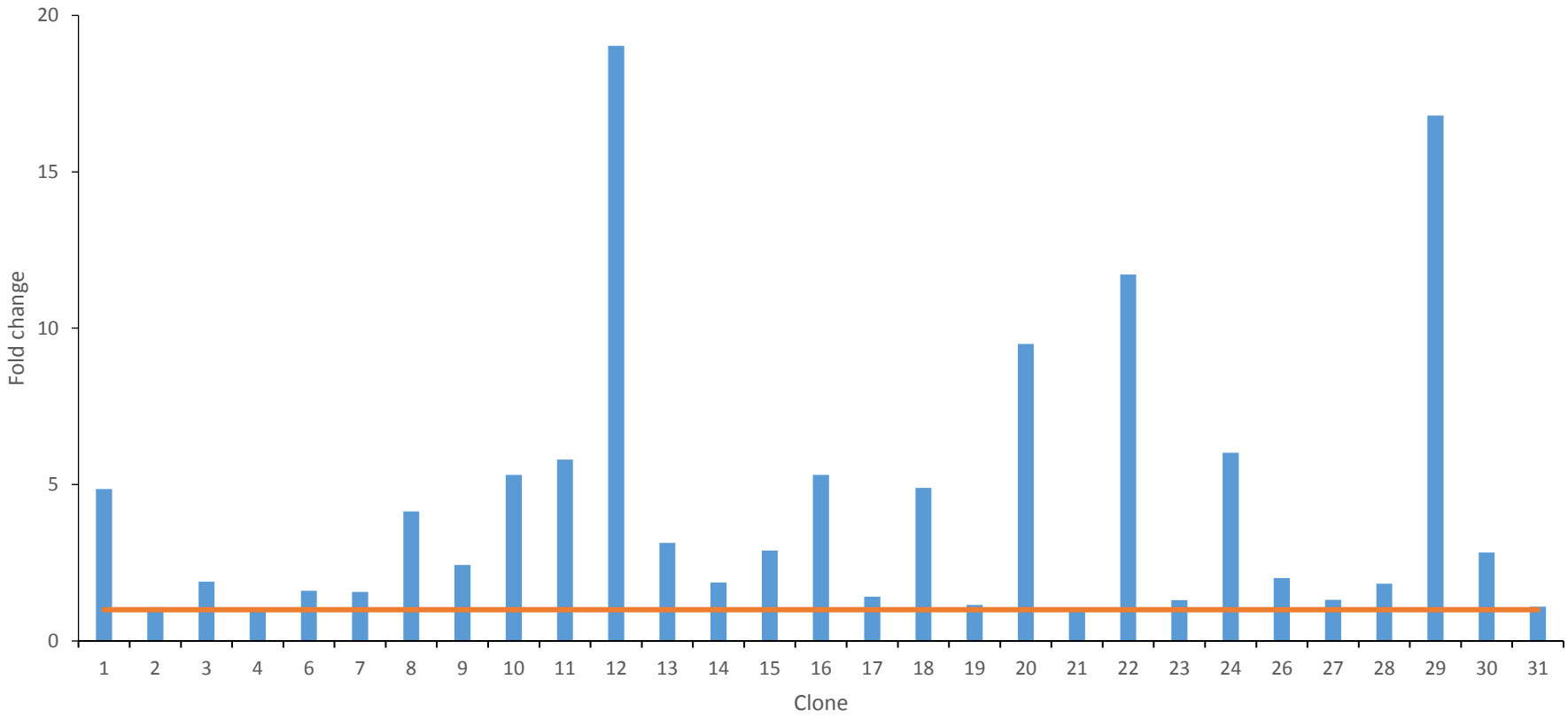
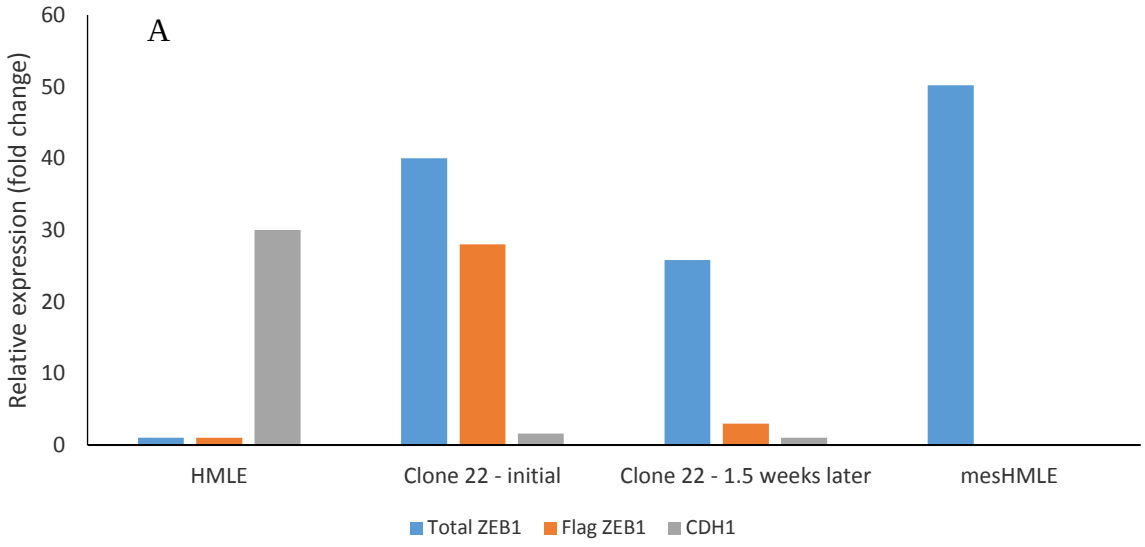


Figure 21 Characterisation of ZEB1 expression in 29 stable clones after doxycycline induction.

Change in ZEB1 expression as measured by qRT-PCR. Data are mean of 3 technical replicates and represent the fold change in ZEB1 expression after doxycycline induction compared to un-induced levels. In each case, expression was first normalised to GAPDH. Orange line represents un-induced level.



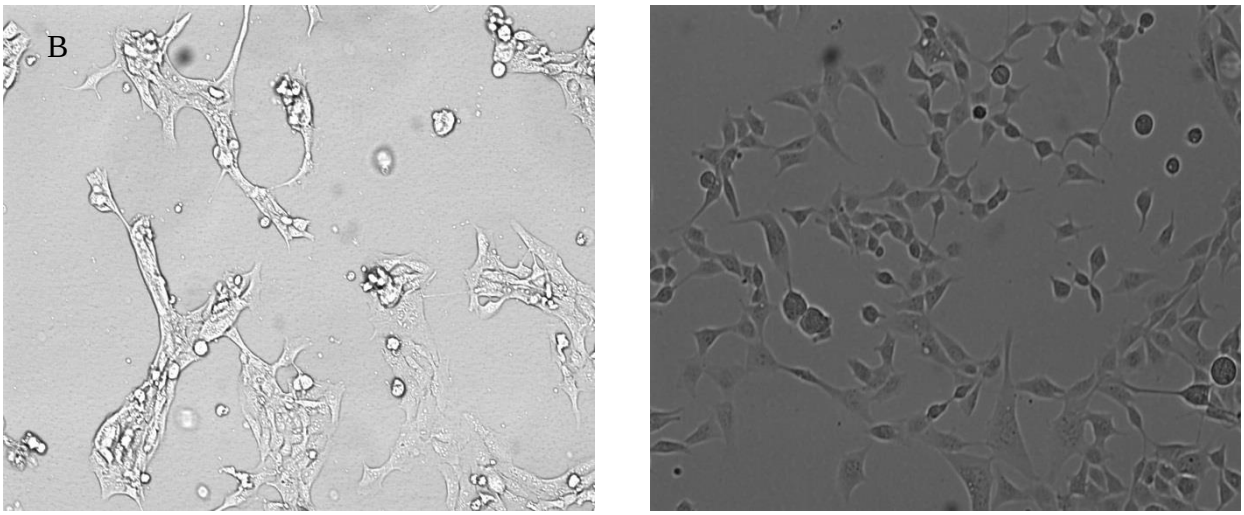


Figure 22 Changes in clone 22 expression and morphology over time

(A) qRT-PCR of mesHMLE, clone 22 1.5 weeks after initial measurement, clone 22 at original measurement and parental HMLE cells. mRNA levels were normalised using GAPDH. ZEB1 and CDH1 levels were normalised to HMLE cells and FLAG levels were normalised to un-induced clone 22. (B) Representative images of clone 22 1.5 weeks after initial measurement in Humec media (left) or Weinberg media (right).

5.2.3 Analysing ENCODE ZEB1 ChIP-seq data

Given the difficulties with performing ZEB1 ChIP-seq (5.2.2) and the problems encountered while making a stable clone (5.2.2), and the lodging of ZEB1 ChIP-seq data as part of the ENCODE project [137], I decided to analyse the publically available data. The ENCODE ZEB1 ChIP-seq data came from a liver (HEPG2) and a B cell line (GM12878) and is available at http://genome.ucsc.edu/cgi-bin/hgFileUi?db=hg19&g=wgEncodeHaibTfbs.

5.2.3.1 Quality control and removing adapters

We used FastQC to assess read quality and number, trim adapters and remove reads that did not meet the quality or length requirements. For HEPG2 this worked efficiently and we easily identified, and removed, the adapter as the most over-represented sequence from the FastQC (Table 16).

Table 16 Summary of the processing and FastQC analysis for the HEPG2 runs

a: percent of total reads; b: percent of processed reads

Run	Trimmed reads	Reads trimmed due to quality (% of total)	Reads trimmed due to length, too short	Reads remaining
ZEB1Rep1	3,456,678 (11.9) [a]	5.96	2,218,520 (7.7) [b]	26,727,568
ZEB1Rep2	4,236,867 (15.1%)	4.56	1,824,247 (6.5)	26,147,746
Control2	3,347,769 (9.9%)	5.32	1,616,968 (4.8)	32,324,924
Control3	5,067,392 (11.1%)	3.80	1,520,547 (3.3)	44,020,652

For GM12878 we encountered significant problems with both the control and experimental data. Many reads were trimmed from the experimental repeats due to either length or quality; however this still left us with at least half the amount of reads as for the HEPG2 samples as there were many more initial reads. This was sufficient for further analysis.

For the controls, there were additional issues. We identified five separate runs merged into the control data. We split these into four that we successfully identified using their read ID and a final combined group of unknown reads (Table 17). We then identified the adapter for each of these runs, removed the adapters and remerged them. After, we went on to map the reads to the human genome (HG19).

Table 17 Summary of the processing and FastQC analysis for the GM12878 runs.

The GM12878 control had some oddities, we identified that it was a combination of multiple runs and separated these accordingly, they are represented individually in the table.

a: percent of total reads; b: percent of processed reads

Run	Number of reads trimmed	Quality trimmed (% of total)	Number of too short reads	Reads remaining	Quality of fastqc after trimming
ZEB1					
Rep1	1,827,096 (8.8)[a]	40.2	9,019,652 (43.2) [b]	11,848,425	Good
Rep2	2,679,784 (9.9)	18.5	5,542,761 (20.5)	21,513,406	Good
Control (RXL)					
HWI-EAS295_4	830,955 (4.4)	0.4	11,657,970 (62.1)	7,108,819	Low
USI-EAS45_5	472,888 (5.3)	21.3	1,633,025 (18.2)	7,361,394	Good
HWI-EAS149_3	1,445,612 (7.9)	20.6	3,774,197 (20.6)	14,567,329	Good
HWI-EAS149_4	1,434,109 (7.2)	23.8	4,699,664 (23.6)	15,247,543	Good
Unknown	432,281 (3.0)	27.4	4,034,865 (28.0)	10,380,870	Low
Total				37,176,266	

5.2.3.2 Mapping reads and removing duplicates

We mapped the reads and converted the file from bwa to bam using command line and fastQ (Table 18).

Table 18 Summary of mapped reads that met the quality threshold for each run for both cell lines.

RXL represents reverse cross linking, that is, the control runs.

Run	Total reads	Mapped (%)
GM12878		
Zeb1Rep1	11,848,425	10,558,779 (89.12)
Zeb1Rep2	21,513,406	20,107,041 (93.46)
RXL lRep4-USI-EAS45	7,361,394	6,668,553 (90.59)
RXL Rep4-HWI-EAS295	7,108,819	4,103,058 (57.72)
RXL unknown	10,380,870	5,148,664 (49.60)
Rxl	37,176,266	33,379,540 (89.79)
HEPG2		
Zeb1Rep1	26,727,568	25,266,321 (94.53)
Zeb1Rep2	26,147,746	23,332,156 (89.23)
Rxl Rep2	32,324,924	31,448,837 (97.29)
Rxl Rep3	44,020,652	43,515,674 (98.85)

The majority of the runs had ~90% map rate while two of the GM12878 control runs had a low mapping rate of ~50%. Despite this, we had at least $2x10^7$ reads for every run, apart from one of the GM12878 runs.

Duplicates are traditionally removed to control for potential PCR bias. We used Picard for this (https://github.com/broadinstitute/picard). Most of the runs had <10% duplicates while the GM12878 experimental repeats had ~40% duplication (Table 19).

Table 19 Summary of duplicates in mapped reads identified and removed using PICARD.

Run	Unpaired reads examined	Unmapped reads	Unpaired read duplicates	Percent duplication
GM12878				
Rxl	33,379,540	3,796,726	1,222,306	3.7
Zeb1 Rep1	10,558,779	1,289,646	4,259,101	40.3
Zeb1 Rep2	20,107,041	1,406,365	8,939,496	44.5
HEPG2				
Rxl Rep2	31,448,837	876,087	3,144,989	10.0
Rxl Rep3	43,515,674	504,978	4,453,529	10.2
Zeb1 Rep1	25,266,321	1,461,247	2,047,334	8.103
Zeb1 Rep2	23,332,156	2,815,590	3,479,007	14.9

5.2.3.3 Calling ZEB1 peaks and ZEB1 motif analysis

We used MACS [86] to identify ZEB1 binding sites. We determined which peaks were linked to a specific gene based on their genomic location. If the peak was within -2.5 to + 0.5kb of a gene TSS, then we linked it to that gene. Using this method we identified 3674 potential ZEB1 targets in GM12878 and 658 genes in HEPG2; this five-fold difference can partially be explained by the read distribution of the negative control for HEPG2. This control had high read density near TSS, potentially obscuring valid target genes.

Two observations from our analysis support the overall reliability of these datasets. Firstly, there was a 60% overlap between the two gene lists. Secondly, after trimming the peaks to ±150bp from the peak summit, the established ZEB1 motif, the Z-box [57, 138, 139], was identified as the most frequent motif using MEME-ChIP [140]; this is the consensus site referred to hereafter.

From these data I first examined the relationship between ZEB1 motifs and score (Figure 23 A). There was a very strong, statistically significant ($p<0.0005$) positive relationship between these two factors. I also identified the distribution of ZEB1 consensus sites in both cell lines.

Relatively, GM12878 had 30% the amount of peaks with two consensus sites and twice as many peaks with no consensus sites compared to HEPG2 (Figure 23 B). However, this overlooks the possibility of low quality peaks or false positives. When only the top 300 ranked genes for each cell line were considered, less than 1% of the peaks contained no consensus sites and there were only marginal differences in motif distribution between HEPG2 and GM12878 (Figure 23 B). This, in addition to the correlation between ZEB1 motif and score strongly supports the notion that a ZEB1 consensus site in a peak increases ZEB1 binding. This further suggests that the differences in ZEB1 motif distribution observed between the complete gene lists are due to the aforementioned differences in read quality.

This provided me with two possible explanations for the large number of lower quality ZEB1 bound promoters with no ZEB1 motifs. Either we had identified a large number of false positives, or these sites were bound by ZEB1 at non-canonical sites and this resulted in lower levels of enrichment. The latter hypothesis accounts for both observations making it the likely explanation.

Interestingly, the large number of peaks that contained a single ZEB1 site suggest that, contrary to what is popularly expressed in the literature [58], ZEB1 does not require a paired binding site to interact with DNA.

In addition to examining the relationship between ZEB1 motifs and peaks I analysed the relationship between ZEB1 motifs. Our data confirmed that the spacing between consensus sites can vary extensively with a range of 6-180bp between ZEB1 sites (Figure 23 C). However, for over 70% of paired sites the distance was between 6 and 50bp (Figure 23 C), consistent with earlier studies [57].

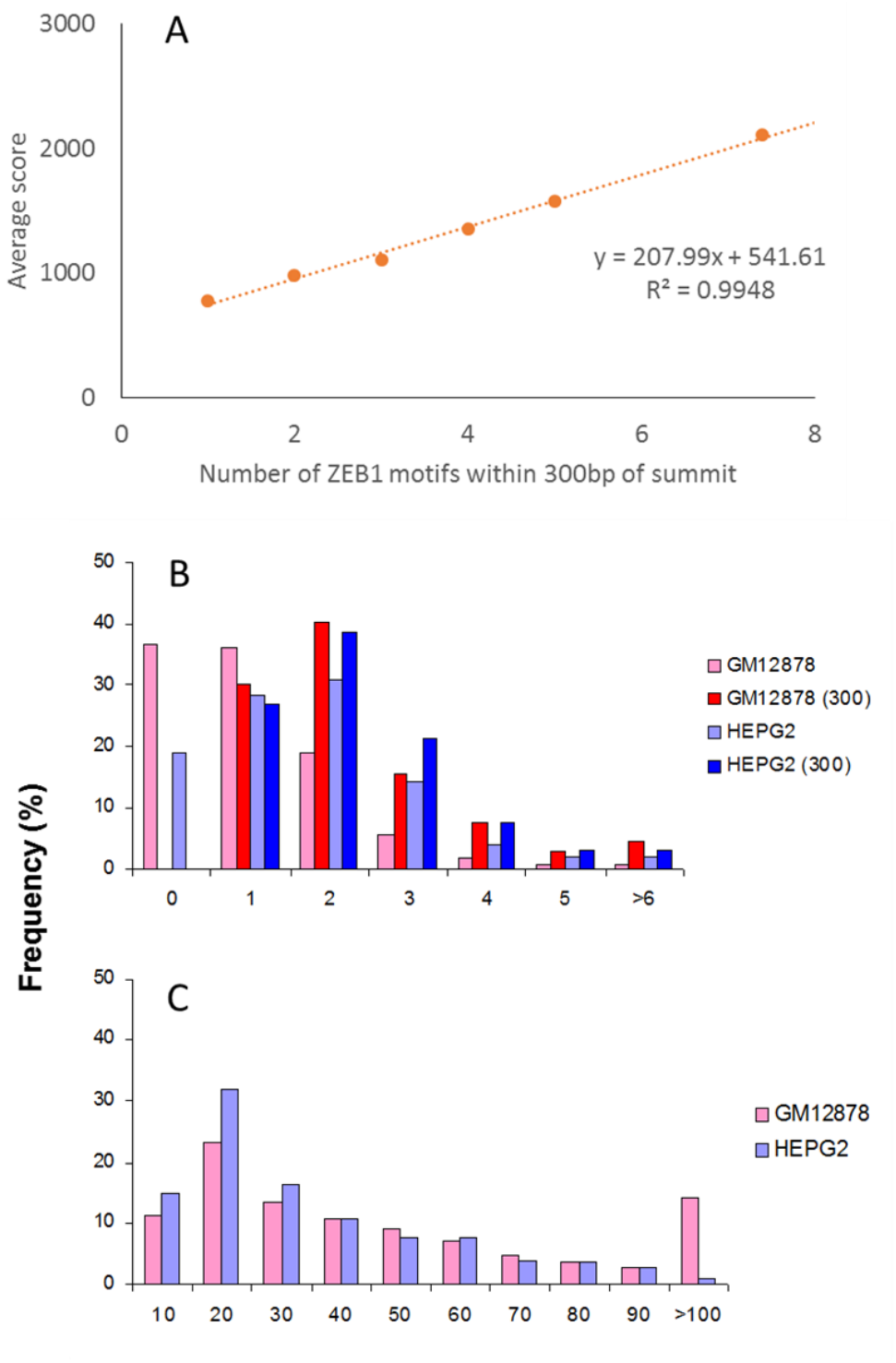


Figure 23. Summary of observations on ZEB1 binding from ChIP-seq analysis.

(A) Relationship between average score and number of ZEB1 motifs. As there were only 19 genes with >6 motifs I calculated the weighted average for motifs between 6-13 (7.4). (B) Frequency distribution of ZEB1 consensus site (CAGGTA/G) in peaks for all genes in GM12878 and HEPG2 and the top 300 of these. (C) Distance between ZEB1 consensus sites in GM12878 and HEPG2 cell lines. Both sets of data are derived from ChIP-Seq. experiments in GM12878 and HEPG2 cell lines performed by Myers R and Pauli F [137].

5.2.4 Potential ZEB1 target genes

To identify novel ZEB1 target genes I designed a list of criteria to maximise the probability that they would be genuine targets. Candidate genes were required to: have at least a single ZEB1 site, be in the top 100 ranked genes for at least one list, not be established ZEB1 targets, have roles in areas that were associated with EMT, such as migration, proliferation or cytoskeleton remodelling and have a negative correlation with ZEB1 in the NCI-60 data set.

I chose 17 genes that met this criteria as well as 9 that did not meet all of these criteria but were established genes important for EMT, cancer or both and were present within the top 300 ranked genes of at least one cell line. This list included PTEN and BRCA1 (Table 20). The reasoning behind these nine lower-probability genes was that if there was a direct link with ZEB1 that would be of interest. Alongside these 26 genes I chose three validated ZEB1 targets as positive controls, CRB3, EPCAM and LLGL2.

Table 20 Summary of ChIP-seq data for candidate ZEB1 target genes separated into high probability candidates and high interest candidates, and positive controls.

Candidate genes	GM12878			HEPG2			Pubmed search results in combination with		
	ZEB1 sites	Rank	Distance from TSS	ZEB1 sites	Rank	Distance from TSS	Cancer	EMT	ZEB1
High probability									
ANKS1A	3	18	-124	3	71	-98	2	1	0
AP3B1	3	57	-1151	3	22	-1139	8	3	0
CXCR3	1	105	-4191		647	#N/A	309	113	0
DEAF1	7	7	1397	10	64	1350	8	2	0
DFFB	2	11	57	2	367	93	24	0	0
ETV7	2	9	266	4	100	279	9	0	0
F11R	2	19	39	2	27	11	16	49	0
FLNA	2	139	870	2	16	888	41	10	0
GAB2	2	100		3	128	386	98	21	1
GGA1	1	26	-1800	1	46	-1798	4	0	0
GSK3A	4	27	-506	4	40	-508	3	0	0
INADL	3	160	21	4	14	14	2	24	0
POLR1D	2	31	791	2	56	850	1	0	0
SPINT1	2	23	181	2	69	177	70	46	1
SYK	1	65	195		637	195	562	85	0
TAF3	3	3	-345	3	31	-311	3	0	0
WTAP	2	58	-160	3	409	-175	36	7	0
High interest									
BAD	1	1535		0	4708	374	0	-643	
BRCA1	1	2688		0	8562	629	1	-90	
CTNNB1		1871		0	7983	3138	43	2	
HIF1A	1	1446		0	3146	498	6	-91	

KRAS	1	3157		0	5228	539	4	-40	
MYC	1	1771		0	15595	1681	16	-1976	
PTEN	2	251	3	340	5045	555	4	-306	-284
BCL2	3	46		0	2457	251	4	1279	
MAK	1	555		0	1043	132	0	135	
Positive controls									
CRB3	4	218	-71	3	11	-46	31	46	3
EPCAM	1	1488	202	1	15	99	831	590	11
LLGL2	2	44	-278	4	4	-252	4	2	0

I subjected the candidate genes to a battery of assays designed to test whether these genes were genuine ZEB1 targets. These assays were performed in a sequential order based on ease and cost. First, I tested whether modulating ZEB1 expression changed their expression at the RNA level. Second, I tested whether ZEB1 affected the promoter directly through luciferase assays. Finally, I performed ChIP-qRT-PCR to verify ZEB1 binding at the promoter.

5.2.5 Validation of potential ZEB1 targets – qRT-PCR

I first evaluated whether modulating ZEB1 expression levels affected expression of the potential target genes through qRT-PCR. To modulate ZEB1 expression I used siRNA knockdown in two mesenchymal cell lines, MDA-MB-231 and mesHMLE, and transient overexpression in two epithelial lines, HMLE and MCF10A. I will first describe the siRNA experiments.

I confirmed ZEB1 knockdown through Western Blot (Figure 24 A) and qRT-PCR (Figure 24 B). I performed the siRNA experiments in triplicate and tested for statistical significance using ANOVA followed by Fisher's PLSD. Of the 26 candidate genes two, F11R and INADL, had a significant negative correlation with ZEB1 expression in both cell lines (Figure 24 C). ANKS1A, CTNNB1, ETV7, MAK, POLR1D and DEAF1 had a significant negative correlation with ZEB1 expression in one of the two cell lines. The remaining 18 genes had no significant correlation with ZEB1.

To supplement the knockdown experiments I used transient ZEB1 over expression in two epithelial cell lines, MCF10A and HMLE cells. Expression was confirmed using Western Blot (Figure 25 A) and qRT-PCR (Figure 25 B).

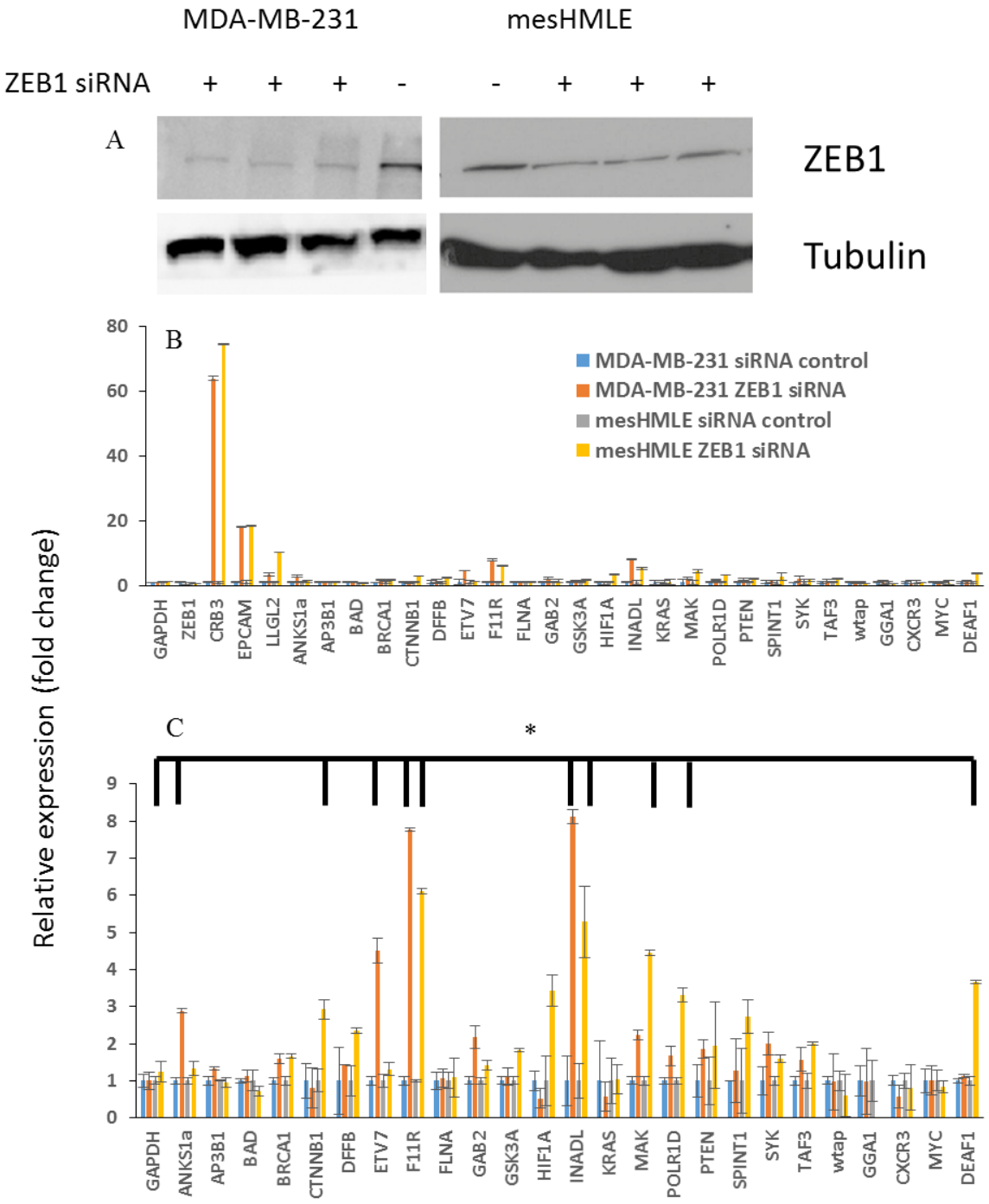


Figure 24 Effect of ZEB1 knockdown on candidate gene expression in MDA-MB-231 and mesHMLE cells.

(A) Western blots showing changes in ZEB1 expression in MDA-MB-231 and mesHMLE cells transiently transfected with ZEB1 siRNA or a negative control siRNA. Alpha tubulin was used as a loading control. (B, C) qRT-PCR of 26 candidate genes with (B) or without (C) the three controls relative to the appropriate siRNA control and normalised to GAPDH. Error bars represent ±SD of three biological replicates. Asterisk represents statistically significant difference at $p<0.05$ (Fisher's PLSD).

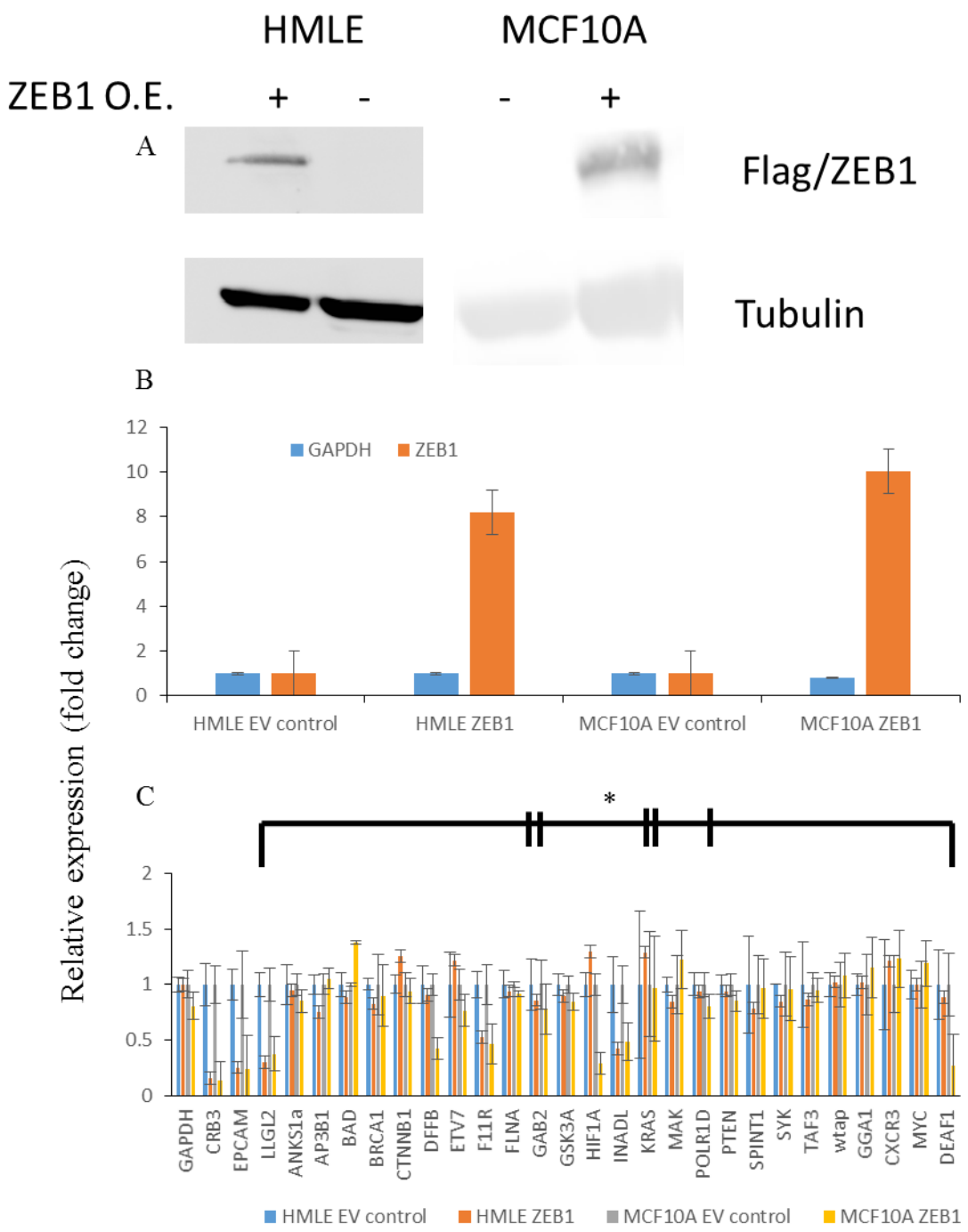


Figure 25 Effect of ZEB1 overexpression on candidate gene expression in HMLE and MCF10A cells.

(A) Western Blots showing changes in ZEB1 expression in transiently transfected HMLE and MCF10A cells. Alpha tubulin was used as a loading control. Flag antibody was used to detect ZEB1. (B) qRT-PCR of ZEB1 normalised to GAPDH and the empty vector control. (C) qRT-PCR of 26 candidate genes relative to the appropriate empty vector control and normalised to GAPDH. Error bars represent ±SD of three biological replicates. Asterisk represents statistically significant difference ($p<0.05$, Fisher's PLSD)

Of the 26 genes tested, F11R and INADL were the only ones that negatively correlated with ZEB1 expression in both cell lines. This, paired with their strong response to ZEB1 knockdown in epithelial cells, marked them as strong candidates. Of the genes that responded in one cell line to ZEB1 knockdown, DEAF1 and MAK also had a negative correlation with ZEB1 expression in one of these mesenchymal lines. This combination of assays gave a high degree of confidence that F11R and INADL were genuine ZEB1 targets and varying degree of confidence in the other genes. From this, I decided to focus on F11R and INADL. Both of these genes are associated with tight-junctions as is the established ZEB1 target CRB3 suggesting a potential for ZEB1 in the regulation of these junctions [141]. F11R and INADL were both identified as potential ZEB1 target genes in a 2007 paper though no further research has been done on this [29].

5.2.6 Validation of potential ZEB1 targets - luciferase

I next performed luciferase assays to measure the effects of modulating ZEB1 expression on isolated gene promoters. To do this I located the promoter and ZEB1 sites for F11R and INADL, and designed primers to amplify this region incorporating restriction enzyme sites. The resulting PCR product was digested and cloned upstream of a luciferase gene in the PGL3 vector. I transfected this vector alongside the RL control vector in HMLE and MCF10A cells co-transfected with ZEB1 overexpression vector as previously described (5.2.5). Similar to the qRT-PCR results, I found a statistically significant negative correlation between ZEB1 and both F11R and INADL (Figure 26). This enhances my earlier evidence that F11R and INADL are repressed by ZEB1.

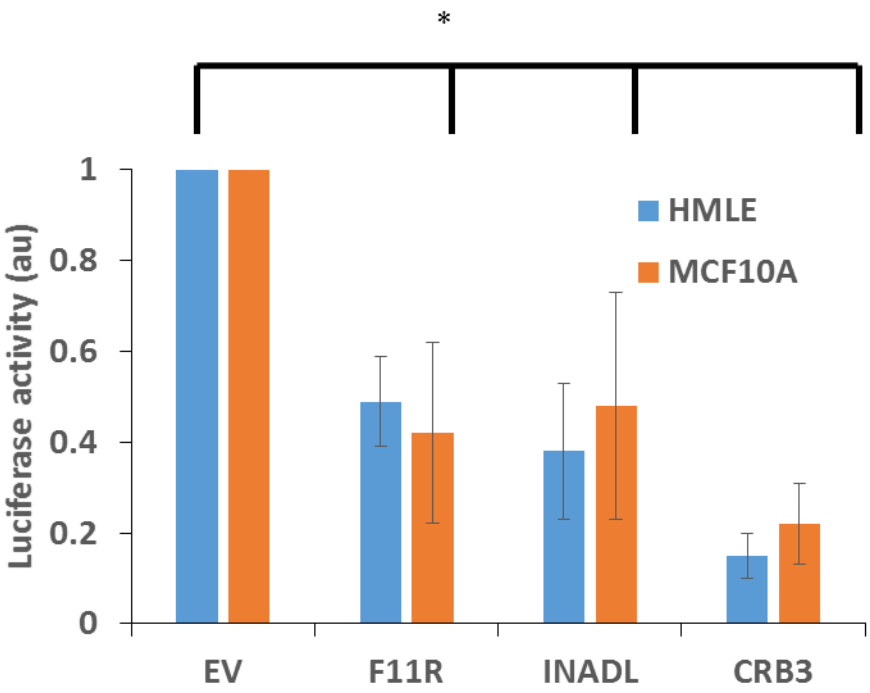


Figure 26 Effect of ZEB1 overexpression on candidate gene luciferase activity in MCF10A and HMLE cells.

Normalised luciferase readings of F11R, INADL and CRB3 promoters in MCF10A and HMLE cells transfected with ZEB1 overexpression vector or EV control normalised to the appropriate luciferase control and relative to the EV control transfection. Error bars represent ±SD of three biological replicates. Asterisk represents statistically significant difference ($p<0.05$, T-test, unpaired).

The final test I undertook to ensure that these two genes are directly regulated by ZEB1 was to confirm ZEB1 binding directly at the promoter region through ChIP-qRT-PCR.

5.2.7 Validation of ZEB1 targets – ChIP-qRT-PCR

I used MDA-MB-231 cells for my ChIP experiments with HPRT1 and GAPDH as negative controls accounting for both basal ZEB1 binding background and non-specific antibody binding. I optimised ZEB1 ChIP-qRT-PCR experiments by testing two antibodies (H102, E20X) at three concentrations. I found that both 5µl of H102 and 0.5µl of E20X gave very

similar enrichment so I carried out subsequent experiments with both. A risk with ChIP experiments is that non-specific antibody binding can confuse the results. As I used two antibodies this makes it likely that the enrichment observed is due to genuine ZEB1 binding and pulldown.

ZEB1 displayed strong binding near the ChIP-seq ZEB1 summit for both INADL and F11R. To ensure that this was not due to open chromatin, I designed primers to tile across the nearby genomic region in 300bp steps from the summit to ±900 bp (Figure 27 A and B). These tiled primers yielded a clear picture of where ZEB1 binds in this region. Near the summit, enrichment was high with levels rapidly falling off by ±300bp and enrichment at a similar level to the negative controls by ±600bp from the summit. This provides strong evidence for the ZEB1 binding site identified from the ChIP-seq data as a legitimate ZEB1 binding site. Interestingly, I saw greater clustering between two samples with different antibodies performed side by side versus the same antibody performed on separate days (Figure 27 C). This suggests that the two antibodies bound ZEB1 with a similar efficiency.

As I had an interest in bivalent genes, I analysed the HMLE-mesHMLE ChIP-seq data to determine whether F11R or INADL were bivalently regulated. In this system, both genes underwent a large reduction in expression, 4.5 fold for F11R and 7 fold for INADL (Table 21). This reduction in expression was accompanied by a loss of H3K4me3 and no gain of H3K27me3 suggesting they are not bivalently regulated.

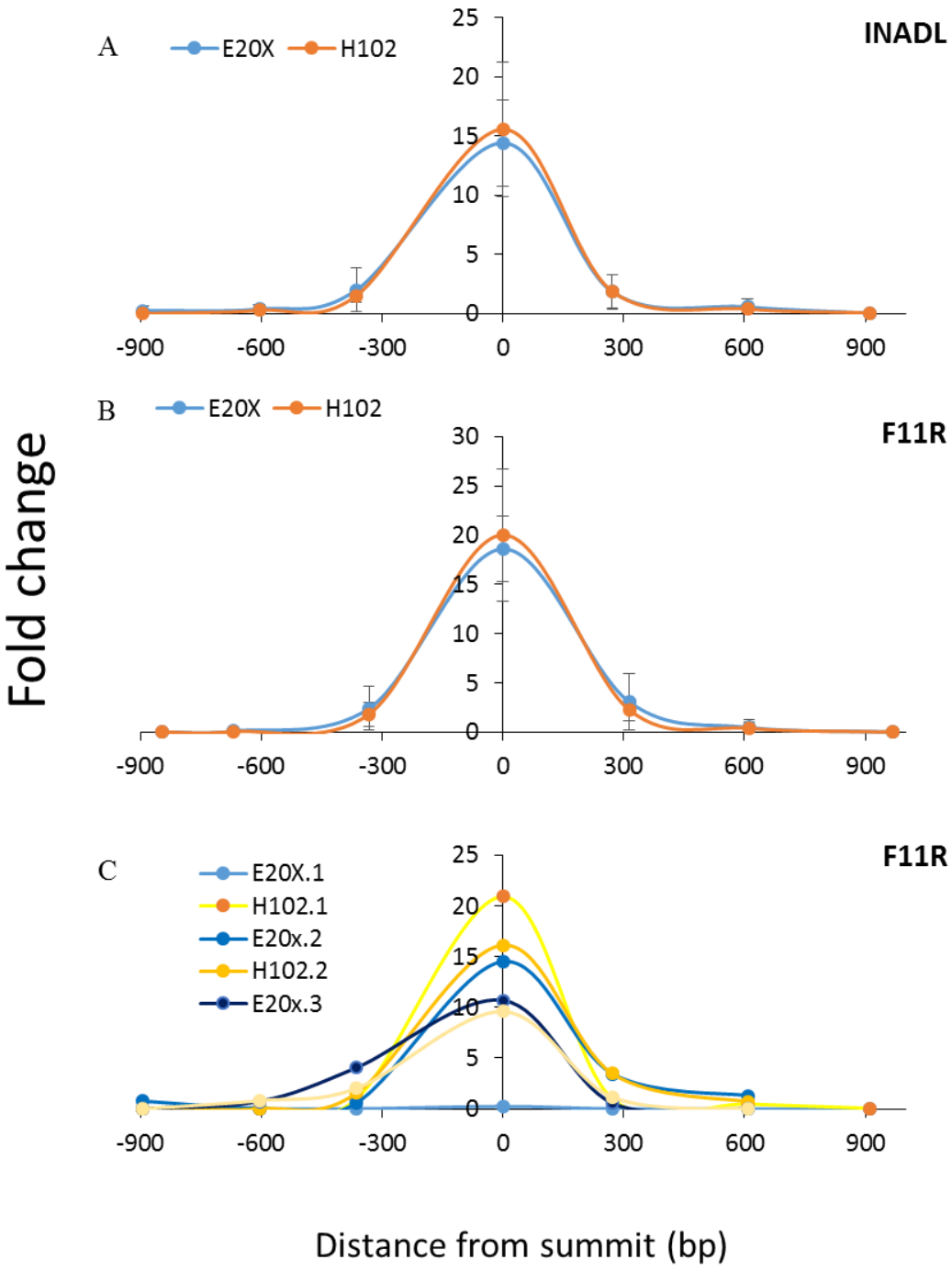


Figure 27 ChIP-qRT-PCR of ZEB1 at tiled INADL and F11R promoters in MDA-MB-231 cells.

qRT-PCR of ZEB1 ChIP using E20X (blue) and H102 (orange) for tiled primers across the INADL (A) and F11R (B) promoter with a ~300bp separation and normalised to the average enrichment at GAPDH and HPRT1. Error bars represent ±SD of three biological replicates. (C) The three individual qRT-PCR runs shown in (B) with E20X ChIPs in blue and H102 ChIPs in yellow.

Note: As the Cts were very low for some of the peripheral sites these only consist of 1-2 repeats.

Table 21 Summary of gene expression and epigenetic state of novel ZEB1 target genes, F11R and INADL. Expression was measured as area under the peak. Fold change in expression or epigenetic marks was measured as HMLE vs mesHMLE cells. Both genes maintained an active only epigenetic signature in both cell lines.

Gene	HMLE expression	MesHMLE expression	Fold change in expression (log2)	q_value $x10^{-4}$	H3K4me3	H3K27me3
F11R	73.65	16.03	-2.20	3.1	-1.03	0
INADL	14.72	2.07	-2.83	1.0	-0.72	0

5.3 Discussion

5.3.1 Use of ZEB1 antibodies in ChIP-seq

One aspect that came became prevalent during my research was the difficulty of obtaining high quality ZEB1 ChIP-seq data. This was observed both in my research work and that of a Ph.D. student from the Brabletz group (personal correspondence). Despite these problems encountered with ZEB1 ChIP-seq, successful experiments were performed as part of the ENCODE project in non-EMT cell-lines [137]. Three possible reason for this difference are: significantly more experience with ChIP-seq in their laboratory or that their specific cell-lines were more amenable to the ChIP-seq process. A third explanation is that they used data that was derived from similar low quality ChIP-seq however they had some additional techniques to generate ChIP-seq quality chromatin from this initial sample.

Tackling this challenge of ZEB1 ChIP-seq in EMT models is important for future analysis and further research into the role of ZEB1. Raising new ZEB1 antibodies is a good, if lengthy, approach to potentially improve our capacity to ChIP-seq ZEB1.

5.3.2 Interpretation of the ENCODE ZEB1 ChIP-seq

My analysis of the ENCODE ZEB1 ChIP-seq data provides insights into ZEB1 binding properties (5.2.3). Specifically, ~30% of all potential ZEB1 target genes had a single ZEB1 binding site, and this percentage remained largely constant for the highest ranked genes (Figure 23). This suggests that, contrary to most published work [30, 31, 57], a single ZEB1 site is sufficient for binding. Of note, the two genes validated in my work, F11R and INADL both had a single ZEB1 site providing some experimental evidence that this is sufficient for ZEB1 binding and function.

The data also support previous findings [57] on the average distance between functional ZEB1 motifs with ~70% having a 10-50bp separation (Figure 23). As previously mentioned, there was no rigorous experimental follow up to determine whether these single sites were sufficient or on the effect of differing linker lengths on ZEB1 binding activity; however, the *in silico* analysis still provides valuable insights. It would be informative to make a model where ZEB1 sites were incorporated upstream of promoters either singly or in pairs to test whether a single site is sufficient.

5.3.3 Validation of ZEB1 targets

I tested 26 potential ZEB1 target genes with a set of rigorous criteria. They had to respond to modulations in ZEB1 expression at the RNA (5.2.5), luciferase (5.2.6) and ChIP for ZEB1 (5.2.7) in two epithelial and two mesenchymal cell lines. Of the 26 genes, only two met these criteria, F11R and INADL. There are a number of potential explanations for this low success rate. As the ChIP-seq data I used were produced in two cell lines that are highly divergent from my model system, B and liver vs mammary cells, it is possible that genes that were bound by ZEB1 in these systems were not targeted in my cell lines. A simple experiment to

test this would be to repeat the assays using the ChIP-seq cell lines. However, this was beyond the scope of my project.

Another potential explanation is that, as my first criterion was whether the potential target genes responded to modulations to ZEB1 expression at the RNA level, some of the potential target genes were bound by ZEB1 but this did not produce a change at the RNA level in my system. This lack of response could be due to missing cofactors or these genes being unaffected by ZEB1 binding. The final possibility is that the ChIP-seq data was of low quality and despite combining two separate runs in different cell lines there was a common bias to the ChIP-sequencing data which led to an abundance of false positives in my list of candidate genes. Of these, I propose that the most probable explanation consists of a combination of these factors with cell line specificity playing a major role. As previously discussed (3.2.2) cell line specificity plays an important role in the importance of different EMT TFs so it is likely also important for the specific targets of said EMT TFs.

5.3.4 Characterisation of ZEB1 targets

Both of my confirmed ZEB1 targets are part of the same tight-junction complex. This was of particular interest as ZEB1 was already known to target another member of this complex, CRB3. This is of note due to the increasing understanding that TFs and miRNAs act by targeting multiple rather than individual members of the same pathway [142]. This combined targeting can result in a stronger effect and helps ensure that critical pathways are effectively regulated.

To determine the effect of ZEB1 expression on tight-junctions an informative set of experiments would be to analyse tight-junctions using IF while modulating ZEB1 expression.

This would confirm whether the presence of key tight-junction members at the membrane was dependent on ZEB1 expression levels. Subsequent experiments could expand on this by evaluating cell-cell adhesion strength. This change, and change in migration, has typically been attributed to ZEB1 mediated repression of CDH1 however, my research suggests that ZEB1 is acting simultaneously on a variety of cell-cell adhesion molecules presumably resulting in a stronger effect than would be obtained based on repressing a single adhesion factor.

Chapter 6 Discussion and Conclusion

The aim of this chapter is to summarise the main findings of this thesis, place these findings in context of our current understanding of the literature, and identify directions for further research.

6.1 Discussion

6.1.1 Emerging picture of the role of bivalently regulated genes in development and EMT

Based on studies of disease incidence, breast cancer remains one of the most clinically relevant diseases [15]. Consequently, determining the molecular mechanisms behind breast cancer progression and devising new strategies to identify novel genes remain a pressing need. I examined this issue through the lens of bivalent genes in EMT.

Epithelial to Mesenchymal Transition (EMT) is an important and complex cellular process in embryonic development, wound healing and tumour progression. It has been proposed that prior to malignant progression a subset of tumour cells undergo an EMT which facilitates the development of key malignant properties such as anoikis resistance, chemoresistance, increased migration and invasion. This combination of properties is hypothesised to allow the now mesenchymal-like cells to separate from the primary tumour, enter the circulatory system and potentially colonise distant sites to produce secondary metastases. This initial EMT is thought to be accompanied by a subsequent MET which is required for the formation of micro, and then macro, metastases. Bivalent genes have only recently been associated with EMT in cancer [83]; however they have an established role within developmental EMT [1]. This raised the question of the extent to which they play a role in cancer EMT.

My study of the role of bivalent genes in EMT began by analysing ChIP-seq data of bivalent epigenetic marks in the HMLE EMT model to identify genes that changed from a bivalent to an active epigenetic state. This analysis identified a number of previously established EMT genes including ZEB1 and TWIST. As I discuss in Section 3.2.2, in all bivalent ChIP-seq data sets examined, some proportion of the master EMT TFs was activated from a bivalent state with the specific changes and EMT TFs varying dependent on the model. Interestingly, I also observed two key epithelial genes, CDH1 and ESRP1, which underwent the opposite transition, from an active to a bivalent epigenetic signature. These observations provide some of the strongest evidence presented for the role of bivalent genes in EMT and support the potential of using bivalent gene lists to identify novel EMT associated genes. To test this, I examined in detail the effect of three such genes, ADM2, PLEKHO1 and RASA3, on the EMT-associated properties, proliferation and migration (Chapter 3). I also increased our understanding of the role of ZEB1 in EMT through characterising ZEB1 isoform expression during EMT (Chapter 4) and the identification of two novel ZEB1 targets with a role in tight-junction formation (Chapter 5).

Malouf *et al* [116] published an interesting study on EMT and bivalence, which is particularly relevant to my hypothesis that bivalent genes are important for malignant EMT. They examined the effect of TWIST overexpression on epigenetic marks in HMLE cells with a focus on general epigenetic trends rather than on specific genes. Unlike our HMLE epigenetic ChIP-seq data (3.1.3) they found that few genes underwent a bivalent to active transformation. However, in agreement with our findings, they reported a high proportion of genes changed from an active to a bivalent signature. Their data support my proposition that bivalent genes are relevant for MET but not my hypothesis that bivalent genes are important

for EMT. A possible reason for the discrepancy is that they identified approximately 3-fold fewer bivalent genes in parental HMLE cells than we did (3.1.3).

6.1.2 Characterisation of ADM2 and PLEKHO1 with an EMT focus

To determine the potential of using bivalent gene lists to identify novel EMT associated genes I characterised the expression and epigenetic status of three novel bivalently regulated genes, ADM2 PLEKHO1 and RASA3, in epithelial and mesenchymal cell lines and assayed their effect on the EMT properties of migration and proliferation in a variety of breast cells. First I used *in vitro* and *in silico* methods to determine their expression level in epithelial vs mesenchymal cell lines and determined that PLEKHO1 and RASA3 were predominantly expressed in epithelial cell lines from a variety of origins while for ADM2 this appeared to be specific for breast cells (3.2.1). Second I confirmed that these genes had genuine bivalent promoters through ChIP-reChIP (3.2.2). Third I showed that there was no clear link between ADM2, PLEKHO1 or RASA3 expression and proliferation (3.2.4). Finally, I identified that both ADM2 and PLEKHO1 expression correlates with migration in a variety of breast cells while RASA3 did not (3.2.5, 3.2.6). Future studies using mouse xenograft models would be crucial to determine the effect of ADM2 and PLEKHO1 on breast cancer metastasis in an *in vivo* setting.

Importantly the co-expression of PLEKHO1 and ADM2 resulted in a synergistic effect on migration (3.2.8, 3.2.9). I hypothesise that this synergistic effect is mediated through PI3K signalling due to the convergence of PLEKHO1 as a downstream effector of this pathway ([143, 144], 3.2.7) and ADM2 as an upstream activator [117]. Correspondingly, an informative experiment to test this would be to use IF to determine whether PLEKHO1 co-localises with PIP3 in MCF10A_PLEKHO1 cells when grown in full media and whether co-

localisation is lost under starvation and rescued with ADM2 conditioned media. As a complement to this, determining the effect of a PI3K inhibitor on migration and PLEKHO1 localisation would also be very informative. If this thorough examination of the relationship between PLEKHO1 and ADM2 supports my initial findings it raises the possibility that bivalent genes that are regulated together might have overlapping function or interactions. This idea is supported by my earlier observations on GO groupings (3.1.3) and further paired examinations of bivalently regulated genes could identify important new relationships between key EMT genes.

My findings support multiple studies published during my research on the role of ADM2 as a prognostic marker for cancer progression. A 2015 analysis of ADM2 plasma levels in breast cancer patients showed that ADM2 is a strong predictor of 5 year mortality rates [103]. In addition to this, ADM2 has been proposed as a prognostic marker in prostate [145], pancreatic [104] and colorectal [146] cancer This combination of studies provide strong evidence for the importance of ADM2 in cancer. My finding on ADM2's role in migration and its potential importance in EGF signalling suggest that ADM2 is not over expressed in tumours as a passenger but as an active player and presents a potential mechanism for this. Additional evidence that ADM2 is a strong effector on migration comes from its role in pregnancy. Two studies [96, 97] showed that ADM2 overexpression increases invasion in human first trimester trophoblasts and stimulates the invasive capacity of spontaneously aborted first-trimester primary extravillous cytotrophoblast cells although it remains unclear how this mechanism aligns with my proposed mechanism.

The role of PLEKHO1 remains unclear and appears to differ depending on the study and system. A 2015 paper showed that PLEKHO1 had a positive correlation with migration,

invasion and tumour formation in prostate cancer cell lines [105]. In direct contrast to this, a 2016 paper examining the role of a lnc-RNA in gastric cancer identified PLEKHO1 as a tumour suppressor which was routinely repressed in gastric cancer patients [147]. They also showed that PLEKHO1 overexpression decreased migration and proliferation. In addition, there was a 2014 paper that described PLEKHO1 as a tumour suppressor in colon cancer [148] based on both expression, prognosis and functional assays; their functional assays showed that PLEKHO1 knockdown resulted in greater tumour size. Importantly, they did not show that PLEKHO1 knockdown increased migration so this evidence is not in direct conflict with my findings. My data support the study of Kim *et al* [143] as PLEKHO1 increased migration in my breast model.

The most likely explanation for the discrepancy in PLEKHO1's role in cancer and migration is that the role of PLEKHO1 is highly cell line dependent. Specifically, as PLEKHO1 acts as a scaffolding protein [149, 150], the particular combination of partner proteins would have a large impact on the role of PLEKHO1.

6.1.3 Characterising expression of ZEB1 isoforms in EMT

In addition to evaluating the role of novel EMT associated genes I examined the established master EMT TF ZEB1.

As previously discussed (4.1.1), isoforms and isoform switches play a key role in a variety of cellular processes, including EMT. As ZEB1 isoforms are almost completely unstudied in the literature I used the HMLE-mesHMLE EMT model to determine whether ZEB1 isoforms were preferentially regulated during EMT using isoform specific primers (4.2.2). To my knowledge this was the first quantification of ZEB1 isoform expression in the literature. I found that the full-length isoform accounted for ~90% of total ZEB1 across the EMT time

course with the other isoforms comprising the, largely constant, remainder. From this I concluded that ZEB1 isoforms were not differentially regulated at the RNA level during EMT although I cannot exclude the possibility that they are regulated at the protein level.

Some interesting experiments that could be performed to assess the importance of these isoforms include making isoform-specific expression constructs and overexpressing these in epithelial cells paired with RNA-seq to determine whether there are differences in the genes that are regulated. My hypothesis is that all four isoforms would induce an EMT due to their structural similarity. Any discrepancies would provide valuable insights into the relationship between ZEB1 domains and function. In addition, a critical experiment that was beyond the scope of my project is to quantify the protein expression of these different isoforms. As they cannot readily be differentiated by size through Western blot and raising antibodies against the unique fragments is fraught with issues, mass spectrometry and 2-dimensional gel-electrophoresis are the most appropriate methods [151].

6.1.4 Bioinformatic insights on ZEB1 binding

I continued to investigate ZEB1 by using publicly available ZEB1 ChIP-seq data to investigate the discrepancies between previous studies on the mode of ZEB1 binding and to identify novel ZEB1 targets [30, 57, 58]. From these previous studies, it was unclear whether ZEB1 required a paired binding site to bind DNA. Franklin *et al* [58] used reporter assays to show that a single ZEB1 site caused an ~40% repression while two sites caused an ~80% repression. Subsequent experiments typically refer to ZEB1 binding to a pair of Z-boxes [57]. To reconcile the differences between these studies I examined both the prevalence of paired Z-boxes in the ChIP-seq data-set as well as the relationship between number of Z-boxes and ChIP-seq score (5.2.3). My data strongly supported the original findings that a single Z-box

was sufficient for ZEB1 binding as peaks with a single site accounted for a large proportion of both total and top ranked genes. In addition, the strong relationship between the number of Z-boxes and score reinforces their findings that more Z-boxes resulted in stronger repression. This is not surprising as more Z-boxes will increase the likelihood that a ZEB1 molecule will bind resulting in repression, even if they do not act as a pair to recruit a single ZEB1 molecule with higher affinity.

A non-controversial, though understudied aspect of ZEB1 binding that I assessed was the distance between paired Z-boxes. Here there are significant intrinsic limitations in my analysis as I lack functional data confirming that the paired site is bound. Despite this, the data support the earlier findings [57] on distance between Z-boxes with the vast majority separated by 10-50bp.

6.1.5 Validation and characterisation of novel ZEB1 targets

My main interest in studying ZEB1 was to identify novel targets. I used the ChIP-seq data-set previously described to identify potential targets and evaluated these using a range of complementary criteria and tests. Through this I identified F11R and INADL as genuine ZEB1 targets based on their negative relationship with ZEB1 as measured through qRT-PCR and luciferase assays. ZEB1 promoter binding was assessed and confirmed through ChIP-qRT-PCR.

One interesting link between F11R and INADL is that they are both members of the tight-junction complex. Tight-junction complexes are well established players for both adhesion and signalling mechanisms and their presence is important to maintain cells in an epithelial state [152]. In addition, ZEB1 has an established history with tight-junction members. In a microarray study, ZEB1 was shown to repress another member of the tight-junction complex,

CRB3 and this has been further validated through my experiments [29]. As previously discussed, there is a growing interest in transcription factors and miRNAs affecting networks through targeting multiple components [142, 153]. With three members of the tight-junction complex now identified as ZEB1 targets, it appears likely that repression of this complex is a key aspect of ZEB1 activity. This is of note as the role of ZEB1 has traditionally been associated with its repression of the miR-200 family [32, 36] and a different adhesion molecule, CDH1 [29, 154]. The combined evidence for a role of ZEB1 on tight-junctions suggests that this paradigm bears re-examination.

During my studies, Chaffer *et al* [115] published an interesting report on the importance of bivalent genes in cancer associated EMT. They showed the importance of a bivalent ZEB1 promoter for the plasticity of human basal breast cells to transition into cancer stem cells (CSC). The transition into a CSC bears some similarities to those exhibited by an EMT and the two often co-occur [155]. Of interest, their initial observation, that ZEB1 is epigenetically controlled by a bivalent promoter in breast cells, was mirrored in my work. Where our studies differ is that they focussed on the role of ZEB1 on the conversion between $CD44^{lo}$ and $CD44^{hi}$ and how this affects the stem-like state in basal breast cells while I provided insights into ZEB1 binding mechanisms and added significant data supporting the role of ZEB1 in the regulation of F11R and INADL, two genes associated with the classic EMT property of cell adhesion. The combination of these studies furthers the importance of ZEB1 in EMT and identifies the bivalent epigenetic regulation of ZEB1 as a key transcriptional regulatory mechanism for the EMT process. Combined, these data present a novel perspective on ZEB1 regulation as this was typically associated with the post-transcriptional regulation executed by the miR-200 family.

6.2 Conclusion

While bivalent genes have not been ignored in the field of EMT they have, so far, been understudied. My work addressed this issue and identified two novel EMT associated genes that play an important role in migration. I also contributed to the establishment of ZEB1 as a bivalently regulated gene and identified two novel ZEB1 targets and used bioinformatics to obtain information about ZEB1 binding. Alongside these experimental insights into the role of bivalent genes in EMT I have provided bioinformatic analysis showing that different master EMT TFs are bivalently regulated in a variety of different systems. Importantly, some combination of these TFs were bivalently regulated in the three models examined, providing an important and clear link between this mode of regulation and EMT.

This combination of observations strongly suggests that there is an important role for bivalent genes in cancer EMT and that bioinformatic data on bivalently regulated genes in EMT represent a valuable, underutilised resource for the identification of novel EMT genes. I anticipate that further examination of bivalent genes that are regulated in EMT will identify novel EMT associated genes.

In addition to this, I hypothesise that the presence of a bivalent signature in association with master EMT TFs in epithelial cell lines or tumours would correlate with the activation and importance of these in EMT or metastasis providing information about downstream signalling and gene signatures. Of clinical relevance, it would be important to test for associations between the epigenetic status of these master EMT TFs and prognosis. My working hypothesis is that prognosis correlates with the number of EMT TFs that are bivalently silenced in the primary tumour, as the more are in this bivalent state the more can rapidly be turned on to contribute to an EMT.

Chapter 7 Supplementary data

Supplementary Table 1 Summary of epigenetic promoter status in HMLE, mesHMLE, mucosa samples and tumour samples.
Table is sorted first by bivalent status in HMLE = bivalent (yellow) then bivalent status in mesHMLE = active only (green). Data from the mucosa and tumour samples is adapted from Hahn *et al*, 2014 [81].

Gene ID	HMLE bivalent status	MesHMLE bivalent status	Bivalent status in 3/4 mucosa samples	Bivalent status in tumour sample
ABCA2	bivalent	active only	N/A	N/A
ABCG4	bivalent	active only	bivalent	bivalent
ACO1	bivalent	active only	N/A	N/A
ADAM12	bivalent	active only	bivalent	active only
ADAMTS7	bivalent	active only	N/A	N/A
ADM2	bivalent	active only	bivalent	bivalent
AGBL4	bivalent	active only	bivalent	bivalent
AK5	bivalent	active only	bivalent	active only
AKD1	bivalent	active only	N/A	N/A
ALDH5A1	bivalent	active only	N/A	N/A
ALS2CR8	bivalent	active only	N/A	N/A
ANGPT1	bivalent	active only	N/A	N/A
ANGPTL2	bivalent	active only	N/A	N/A
ANKRD35	bivalent	active only	N/A	N/A
ANXA6	bivalent	active only	N/A	N/A
ARHGAP24	bivalent	active only	N/A	N/A
ARHGEF10L	bivalent	active only	N/A	N/A
ARHGEF25	bivalent	active only	N/A	N/A
ARHGEF40	bivalent	active only	N/A	N/A
ARMC3	bivalent	active only	bivalent	bivalent
ARRB1	bivalent	active only	N/A	N/A
ARVCF	bivalent	active only	bivalent	bivalent
AS3MT	bivalent	active only	N/A	N/A
ASAH2B	bivalent	active only	N/A	N/A
ASGR1	bivalent	active only	N/A	N/A
ATP1B1	bivalent	active only	N/A	N/A
ATP8B2	bivalent	active only	N/A	N/A
ATXN7L2	bivalent	active only	N/A	N/A
B3GNT7	bivalent	active only	bivalent	bivalent
BAI2	bivalent	active only	bivalent	bivalent
BAMBI	bivalent	active only	bivalent	active only
BBC3	bivalent	active only	N/A	N/A
BCOR	bivalent	active only	N/A	N/A
BLMH	bivalent	active only	N/A	N/A
BLVRB	bivalent	active only	N/A	N/A
BMP1	bivalent	active only	bivalent	bivalent
BNC2	bivalent	active only	N/A	N/A

BRD3	bivalent	active only	bivalent	bivalent
BUB3	bivalent	active only	bivalent	active only
C11orf95	bivalent	active only	N/A	N/A
C12orf34	bivalent	active only	N/A	N/A
C13orf33	bivalent	active only	bivalent	bivalent
C14orf132	bivalent	active only	bivalent	bivalent
C16orf5	bivalent	active only	N/A	N/A
C19orf29-AS1	bivalent	active only	N/A	N/A
C19orf76	bivalent	active only	N/A	N/A
C1orf115	bivalent	active only	N/A	N/A
C1QTNF2	bivalent	active only	N/A	N/A
C21orf56	bivalent	active only	N/A	N/A
C4orf48	bivalent	active only	bivalent	bivalent
C9orf139	bivalent	active only	N/A	N/A
C9orf50	bivalent	active only	N/A	N/A
CAMK1D	bivalent	active only	N/A	N/A
CAMK4	bivalent	active only	bivalent	bivalent
CAMKK1	bivalent	active only	N/A	N/A
CBS	bivalent	active only	bivalent	bivalent
CCDC136	bivalent	active only	N/A	N/A
CCDC85A	bivalent	active only	bivalent	bivalent
CD151	bivalent	active only	bivalent	bivalent
CD72	bivalent	active only	N/A	N/A
CDC37	bivalent	active only	bivalent	bivalent
CDH12	bivalent	active only	bivalent	bivalent
CDH2	bivalent	active only	bivalent	bivalent
CDH4	bivalent	active only	bivalent	bivalent
CELF6	bivalent	active only	N/A	N/A
CHN1	bivalent	active only	N/A	N/A
CHRDL1	bivalent	active only	bivalent	bivalent
CHST2	bivalent	active only	bivalent	bivalent
CLDN11	bivalent	active only	bivalent	bivalent
CLEC14A	bivalent	active only	bivalent	bivalent
CLGN	bivalent	active only	N/A	N/A
CMTM8	bivalent	active only	bivalent	active only
CNKSR3	bivalent	active only	N/A	N/A
CNRIP1	bivalent	active only	N/A	N/A
COL11A1	bivalent	active only	N/A	N/A
COL5A1	bivalent	active only	bivalent	bivalent
CPNE7	bivalent	active only	bivalent	active only
CPT1C	bivalent	active only	bivalent	bivalent
CREB5	bivalent	active only	N/A	N/A
CRISPLD1	bivalent	active only	N/A	N/A
CTSF	bivalent	active only	N/A	N/A
CXXC4	bivalent	active only	N/A	N/A
CXXC5	bivalent	active only	N/A	N/A

DAGLA	bivalent	active only	N/A	N/A
DBP	bivalent	active only	N/A	N/A
DCC	bivalent	active only	N/A	N/A
DCDC1	bivalent	active only	N/A	N/A
DCLK2	bivalent	active only	N/A	N/A
DEXI	bivalent	active only	N/A	N/A
DGKI	bivalent	active only	bivalent	bivalent
DNAH6	bivalent	active only	N/A	N/A
DNAJC24	bivalent	active only	N/A	N/A
DNER	bivalent	active only	bivalent	bivalent
DNM1P46	bivalent	active only	N/A	N/A
DOCK10	bivalent	active only	bivalent	bivalent
DOCK11	bivalent	active only	N/A	N/A
DOCK2	bivalent	active only	N/A	N/A
E2F2	bivalent	active only	N/A	N/A
EEF1DP3	bivalent	active only	bivalent	bivalent
EFCAB4B	bivalent	active only	N/A	N/A
EFR3B	bivalent	active only	N/A	N/A
EMID1	bivalent	active only	bivalent	bivalent
ENDOV	bivalent	active only	N/A	N/A
ENPP1	bivalent	active only	N/A	N/A
EPB41L1	bivalent	active only	N/A	N/A
EPDR1	bivalent	active only	bivalent	bivalent
EPHA3	bivalent	active only	bivalent	bivalent
ERICH1	bivalent	active only	N/A	N/A
ESAM	bivalent	active only	bivalent	bivalent
FAM134B	bivalent	active only	N/A	N/A
FAM171A2	bivalent	active only	N/A	N/A
FAM20C	bivalent	active only	N/A	N/A
FAM212B	bivalent	active only	N/A	N/A
FAM26F	bivalent	active only	bivalent	bivalent
FAM5C	bivalent	active only	bivalent	bivalent
FAM70B	bivalent	active only	bivalent	active only
FANCE	bivalent	active only	N/A	N/A
FGF20	bivalent	active only	bivalent	bivalent
FGFR1	bivalent	active only	N/A	N/A
FHOD1	bivalent	active only	bivalent	bivalent
FIG4	bivalent	active only	N/A	N/A
FLI1	bivalent	active only	bivalent	bivalent
FOXC2	bivalent	active only	bivalent	bivalent
FOXE3	bivalent	active only	bivalent	bivalent
FOXH1	bivalent	active only	bivalent	bivalent
FOXJ1	bivalent	active only	N/A	N/A
FOXO4	bivalent	active only	N/A	N/A
FRG1B	bivalent	active only	N/A	N/A
FRY	bivalent	active only	N/A	N/A

FRZB	bivalent	active only	N/A	N/A
FZD1	bivalent	active only	N/A	N/A
FZD7	bivalent	active only	N/A	N/A
GAB3	bivalent	active only	bivalent	bivalent
GALNT10	bivalent	active only	N/A	N/A
GAMT	bivalent	active only	bivalent	bivalent
GAS1	bivalent	active only	N/A	N/A
GBGT1	bivalent	active only	N/A	N/A
GFPT2	bivalent	active only	bivalent	bivalent
GIPC2	bivalent	active only	bivalent	active only
GJA3	bivalent	active only	bivalent	bivalent
GJC3	bivalent	active only	N/A	N/A
GLI1	bivalent	active only	N/A	N/A
GMNC	bivalent	active only	N/A	N/A
GNA14	bivalent	active only	N/A	N/A
GNAS-AS1	bivalent	active only	N/A	N/A
GPR124	bivalent	active only	N/A	N/A
GPR137B	bivalent	active only	N/A	N/A
GPR173	bivalent	active only	bivalent	bivalent
GPR62	bivalent	active only	N/A	N/A
GPRASP2	bivalent	active only	bivalent	active only
GRB10	bivalent	active only	N/A	N/A
GREM1	bivalent	active only	bivalent	bivalent
GRIA3	bivalent	active only	N/A	N/A
GRWD1	bivalent	active only	bivalent	bivalent
GYPC	bivalent	active only	N/A	N/A
HBZ	bivalent	active only	N/A	N/A
HECW2	bivalent	active only	bivalent	bivalent
HLX	bivalent	active only	N/A	N/A
HMGCLL1	bivalent	active only	N/A	N/A
HOXB-AS3	bivalent	active only	N/A	N/A
HOXC10	bivalent	active only	bivalent	bivalent
HS3ST3A1	bivalent	active only	bivalent	bivalent
HS3ST3B1	bivalent	active only	bivalent	bivalent
HTR1B	bivalent	active only	bivalent	bivalent
HTR2A	bivalent	active only	N/A	N/A
IGFBP2	bivalent	active only	N/A	N/A
IL13	bivalent	active only	bivalent	bivalent
IL17RD	bivalent	active only	N/A	N/A
IRF2BP1	bivalent	active only	N/A	N/A
ITGA4	bivalent	active only	bivalent	bivalent
ITGA7	bivalent	active only	N/A	N/A
ITPKA	bivalent	active only	bivalent	bivalent
JAM3	bivalent	active only	N/A	N/A
JDP2	bivalent	active only	N/A	N/A
KAL1	bivalent	active only	bivalent	bivalent

KCNJ10	bivalent	active only	N/A	N/A
KCNMA1	bivalent	active only	bivalent	bivalent
KCNT2	bivalent	active only	N/A	N/A
KDR	bivalent	active only	N/A	N/A
KIAA1614	bivalent	active only	N/A	N/A
KIAA1797	bivalent	active only	N/A	N/A
KY	bivalent	active only	N/A	N/A
LEPREL2	bivalent	active only	bivalent	bivalent
LIFR	bivalent	active only	N/A	N/A
LIMD2	bivalent	active only	N/A	N/A
LIN7A	bivalent	active only	bivalent	bivalent
LMOD1	bivalent	active only	N/A	N/A
LOC100126784	bivalent	active only	N/A	N/A
LOC100128511	bivalent	active only	N/A	N/A
LOC100131208	bivalent	active only	N/A	N/A
LOC100287216	bivalent	active only	N/A	N/A
LOC100287846	bivalent	active only	N/A	N/A
LOC100294362	bivalent	active only	N/A	N/A
LOC100506343	bivalent	active only	N/A	N/A
LOC151174	bivalent	active only	N/A	N/A
LOC386597	bivalent	active only	N/A	N/A
LOC388588	bivalent	active only	N/A	N/A
LOC643387	bivalent	active only	N/A	N/A
LOC654342	bivalent	active only	N/A	N/A
LOC728392	bivalent	active only	N/A	N/A
LOC729683	bivalent	active only	N/A	N/A
LRRC32	bivalent	active only	bivalent	bivalent
LRRC33	bivalent	active only	bivalent	bivalent
LRRC4C	bivalent	active only	N/A	N/A
LRRC6	bivalent	active only	N/A	N/A
MAP1A	bivalent	active only	N/A	N/A
MARCKSL1	bivalent	active only	N/A	N/A
MCF2L2	bivalent	active only	bivalent	bivalent
MEGF8	bivalent	active only	bivalent	bivalent
MEX3B	bivalent	active only	N/A	N/A
MIR1181	bivalent	active only	N/A	N/A
MIR196A2	bivalent	active only	N/A	N/A
MIR3610	bivalent	active only	N/A	N/A
MIR3650	bivalent	active only	N/A	N/A
MIR618	bivalent	active only	bivalent	active only
MLPH	bivalent	active only	bivalent	active only
MMP11	bivalent	active only	N/A	N/A
MMP17	bivalent	active only	bivalent	bivalent
MOGAT1	bivalent	active only	bivalent	active only
MOXD1	bivalent	active only	bivalent	bivalent
MRAP2	bivalent	active only	bivalent	bivalent

MSX1	bivalent	active only	bivalent	bivalent
MTMR7	bivalent	active only	bivalent	bivalent
MUC1	bivalent	active only	N/A	N/A
MYADM	bivalent	active only	N/A	N/A
MYB	bivalent	active only	N/A	N/A
MYBPC2	bivalent	active only	bivalent	bivalent
MYL9	bivalent	active only	N/A	N/A
NAV1	bivalent	active only	N/A	N/A
NAV2	bivalent	active only	N/A	N/A
NFASC	bivalent	active only	N/A	N/A
NFATC4	bivalent	active only	N/A	N/A
NFIC	bivalent	active only	N/A	N/A
NID1	bivalent	active only	N/A	N/A
NIM1	bivalent	active only	N/A	N/A
NME4	bivalent	active only	N/A	N/A
NPDC1	bivalent	active only	bivalent	bivalent
NPL	bivalent	active only	N/A	N/A
NR2C2AP	bivalent	active only	N/A	N/A
NR2F1	bivalent	active only	bivalent	bivalent
NRK	bivalent	active only	bivalent	bivalent
NUDT14	bivalent	active only	N/A	N/A
OPRL1	bivalent	active only	N/A	N/A
OSBPL5	bivalent	active only	N/A	N/A
PANX2	bivalent	active only	bivalent	bivalent
PBX1	bivalent	active only	N/A	N/A
PCDHGA11	bivalent	active only	N/A	N/A
PCDHGA7	bivalent	active only	N/A	N/A
PCDHGB2	bivalent	active only	N/A	N/A
PCDHGB6	bivalent	active only	N/A	N/A
PDGFRL	bivalent	active only	bivalent	bivalent
PHOSPHO1	bivalent	active only	N/A	N/A
PKN1	bivalent	active only	N/A	N/A
PLAC9	bivalent	active only	N/A	N/A
PLCE1	bivalent	active only	N/A	N/A
PLEKHO1	bivalent	active only	N/A	N/A
PODNL1	bivalent	active only	N/A	N/A
PODXL	bivalent	active only	N/A	N/A
PPAPDC1A	bivalent	active only	bivalent	active only
PPM1H	bivalent	active only	N/A	N/A
PPM1L	bivalent	active only	N/A	N/A
PPM1N	bivalent	active only	N/A	N/A
PRDM11	bivalent	active only	N/A	N/A
PREX1	bivalent	active only	bivalent	bivalent
PRKCD	bivalent	active only	N/A	N/A
PRKD1	bivalent	active only	N/A	N/A
PRRT2	bivalent	active only	N/A	N/A

PRRX1	bivalent	active only	bivalent	active only
PRTG	bivalent	active only	bivalent	bivalent
PRUNE2	bivalent	active only	N/A	N/A
PTCH2	bivalent	active only	N/A	N/A
PTGDS	bivalent	active only	N/A	N/A
PVRL3	bivalent	active only	N/A	N/A
PYCR1	bivalent	active only	N/A	N/A
PYGO1	bivalent	active only	N/A	N/A
RAB40B	bivalent	active only	N/A	N/A
RAD21	bivalent	active only	N/A	N/A
RAD21-AS1	bivalent	active only	N/A	N/A
RANBP17	bivalent	active only	N/A	N/A
RASA3	bivalent	active only	N/A	N/A
RASD2	bivalent	active only	bivalent	bivalent
RCOR2	bivalent	active only	N/A	N/A
RGS7	bivalent	active only	bivalent	bivalent
RIMBP3	bivalent	active only	N/A	N/A
RNF144A	bivalent	active only	N/A	N/A
RNF208	bivalent	active only	N/A	N/A
RNF220	bivalent	active only	bivalent	bivalent
RRAD	bivalent	active only	bivalent	bivalent
RSRC1	bivalent	active only	bivalent	bivalent
RTN2	bivalent	active only	bivalent	bivalent
SARM1	bivalent	active only	N/A	N/A
SATB2	bivalent	active only	bivalent	active only
SCARF2	bivalent	active only	N/A	N/A
SCN8A	bivalent	active only	bivalent	bivalent
SDC3	bivalent	active only	N/A	N/A
SEC1	bivalent	active only	N/A	N/A
SEMA4C	bivalent	active only	N/A	N/A
SERHL2	bivalent	active only	N/A	N/A
SFRP4	bivalent	active only	bivalent	active only
SH3GL3	bivalent	active only	N/A	N/A
SH3RF3	bivalent	active only	N/A	N/A
SHOX2	bivalent	active only	bivalent	bivalent
SIX2	bivalent	active only	bivalent	bivalent
SIX3	bivalent	active only	bivalent	bivalent
SLC16A12	bivalent	active only	bivalent	bivalent
SLC2A10	bivalent	active only	N/A	N/A
SLC43A1	bivalent	active only	N/A	N/A
SLC4A8	bivalent	active only	bivalent	bivalent
SLC7A2	bivalent	active only	N/A	N/A
SLC9A5	bivalent	active only	bivalent	bivalent
SLFN11	bivalent	active only	bivalent	bivalent
SLIT2	bivalent	active only	bivalent	bivalent
SLIT3	bivalent	active only	N/A	N/A

SNAPC2	bivalent	active only	bivalent	bivalent
SNED1	bivalent	active only	N/A	N/A
SNX31	bivalent	active only	N/A	N/A
SOX12	bivalent	active only	N/A	N/A
SPATA6	bivalent	active only	N/A	N/A
SPTBN1	bivalent	active only	N/A	N/A
SPTBN4	bivalent	active only	bivalent	bivalent
SSBP4	bivalent	active only	N/A	N/A
SSTR1	bivalent	active only	N/A	N/A
ST8SIA6	bivalent	active only	bivalent	bivalent
STAT5A	bivalent	active only	N/A	N/A
STK32B	bivalent	active only	N/A	N/A
STK33	bivalent	active only	bivalent	bivalent
SUSD5	bivalent	active only	bivalent	bivalent
SV2A	bivalent	active only	N/A	N/A
SYNGR1	bivalent	active only	N/A	N/A
SYNM	bivalent	active only	N/A	N/A
SYT11	bivalent	active only	N/A	N/A
SYTL4	bivalent	active only	bivalent	bivalent
TACR3	bivalent	active only	N/A	N/A
TBX18	bivalent	active only	bivalent	bivalent
TBXA2R	bivalent	active only	N/A	N/A
TDRD7	bivalent	active only	N/A	N/A
TET1	bivalent	active only	bivalent	bivalent
TGFBR3	bivalent	active only	N/A	N/A
THY1	bivalent	active only	bivalent	bivalent
TMEM158	bivalent	active only	N/A	N/A
TMEM47	bivalent	active only	bivalent	bivalent
TMIE	bivalent	active only	N/A	N/A
TOX	bivalent	active only	N/A	N/A
TPTEP1	bivalent	active only	N/A	N/A
TRO	bivalent	active only	N/A	N/A
TSPAN2	bivalent	active only	N/A	N/A
TSPYL5	bivalent	active only	N/A	N/A
TTC28	bivalent	active only	N/A	N/A
TTPA	bivalent	active only	bivalent	bivalent
TUBA1A	bivalent	active only	N/A	N/A
TUBA8	bivalent	active only	bivalent	bivalent
TUSC1	bivalent	active only	N/A	N/A
TWIST1	bivalent	active only	bivalent	bivalent
UBE2E2	bivalent	active only	N/A	N/A
ULBP2	bivalent	active only	bivalent	bivalent
VTN	bivalent	active only	N/A	N/A
VWDE	bivalent	active only	bivalent	bivalent
WASH3P	bivalent	active only	N/A	N/A
WASH5P	bivalent	active only	N/A	N/A

WDR12	bivalent	active only	N/A	N/A
WDR24	bivalent	active only	bivalent	bivalent
WIPF1	bivalent	active only	N/A	N/A
WTIP	bivalent	active only	N/A	N/A
XYLT1	bivalent	active only	N/A	N/A
YIF1A	bivalent	active only	bivalent	bivalent
YPEL1	bivalent	active only	N/A	N/A
YPEL2	bivalent	active only	N/A	N/A
ZBTB16	bivalent	active only	N/A	N/A
ZBTB7A	bivalent	active only	N/A	N/A
ZCCHC12	bivalent	active only	bivalent	bivalent
ZCCHC18	bivalent	active only	bivalent	bivalent
ZDHHC14	bivalent	active only	N/A	N/A
ZDHHC2	bivalent	active only [118]	N/A	N/A
ZEB1	bivalent	active only	N/A	N/A
ZEB1-AS1	bivalent	active only	N/A	N/A
ZFHX2	bivalent	active only	N/A	N/A
ZMAT1	bivalent	active only	bivalent	bivalent
ZNF763	bivalent	active only	N/A	N/A
ZYG11A	bivalent	active only	bivalent	bivalent
AATF	bivalent	bivalent	bivalent	active only
ABCC6P1	bivalent	bivalent	N/A	N/A
ABCC6P2	bivalent	bivalent	N/A	N/A
ABHD12B	bivalent	bivalent	N/A	N/A
ABLIM2	bivalent	bivalent	bivalent	bivalent
ACADL	bivalent	repressive only	bivalent	bivalent
ACCN1	bivalent	bivalent	bivalent	bivalent
ACCN2	bivalent	bivalent	N/A	N/A
ACE	bivalent	bivalent	N/A	N/A
ACOT12	bivalent	bivalent	bivalent	bivalent
ACOXL	bivalent	repressive only	N/A	N/A
ACSF3	bivalent	bivalent	bivalent	bivalent
ACSL6	bivalent	bivalent	bivalent	active only
ACSS1	bivalent	bivalent	N/A	N/A
ACSS3	bivalent	bivalent	bivalent	bivalent
ACTN3	bivalent	bivalent	N/A	N/A
ACVR1C	bivalent	bivalent	N/A	N/A
ADAM22	bivalent	bivalent	N/A	N/A
ADAM23	bivalent	bivalent	bivalent	bivalent
ADAMTS1	bivalent	bivalent	N/A	N/A
ADAMTS17	bivalent	bivalent	N/A	N/A
ADAMTS19	bivalent	bivalent	bivalent	bivalent
ADAMTS5	bivalent	bivalent	bivalent	bivalent
ADAMTS9	bivalent	bivalent	N/A	N/A
ADAMTS9-AS2	bivalent	bivalent	N/A	N/A
ADAMTSL3	bivalent	bivalent	N/A	N/A

ADAP2	bivalent	bivalent	bivalent	bivalent
ADCY1	bivalent	bivalent	bivalent	bivalent
ADCY4	bivalent	bivalent	bivalent	bivalent
ADCY5	bivalent	bivalent	bivalent	bivalent
ADD2	bivalent	bivalent	bivalent	bivalent
ADRA1D	bivalent	bivalent	bivalent	bivalent
ADRA2A	bivalent	bivalent	bivalent	bivalent
ADRB1	bivalent	bivalent	bivalent	bivalent
AFF2	bivalent	bivalent	bivalent	bivalent
AGAP2	bivalent	bivalent	N/A	N/A
AGTR1	bivalent	bivalent	N/A	N/A
AGXT2L1	bivalent	bivalent	bivalent	bivalent
AIF1L	bivalent	bivalent	N/A	N/A
AJAP1	bivalent	repressive only	bivalent	bivalent
ALDH1A2	bivalent	bivalent	bivalent	bivalent
ALOX15	bivalent	bivalent	N/A	N/A
ALOX15P1	bivalent	bivalent	N/A	N/A
ALOX5	bivalent	bivalent	bivalent	bivalent
AMIGO1	bivalent	bivalent	N/A	N/A
AMPH	bivalent	bivalent	bivalent	active only
ANKRD18DP	bivalent	bivalent	N/A	N/A
ANKRD19P	bivalent	bivalent	N/A	N/A
ANKRD27	bivalent	bivalent	bivalent	bivalent
ANKRD34B	bivalent	bivalent	bivalent	bivalent
ANKRD36BP2	bivalent	bivalent	N/A	N/A
ANKRD63	bivalent	bivalent	N/A	N/A
ANKS1B	bivalent	bivalent	N/A	N/A
APBA1	bivalent	bivalent	N/A	N/A
APCDD1	bivalent	bivalent	bivalent	bivalent
APCDD1L	bivalent	bivalent	bivalent	bivalent
APLP1	bivalent	bivalent	bivalent	bivalent
ARHGAP20	bivalent	bivalent	N/A	N/A
ARHGAP44	bivalent	bivalent	N/A	N/A
ARHGEF7	bivalent	bivalent	bivalent	bivalent
ARID3C	bivalent	bivalent	bivalent	bivalent
ARMC4	bivalent	bivalent	N/A	N/A
ARNT2	bivalent	bivalent	bivalent	bivalent
ASRGL1	bivalent	bivalent	N/A	N/A
ASTN2	bivalent	bivalent	bivalent	bivalent
ASXL3	bivalent	bivalent	N/A	N/A
ATF3	bivalent	bivalent	bivalent	bivalent
ATHL1	bivalent	bivalent	bivalent	bivalent
ATP12A	bivalent	bivalent	bivalent	bivalent
ATP2A3	bivalent	bivalent	N/A	N/A
ATP6V1C2	bivalent	bivalent	bivalent	active only
ATP8A2	bivalent	bivalent	N/A	N/A

ATRNL1	bivalent	bivalent	bivalent	bivalent
AVPR1A	bivalent	bivalent	N/A	N/A
AXIN2	bivalent	bivalent	N/A	N/A
B3GAT2	bivalent	bivalent	bivalent	bivalent
B3GNT4	bivalent	bivalent	bivalent	bivalent
B4GALNT3	bivalent	bivalent	bivalent	active only
B4GALNT4	bivalent	bivalent	N/A	N/A
BAALC	bivalent	bivalent	bivalent	bivalent
BAI3	bivalent	bivalent	bivalent	bivalent
BARX2	bivalent	bivalent	bivalent	bivalent
BATF3	bivalent	bivalent	N/A	N/A
BCL11A	bivalent	bivalent	N/A	N/A
BCL2	bivalent	bivalent	N/A	N/A
BEAN1	bivalent	bivalent	N/A	N/A
BEGAIN	bivalent	bivalent	bivalent	bivalent
BEND4	bivalent	bivalent	bivalent	bivalent
BEX5	bivalent	no mark	bivalent	bivalent
BMI1	bivalent	bivalent	N/A	N/A
BMP4	bivalent	bivalent	bivalent	active only
BMP6	bivalent	bivalent	bivalent	bivalent
BMP8A	bivalent	bivalent	bivalent	bivalent
BMP8B	bivalent	bivalent	bivalent	bivalent
BMPER	bivalent	bivalent	bivalent	bivalent
BMPR1B	bivalent	bivalent	bivalent	bivalent
BOC	bivalent	bivalent	N/A	N/A
BRSK2	bivalent	bivalent	bivalent	bivalent
BSN	bivalent	bivalent	N/A	N/A
BSN-AS2	bivalent	bivalent	N/A	N/A
BSPRY	bivalent	bivalent	N/A	N/A
C10orf53	bivalent	repressive only	N/A	N/A
C11orf20	bivalent	bivalent	bivalent	bivalent
C11orf87	bivalent	bivalent	N/A	N/A
C11orf92	bivalent	bivalent	N/A	N/A
C11orf93	bivalent	bivalent	N/A	N/A
C11orf96	bivalent	bivalent	N/A	N/A
C12orf53	bivalent	bivalent	N/A	N/A
C12orf68	bivalent	bivalent	bivalent	bivalent
C13orf15	bivalent	bivalent	bivalent	bivalent
C14orf37	bivalent	bivalent	N/A	N/A
C15orf27	bivalent	bivalent	bivalent	bivalent
C16orf45	bivalent	bivalent	N/A	N/A
C17orf28	bivalent	bivalent	bivalent	bivalent
C17orf72	bivalent	bivalent	bivalent	bivalent
C17orf96	bivalent	bivalent	bivalent	bivalent
C18orf1	bivalent	bivalent	N/A	N/A
C19orf73	bivalent	bivalent	bivalent	bivalent

C1orf114	bivalent	bivalent	N/A	N/A
C1orf194	bivalent	bivalent	bivalent	bivalent
C1orf233	bivalent	bivalent	N/A	N/A
C1orf65	bivalent	repressive only	N/A	N/A
C1orf95	bivalent	bivalent	bivalent	bivalent
C1QL3	bivalent	bivalent	bivalent	bivalent
C20orf112	bivalent	bivalent	N/A	N/A
C20orf201	bivalent	bivalent	bivalent	bivalent
C2CD4A	bivalent	bivalent	N/A	N/A
C2CD4B	bivalent	bivalent	bivalent	bivalent
C2orf70	bivalent	bivalent	bivalent	active only
C3orf70	bivalent	bivalent	N/A	N/A
C3orf72	bivalent	bivalent	bivalent	bivalent
C3orf80	bivalent	bivalent	N/A	N/A
C4orf19	bivalent	bivalent	bivalent	bivalent
C5orf25	bivalent	bivalent	N/A	N/A
C5orf49	bivalent	bivalent	bivalent	bivalent
C6orf97	bivalent	bivalent	bivalent	active only
C8orf34	bivalent	bivalent	N/A	N/A
C8orf56	bivalent	bivalent	bivalent	bivalent
C8orf84	bivalent	bivalent	bivalent	bivalent
C9orf172	bivalent	bivalent	bivalent	bivalent
C9orf4	bivalent	bivalent	N/A	N/A
CA8	bivalent	bivalent	bivalent	bivalent
CABLES1	bivalent	bivalent	bivalent	active only
CABP1	bivalent	bivalent	bivalent	bivalent
CACNA1D	bivalent	repressive only	N/A	N/A
CACNA1G	bivalent	bivalent	bivalent	bivalent
CACNA2D1	bivalent	bivalent	N/A	N/A
CACNA2D3	bivalent	bivalent	bivalent	bivalent
CACNG4	bivalent	bivalent	bivalent	bivalent
CADM1	bivalent	bivalent	N/A	N/A
CADM2	bivalent	bivalent	N/A	N/A
CADM4	bivalent	bivalent	N/A	N/A
CADPS2	bivalent	bivalent	N/A	N/A
CALCB	bivalent	bivalent	bivalent	bivalent
CAMK2B	bivalent	bivalent	bivalent	bivalent
CAMK2N2	bivalent	bivalent	bivalent	bivalent
CASR	bivalent	repressive only	bivalent	bivalent
CASZ1	bivalent	bivalent	N/A	N/A
CBLN2	bivalent	bivalent	N/A	N/A
CBWD1	bivalent	bivalent	N/A	N/A
CBWD2	bivalent	bivalent	N/A	N/A
CBWD3	bivalent	bivalent	N/A	N/A
CBWD5	bivalent	bivalent	N/A	N/A
CBWD6	bivalent	bivalent	N/A	N/A

CBX2	bivalent	bivalent	bivalent	bivalent
CBX4	bivalent	bivalent	bivalent	active only
CBX6	bivalent	bivalent	N/A	N/A
CBX7	bivalent	bivalent	N/A	N/A
CBX8	bivalent	bivalent	bivalent	bivalent
CCDC108	bivalent	bivalent	N/A	N/A
CCDC39	bivalent	bivalent	bivalent	bivalent
CCDC64B	bivalent	repressive only	N/A	N/A
CCDC88C	bivalent	bivalent	N/A	N/A
CCNJL	bivalent	bivalent	bivalent	bivalent
CD38	bivalent	bivalent	N/A	N/A
CD70	bivalent	bivalent	N/A	N/A
CDH11	bivalent	bivalent	N/A	N/A
CDHR1	bivalent	bivalent	N/A	N/A
CDK5R2	bivalent	bivalent	bivalent	bivalent
CDKN1C	bivalent	bivalent	N/A	N/A
CDO1	bivalent	bivalent	bivalent	bivalent
CDON	bivalent	bivalent	N/A	N/A
CDYL2	bivalent	bivalent	bivalent	bivalent
CEBPA	bivalent	bivalent	N/A	N/A
CELF4	bivalent	bivalent	N/A	N/A
CELSR3	bivalent	bivalent	bivalent	bivalent
CERKL	bivalent	bivalent	bivalent	active only
CERS1	bivalent	bivalent	N/A	N/A
CERS4	bivalent	bivalent	N/A	N/A
CES1P1	bivalent	no mark	N/A	N/A
CGNL1	bivalent	bivalent	bivalent	bivalent
CGREF1	bivalent	bivalent	N/A	N/A
CH25H	bivalent	bivalent	bivalent	bivalent
CHDH	bivalent	bivalent	bivalent	active only
CHGB	bivalent	bivalent	N/A	N/A
CHN2	bivalent	bivalent	bivalent	bivalent
CHRNA7	bivalent	bivalent	bivalent	bivalent
CHST1	bivalent	bivalent	N/A	N/A
CHSY3	bivalent	bivalent	N/A	N/A
CIITA	bivalent	no mark	N/A	N/A
CITED1	bivalent	bivalent	bivalent	bivalent
CKB	bivalent	bivalent	N/A	N/A
CLEC11A	bivalent	bivalent	N/A	N/A
CLIC6	bivalent	bivalent	bivalent	bivalent
CLSTN2	bivalent	bivalent	bivalent	bivalent
CLUL1	bivalent	bivalent	N/A	N/A
CMPK2	bivalent	bivalent	N/A	N/A
CNIH3	bivalent	bivalent	bivalent	bivalent
CNNM1	bivalent	bivalent	bivalent	bivalent
CNR1	bivalent	bivalent	bivalent	bivalent

CNTN1	bivalent	repressive only	bivalent	bivalent
COBL	bivalent	bivalent	bivalent	bivalent
COCH	bivalent	bivalent	bivalent	bivalent
COL14A1	bivalent	bivalent	bivalent	bivalent
COL19A1	bivalent	bivalent	bivalent	bivalent
COL23A1	bivalent	bivalent	bivalent	bivalent
COL25A1	bivalent	bivalent	bivalent	bivalent
COL27A1	bivalent	bivalent	bivalent	bivalent
COL2A1	bivalent	repressive only	bivalent	bivalent
COL4A3	bivalent	bivalent	bivalent	bivalent
COL4A4	bivalent	bivalent	bivalent	bivalent
COL5A3	bivalent	bivalent	N/A	N/A
COL9A2	bivalent	bivalent	bivalent	bivalent
CPNE5	bivalent	bivalent	bivalent	bivalent
CPNE9	bivalent	bivalent	bivalent	bivalent
CPT1A	bivalent	bivalent	N/A	N/A
CPVL	bivalent	bivalent	bivalent	active only
CR2	bivalent	repressive only	bivalent	active only
CRH	bivalent	bivalent	N/A	N/A
CRNDE	bivalent	bivalent	N/A	N/A
CRYAB	bivalent	repressive only	N/A	N/A
CRYBA2	bivalent	bivalent	bivalent	bivalent
CSDAP1	bivalent	bivalent	N/A	N/A
CSMD3	bivalent	bivalent	bivalent	bivalent
CSPG5	bivalent	bivalent	bivalent	bivalent
CTNND2	bivalent	bivalent	bivalent	bivalent
CTSH	bivalent	bivalent	N/A	N/A
CTTNBP2	bivalent	bivalent	bivalent	bivalent
CWH43	bivalent	bivalent	bivalent	bivalent
CXCL14	bivalent	bivalent	N/A	N/A
CXCL16	bivalent	bivalent	N/A	N/A
CYFIP2	bivalent	bivalent	N/A	N/A
CYP1B1	bivalent	bivalent	bivalent	bivalent
CYP26B1	bivalent	bivalent	bivalent	bivalent
CYP2E1	bivalent	bivalent	N/A	N/A
CYP2J2	bivalent	bivalent	bivalent	bivalent
CYP2S1	bivalent	bivalent	N/A	N/A
CYP46A1	bivalent	bivalent	N/A	N/A
CYP7B1	bivalent	bivalent	N/A	N/A
CYS1	bivalent	bivalent	bivalent	bivalent
D4S234E	bivalent	bivalent	bivalent	bivalent
DAAM2	bivalent	bivalent	bivalent	bivalent
DACH1	bivalent	bivalent	N/A	N/A
DACH2	bivalent	bivalent	bivalent	bivalent
DACT1	bivalent	bivalent	N/A	N/A
DAPK1	bivalent	bivalent	N/A	N/A

DCLK1	bivalent	bivalent	N/A	N/A
DDX25	bivalent	bivalent	bivalent	bivalent
DEGS2	bivalent	bivalent	bivalent	bivalent
DFNB31	bivalent	bivalent	N/A	N/A
DGCR5	bivalent	bivalent	N/A	N/A
DGKG	bivalent	bivalent	bivalent	bivalent
DHH	bivalent	bivalent	N/A	N/A
DIRAS2	bivalent	bivalent	bivalent	bivalent
DLL4	bivalent	bivalent	bivalent	active only
DLX3	bivalent	bivalent	bivalent	bivalent
DLX4	bivalent	bivalent	bivalent	bivalent
DNAH11	bivalent	bivalent	bivalent	bivalent
DNAH8	bivalent	repressive only	N/A	N/A
DNAJC22	bivalent	bivalent	bivalent	active only
DOC2GP	bivalent	no mark	N/A	N/A
DOCK3	bivalent	bivalent	N/A	N/A
DPY19L2	bivalent	bivalent	N/A	N/A
DPY19L2P1	bivalent	bivalent	N/A	N/A
DPY19L2P2	bivalent	bivalent	bivalent	bivalent
DPY19L2P4	bivalent	bivalent	bivalent	bivalent
DPYSL2	bivalent	bivalent	N/A	N/A
DPYSL4	bivalent	bivalent	bivalent	bivalent
DPYSL5	bivalent	bivalent	bivalent	bivalent
DRD2	bivalent	bivalent	bivalent	bivalent
DUSP15	bivalent	bivalent	bivalent	bivalent
DUSP2	bivalent	bivalent	N/A	N/A
DUSP8	bivalent	bivalent	N/A	N/A
EDIL3	bivalent	bivalent	bivalent	bivalent
EDNRB	bivalent	bivalent	bivalent	bivalent
EFHD1	bivalent	bivalent	bivalent	bivalent
EFNA3	bivalent	bivalent	N/A	N/A
EFNB3	bivalent	bivalent	bivalent	bivalent
EFS	bivalent	bivalent	N/A	N/A
EGFEM1P	bivalent	bivalent	N/A	N/A
EGFLAM	bivalent	bivalent	bivalent	bivalent
EGLN3	bivalent	bivalent	N/A	N/A
EGR2	bivalent	bivalent	bivalent	bivalent
EGR3	bivalent	bivalent	bivalent	bivalent
EGR4	bivalent	bivalent	bivalent	bivalent
EIF4E3	bivalent	bivalent	bivalent	bivalent
ELFN2	bivalent	bivalent	N/A	N/A
ELMOD1	bivalent	bivalent	bivalent	bivalent
ELOVL2	bivalent	bivalent	bivalent	bivalent
EML6	bivalent	bivalent	bivalent	bivalent
EMX1	bivalent	bivalent	bivalent	bivalent
EN1	bivalent	bivalent	N/A	N/A

ENPEP	bivalent	bivalent	N/A	N/A
EOMES	bivalent	bivalent	bivalent	bivalent
EPB41L4A	bivalent	bivalent	N/A	N/A
EPHA4	bivalent	bivalent	N/A	N/A
EPHA5	bivalent	bivalent	bivalent	bivalent
EPHB1	bivalent	bivalent	bivalent	active only
EPHB6	bivalent	bivalent	N/A	N/A
ERBB4	bivalent	bivalent	bivalent	bivalent
ERC2	bivalent	bivalent	N/A	N/A
ERG	bivalent	bivalent	bivalent	bivalent
ERO1LB	bivalent	bivalent	N/A	N/A
ESPN	bivalent	bivalent	bivalent	bivalent
ESX1	bivalent	bivalent	bivalent	bivalent
EYA4	bivalent	bivalent	bivalent	bivalent
FAM105A	bivalent	bivalent	N/A	N/A
FAM110B	bivalent	bivalent	N/A	N/A
FAM124A	bivalent	bivalent	N/A	N/A
FAM131B	bivalent	bivalent	bivalent	bivalent
FAM133A	bivalent	bivalent	bivalent	bivalent
FAM149A	bivalent	bivalent	N/A	N/A
FAM150A	bivalent	bivalent	bivalent	bivalent
FAM150B	bivalent	bivalent	bivalent	bivalent
FAM153C	bivalent	bivalent	bivalent	bivalent
FAM155A	bivalent	bivalent	bivalent	bivalent
FAM155B	bivalent	repressive only	bivalent	bivalent
FAM159A	bivalent	bivalent	bivalent	bivalent
FAM159B	bivalent	bivalent	bivalent	bivalent
FAM162B	bivalent	bivalent	bivalent	bivalent
FAM181B	bivalent	bivalent	bivalent	bivalent
FAM184A	bivalent	bivalent	N/A	N/A
FAM184B	bivalent	bivalent	bivalent	bivalent
FAM190A	bivalent	bivalent	bivalent	bivalent
FAM196A	bivalent	bivalent	bivalent	bivalent
FAM198A	bivalent	bivalent	bivalent	bivalent
FAM19A2	bivalent	bivalent	bivalent	bivalent
FAM20A	bivalent	bivalent	bivalent	bivalent
FAM211A	bivalent	bivalent	N/A	N/A
FAM43A	bivalent	bivalent	N/A	N/A
FAM46C	bivalent	bivalent	N/A	N/A
FAM47E-STBD1	bivalent	bivalent	N/A	N/A
FAM49A	bivalent	bivalent	N/A	N/A
FAM78A	bivalent	bivalent	bivalent	bivalent
FAM78B	bivalent	bivalent	N/A	N/A
FAM81A	bivalent	bivalent	N/A	N/A
FARP1	bivalent	bivalent	N/A	N/A
FBLN5	bivalent	bivalent	N/A	N/A

FBLN7	bivalent	bivalent	bivalent	bivalent
FBP1	bivalent	bivalent	N/A	N/A
FBXL15	bivalent	bivalent	bivalent	bivalent
FBXL7	bivalent	bivalent	N/A	N/A
FGF12	bivalent	bivalent	bivalent	bivalent
FGF14	bivalent	bivalent	bivalent	bivalent
FGF9	bivalent	bivalent	bivalent	bivalent
FGFR3	bivalent	bivalent	N/A	N/A
FIBCD1	bivalent	bivalent	bivalent	bivalent
FIBIN	bivalent	bivalent	N/A	N/A
FLJ11235	bivalent	bivalent	N/A	N/A
FLJ23152	bivalent	bivalent	N/A	N/A
FLJ35024	bivalent	bivalent	bivalent	bivalent
FLJ43390	bivalent	bivalent	bivalent	bivalent
FLJ44511	bivalent	bivalent	N/A	N/A
FNDC1	bivalent	bivalent	bivalent	bivalent
FNDC5	bivalent	bivalent	N/A	N/A
FOXA1	bivalent	bivalent	bivalent	bivalent
FOXA2	bivalent	repressive only	bivalent	active only
FOXA3	bivalent	bivalent	N/A	N/A
FOXB1	bivalent	bivalent	bivalent	bivalent
FOXB2	bivalent	bivalent	bivalent	bivalent
FOXD3	bivalent	bivalent	bivalent	bivalent
FOXE1	bivalent	bivalent	bivalent	bivalent
FOXI3	bivalent	bivalent	N/A	N/A
FOXL2	bivalent	bivalent	bivalent	bivalent
FOXN4	bivalent	bivalent	bivalent	bivalent
FREM2	bivalent	bivalent	bivalent	bivalent
FRMD3	bivalent	bivalent	bivalent	bivalent
FRMPD1	bivalent	bivalent	bivalent	bivalent
GABRA2	bivalent	repressive only	bivalent	bivalent
GAD1	bivalent	bivalent	bivalent	active only
GADD45G	bivalent	bivalent	N/A	N/A
GADL1	bivalent	bivalent	bivalent	active only
GAL	bivalent	bivalent	bivalent	bivalent
GALNT12	bivalent	bivalent	N/A	N/A
GALNTL1	bivalent	bivalent	bivalent	bivalent
GALNTL6	bivalent	bivalent	bivalent	bivalent
GATA2	bivalent	bivalent	bivalent	bivalent
GATA6	bivalent	bivalent	bivalent	active only
GATM	bivalent	bivalent	N/A	N/A
GBP4	bivalent	no mark	N/A	N/A
GBX2	bivalent	bivalent	bivalent	bivalent
GDA	bivalent	repressive only	bivalent	bivalent
GDF1	bivalent	bivalent	bivalent	bivalent
GDNF	bivalent	bivalent	bivalent	bivalent

GFI1	bivalent	bivalent	bivalent	bivalent
GFRA1	bivalent	bivalent	bivalent	bivalent
GGTA1P	bivalent	bivalent	N/A	N/A
GHR	bivalent	bivalent	N/A	N/A
GJB6	bivalent	bivalent	bivalent	bivalent
GKAP1	bivalent	bivalent	N/A	N/A
GLB1L2	bivalent	repressive only	N/A	N/A
GLB1L3	bivalent	bivalent	bivalent	bivalent
GLDC	bivalent	bivalent	bivalent	bivalent
GLS2	bivalent	bivalent	bivalent	bivalent
GLT25D2	bivalent	bivalent	N/A	N/A
GNAO1	bivalent	bivalent	bivalent	bivalent
GNAS	bivalent	bivalent	bivalent	bivalent
GNG4	bivalent	bivalent	bivalent	active only
GOLGA7B	bivalent	bivalent	bivalent	bivalent
GPC5	bivalent	bivalent	bivalent	bivalent
GPR158	bivalent	repressive only	bivalent	active only
GPR160	bivalent	bivalent	N/A	N/A
GPR27	bivalent	bivalent	bivalent	bivalent
GPR37	bivalent	bivalent	bivalent	bivalent
GPR63	bivalent	bivalent	N/A	N/A
GPR64	bivalent	bivalent	bivalent	bivalent
GPR88	bivalent	bivalent	N/A	N/A
GPR98	bivalent	bivalent	bivalent	bivalent
GPRC5B	bivalent	bivalent	bivalent	bivalent
GPRC5C	bivalent	bivalent	bivalent	bivalent
GPRIN2	bivalent	bivalent	bivalent	bivalent
GPRIN3	bivalent	bivalent	N/A	N/A
GREM2	bivalent	bivalent	N/A	N/A
GRIA2	bivalent	bivalent	bivalent	bivalent
GRIN2D	bivalent	bivalent	N/A	N/A
GRM4	bivalent	bivalent	N/A	N/A
GRM8	bivalent	bivalent	bivalent	active only
GSC	bivalent	bivalent	bivalent	bivalent
GSX2	bivalent	bivalent	bivalent	bivalent
GUCY1A2	bivalent	bivalent	bivalent	bivalent
GUCY1A3	bivalent	bivalent	N/A	N/A
GUCY1B3	bivalent	bivalent	N/A	N/A
H2AFY2	bivalent	bivalent	bivalent	bivalent
HAAO	bivalent	bivalent	N/A	N/A
HAP1	bivalent	bivalent	bivalent	bivalent
HAPLN1	bivalent	bivalent	bivalent	bivalent
HAR1A	bivalent	bivalent	N/A	N/A
HAR1B	bivalent	bivalent	N/A	N/A
HBB	bivalent	repressive only	N/A	N/A
HBM	bivalent	bivalent	bivalent	bivalent

HCN1	bivalent	bivalent	bivalent	bivalent
HCRTR1	bivalent	bivalent	bivalent	bivalent
HDAC5	bivalent	bivalent	N/A	N/A
HERC5	bivalent	bivalent	N/A	N/A
HES4	bivalent	bivalent	bivalent	bivalent
HES6	bivalent	bivalent	N/A	N/A
HEY2	bivalent	bivalent	bivalent	bivalent
HFM1	bivalent	bivalent	N/A	N/A
HLA-DQB1	bivalent	repressive only	N/A	N/A
HMHA1	bivalent	repressive only	bivalent	bivalent
HMX2	bivalent	bivalent	N/A	N/A
HOTTIP	bivalent	bivalent	N/A	N/A
HOXA11	bivalent	bivalent	N/A	N/A
HOXA13	bivalent	bivalent	bivalent	bivalent
HOXA-AS5	bivalent	bivalent	N/A	N/A
HOXB7	bivalent	bivalent	N/A	N/A
HOXB8	bivalent	repressive only	bivalent	active only
HOXB9	bivalent	bivalent	bivalent	active only
HOXD1	bivalent	bivalent	N/A	N/A
HOXD8	bivalent	bivalent	bivalent	bivalent
HOXD-AS1	bivalent	bivalent	N/A	N/A
HOXD-AS2	bivalent	bivalent	N/A	N/A
HPCAL4	bivalent	bivalent	bivalent	bivalent
HRASLS	bivalent	bivalent	N/A	N/A
HRK	bivalent	bivalent	bivalent	bivalent
HS6ST3	bivalent	bivalent	bivalent	active only
HSD11B2	bivalent	bivalent	N/A	N/A
HSF4	bivalent	bivalent	bivalent	bivalent
HSPA12A	bivalent	bivalent	bivalent	bivalent
HTR6	bivalent	bivalent	bivalent	bivalent
HUNK	bivalent	bivalent	bivalent	active only
HVCN1	bivalent	bivalent	N/A	N/A
ICA1	bivalent	bivalent	bivalent	active only
ICAM5	bivalent	bivalent	bivalent	bivalent
ICOSLG	bivalent	bivalent	N/A	N/A
ID2	bivalent	bivalent	N/A	N/A
IGDCC3	bivalent	bivalent	bivalent	bivalent
IGDCC4	bivalent	bivalent	N/A	N/A
IGF2BP3	bivalent	bivalent	bivalent	active only
IGFBP3	bivalent	bivalent	bivalent	bivalent
IGFBP5	bivalent	bivalent	N/A	N/A
IGFBPL1	bivalent	bivalent	bivalent	bivalent
IGSF9B	bivalent	bivalent	N/A	N/A
IL10RA	bivalent	bivalent	N/A	N/A
IL17RB	bivalent	bivalent	bivalent	active only
IL1RAPL2	bivalent	bivalent	N/A	N/A

IL20RA	bivalent	bivalent	bivalent	active only
ILDR2	bivalent	bivalent	bivalent	bivalent
INHBB	bivalent	bivalent	bivalent	bivalent
INSM2	bivalent	bivalent	bivalent	bivalent
INSR	bivalent	bivalent	N/A	N/A
INSRR	bivalent	bivalent	bivalent	bivalent
INTS4L1	bivalent	bivalent	bivalent	bivalent
INTS4L2	bivalent	bivalent	bivalent	bivalent
IQGAP2	bivalent	bivalent	N/A	N/A
IQUB	bivalent	bivalent	N/A	N/A
IRAK3	bivalent	bivalent	bivalent	bivalent
IRS4	bivalent	bivalent	bivalent	bivalent
IRX3	bivalent	bivalent	bivalent	bivalent
IRX5	bivalent	bivalent	bivalent	bivalent
ISL1	bivalent	bivalent	bivalent	bivalent
ISL2	bivalent	bivalent	bivalent	bivalent
ISLR2	bivalent	repressive only	bivalent	bivalent
ISM1	bivalent	bivalent	bivalent	bivalent
ISM2	bivalent	bivalent	bivalent	bivalent
ITGA11	bivalent	bivalent	bivalent	bivalent
ITGA9	bivalent	bivalent	N/A	N/A
JAM2	bivalent	bivalent	bivalent	bivalent
KBTBD11	bivalent	bivalent	bivalent	bivalent
KCNA2	bivalent	bivalent	bivalent	bivalent
KCNB2	bivalent	bivalent	bivalent	bivalent
KCNC4	bivalent	bivalent	N/A	N/A
KCND2	bivalent	bivalent	bivalent	bivalent
KCND3	bivalent	bivalent	bivalent	bivalent
KCNG1	bivalent	bivalent	bivalent	bivalent
KCNG3	bivalent	bivalent	bivalent	bivalent
KCNH3	bivalent	bivalent	bivalent	bivalent
KCNH5	bivalent	bivalent	N/A	N/A
KCNJ11	bivalent	bivalent	bivalent	bivalent
KCNJ12	bivalent	bivalent	bivalent	bivalent
KCNJ2	bivalent	bivalent	N/A	N/A
KCNJ2-AS1	bivalent	bivalent	N/A	N/A
KCNJ6	bivalent	bivalent	N/A	N/A
KCNK4	bivalent	bivalent	bivalent	bivalent
KCNMB4	bivalent	bivalent	bivalent	bivalent
KCNS2	bivalent	bivalent	bivalent	bivalent
KCNS3	bivalent	bivalent	bivalent	bivalent
KCNV1	bivalent	bivalent	bivalent	bivalent
KCTD8	bivalent	bivalent	bivalent	active only
KHDRBS3	bivalent	bivalent	N/A	N/A
KIAA0226L	bivalent	bivalent	N/A	N/A
KIAA1024	bivalent	bivalent	bivalent	bivalent

KIAA1045	bivalent	bivalent	bivalent	bivalent
KIAA1199	bivalent	bivalent	bivalent	active only
KIAA1244	bivalent	bivalent	bivalent	active only
KIAA1274	bivalent	bivalent	N/A	N/A
KIAA1324	bivalent	bivalent	bivalent	active only
KIF21B	bivalent	bivalent	N/A	N/A
KIF26B	bivalent	bivalent	bivalent	active only
KIRREL3	bivalent	bivalent	bivalent	bivalent
KIT	bivalent	bivalent	bivalent	bivalent
KLF14	bivalent	bivalent	bivalent	bivalent
KLF15	bivalent	bivalent	bivalent	bivalent
KLF2	bivalent	bivalent	N/A	N/A
KLHL14	bivalent	repressive only	N/A	N/A
KLHL32	bivalent	bivalent	N/A	N/A
KLHL35	bivalent	bivalent	bivalent	bivalent
KLRG2	bivalent	bivalent	bivalent	bivalent
LAMA1	bivalent	bivalent	bivalent	bivalent
LAMA2	bivalent	bivalent	N/A	N/A
LANCL3	bivalent	bivalent	bivalent	bivalent
LBH	bivalent	bivalent	N/A	N/A
LEF1	bivalent	bivalent	bivalent	bivalent
LGI2	bivalent	bivalent	bivalent	bivalent
LGI3	bivalent	bivalent	bivalent	bivalent
LHX1	bivalent	bivalent	bivalent	bivalent
LHX4	bivalent	bivalent	bivalent	bivalent
LHX9	bivalent	bivalent	bivalent	bivalent
LIMCH1	bivalent	bivalent	N/A	N/A
LINC00092	bivalent	bivalent	N/A	N/A
LINC00246B	bivalent	repressive only	N/A	N/A
LINC00340	bivalent	bivalent	N/A	N/A
LINC00466	bivalent	bivalent	N/A	N/A
LINC00478	bivalent	no mark	N/A	N/A
LINGO1	bivalent	bivalent	bivalent	bivalent
LMF1	bivalent	bivalent	N/A	N/A
LMTK3	bivalent	bivalent	bivalent	bivalent
LMX1B	bivalent	bivalent	bivalent	bivalent
LOC100127983	bivalent	bivalent	N/A	N/A
LOC100128811	bivalent	repressive only	bivalent	bivalent
LOC100129620	bivalent	bivalent	N/A	N/A
LOC100131320	bivalent	bivalent	N/A	N/A
LOC100132891	bivalent	bivalent	N/A	N/A
LOC100144602	bivalent	bivalent	N/A	N/A
LOC100190938	bivalent	bivalent	bivalent	bivalent
LOC100289656	bivalent	bivalent	N/A	N/A
LOC100499405	bivalent	bivalent	N/A	N/A
LOC100505483	bivalent	bivalent	N/A	N/A

LOC100506409	bivalent	bivalent	N/A	N/A
LOC100506779	bivalent	bivalent	N/A	N/A
LOC100507050	bivalent	bivalent	N/A	N/A
LOC100507218	bivalent	bivalent	N/A	N/A
LOC100507588	bivalent	bivalent	N/A	N/A
LOC100652791	bivalent	bivalent	N/A	N/A
LOC145663	bivalent	bivalent	N/A	N/A
LOC149773	bivalent	bivalent	N/A	N/A
LOC202181	bivalent	bivalent	N/A	N/A
LOC253962	bivalent	bivalent	N/A	N/A
LOC283143	bivalent	bivalent	N/A	N/A
LOC283194	bivalent	bivalent	N/A	N/A
LOC283731	bivalent	repressive only	bivalent	bivalent
LOC283856	bivalent	bivalent	bivalent	bivalent
LOC284648	bivalent	bivalent	N/A	N/A
LOC285084	bivalent	bivalent	N/A	N/A
LOC286002	bivalent	bivalent	bivalent	bivalent
LOC286189	bivalent	bivalent	N/A	N/A
LOC338651	bivalent	bivalent	N/A	N/A
LOC389458	bivalent	bivalent	N/A	N/A
LOC389634	bivalent	bivalent	bivalent	bivalent
LOC389765	bivalent	bivalent	N/A	N/A
LOC402160	bivalent	bivalent	N/A	N/A
LOC440925	bivalent	bivalent	bivalent	bivalent
LOC572558	bivalent	bivalent	bivalent	bivalent
LOC641518	bivalent	bivalent	bivalent	bivalent
LOC642366	bivalent	bivalent	N/A	N/A
LOC643201	bivalent	bivalent	N/A	N/A
LOC643923	bivalent	bivalent	bivalent	bivalent
LOC646278	bivalent	bivalent	N/A	N/A
LOC728739	bivalent	bivalent	N/A	N/A
LOC80054	bivalent	bivalent	bivalent	bivalent
LOC91149	bivalent	bivalent	bivalent	active only
LPHN1	bivalent	bivalent	N/A	N/A
LPL	bivalent	repressive only	bivalent	bivalent
LPPR1	bivalent	bivalent	bivalent	bivalent
LPPR4	bivalent	bivalent	bivalent	bivalent
LPPR5	bivalent	bivalent	bivalent	bivalent
LRCH2	bivalent	bivalent	bivalent	bivalent
LRFN2	bivalent	bivalent	bivalent	bivalent
LRFN5	bivalent	bivalent	bivalent	bivalent
LRP2	bivalent	bivalent	bivalent	bivalent
LRP3	bivalent	bivalent	N/A	N/A
LRRC16B	bivalent	bivalent	N/A	N/A
LRRC3	bivalent	bivalent	bivalent	bivalent
LRRN2	bivalent	bivalent	N/A	N/A

LUZP2	bivalent	bivalent	bivalent	bivalent
LYPD1	bivalent	bivalent	bivalent	bivalent
MAB21L1	bivalent	bivalent	N/A	N/A
MACF1	bivalent	bivalent	bivalent	bivalent
MAF	bivalent	bivalent	N/A	N/A
MAN1C1	bivalent	bivalent	N/A	N/A
MAP6	bivalent	bivalent	bivalent	bivalent
MAPK4	bivalent	bivalent	bivalent	bivalent
MAPRE2	bivalent	bivalent	N/A	N/A
MAPT	bivalent	bivalent	bivalent	bivalent
MAPT-IT1	bivalent	bivalent	N/A	N/A
MAST3	bivalent	bivalent	N/A	N/A
MC5R	bivalent	bivalent	bivalent	bivalent
MCOLN2	bivalent	bivalent	bivalent	bivalent
MCOLN3	bivalent	bivalent	bivalent	bivalent
MCTP1	bivalent	bivalent	bivalent	bivalent
MDGA1	bivalent	bivalent	N/A	N/A
MDK	bivalent	bivalent	N/A	N/A
MED12L	bivalent	bivalent	bivalent	active only
MEF2C	bivalent	bivalent	N/A	N/A
MEGF11	bivalent	bivalent	N/A	N/A
MERTK	bivalent	bivalent	N/A	N/A
METRNL	bivalent	bivalent	N/A	N/A
MGAT4A	bivalent	bivalent	N/A	N/A
MGC16275	bivalent	bivalent	N/A	N/A
MGC2889	bivalent	bivalent	N/A	N/A
MGC45922	bivalent	bivalent	N/A	N/A
MIR1258	bivalent	bivalent	N/A	N/A
MIR129-2	bivalent	bivalent	bivalent	bivalent
MIR17HG	bivalent	bivalent	N/A	N/A
MIR203	bivalent	repressive only	N/A	N/A
MIR2276	bivalent	bivalent	N/A	N/A
MIR3545	bivalent	repressive only	N/A	N/A
MIR3665	bivalent	bivalent	N/A	N/A
MIR4497	bivalent	bivalent	N/A	N/A
MIR4520A	bivalent	bivalent	N/A	N/A
MIR4520B	bivalent	bivalent	N/A	N/A
MIR4526	bivalent	bivalent	N/A	N/A
MIR4787	bivalent	bivalent	N/A	N/A
MIR9-1	bivalent	bivalent	bivalent	bivalent
MIXL1	bivalent	bivalent	bivalent	bivalent
MLNR	bivalent	bivalent	bivalent	bivalent
MLXIPL	bivalent	bivalent	N/A	N/A
MMP16	bivalent	bivalent	N/A	N/A
MMP25	bivalent	bivalent	bivalent	bivalent
MN1	bivalent	bivalent	bivalent	bivalent

MPP2	bivalent	bivalent	N/A	N/A
MSC	bivalent	bivalent	bivalent	bivalent
MSX2	bivalent	bivalent	bivalent	bivalent
MT1G	bivalent	repressive only	N/A	N/A
MT1H	bivalent	no mark	N/A	N/A
MT1JP	bivalent	bivalent	N/A	N/A
MT1M	bivalent	bivalent	N/A	N/A
MTNR1B	bivalent	repressive only	N/A	N/A
MYH11	bivalent	bivalent	N/A	N/A
MYH14	bivalent	bivalent	bivalent	active only
MYO15B	bivalent	bivalent	N/A	N/A
MYO5B	bivalent	bivalent	bivalent	active only
MYRIP	bivalent	bivalent	bivalent	bivalent
N4BP3	bivalent	bivalent	bivalent	bivalent
NAALAD2	bivalent	bivalent	N/A	N/A
NAP1L3	bivalent	bivalent	bivalent	bivalent
NBEA	bivalent	bivalent	N/A	N/A
NCALD	bivalent	bivalent	N/A	N/A
NCAM1	bivalent	bivalent	N/A	N/A
NCAM2	bivalent	bivalent	N/A	N/A
NDNF	bivalent	bivalent	N/A	N/A
NDRG2	bivalent	bivalent	N/A	N/A
NDRG4	bivalent	bivalent	bivalent	bivalent
NDST3	bivalent	bivalent	bivalent	bivalent
NECAB1	bivalent	bivalent	bivalent	bivalent
NEFL	bivalent	bivalent	bivalent	bivalent
NEFM	bivalent	bivalent	bivalent	bivalent
NELL2	bivalent	bivalent	bivalent	bivalent
NEO1	bivalent	bivalent	N/A	N/A
NETO1	bivalent	bivalent	bivalent	bivalent
NEURL	bivalent	bivalent	bivalent	bivalent
NEURL1B	bivalent	bivalent	N/A	N/A
NEUROG2	bivalent	bivalent	bivalent	bivalent
NFATC1	bivalent	bivalent	bivalent	bivalent
NGFR	bivalent	bivalent	N/A	N/A
NHLH2	bivalent	bivalent	N/A	N/A
NHSL2	bivalent	bivalent	bivalent	bivalent
NKAIN1	bivalent	bivalent	N/A	N/A
NKAIN2	bivalent	bivalent	bivalent	bivalent
NKAIN3	bivalent	bivalent	bivalent	bivalent
NKX2-8	bivalent	bivalent	bivalent	bivalent
NKX3-2	bivalent	bivalent	bivalent	bivalent
NKX6-1	bivalent	bivalent	bivalent	bivalent
NMNAT2	bivalent	bivalent	bivalent	bivalent
NOL4	bivalent	bivalent	bivalent	bivalent
NOTCH3	bivalent	bivalent	bivalent	bivalent

NOVA1	bivalent	bivalent	N/A	N/A
NPAS3	bivalent	bivalent	N/A	N/A
NPBWR1	bivalent	bivalent	bivalent	bivalent
NPFFR2	bivalent	bivalent	N/A	N/A
NPNT	bivalent	bivalent	bivalent	bivalent
NPPC	bivalent	bivalent	bivalent	bivalent
NPR1	bivalent	bivalent	bivalent	bivalent
NPR3	bivalent	bivalent	bivalent	bivalent
NPTXR	bivalent	bivalent	N/A	N/A
NPY1R	bivalent	bivalent	bivalent	bivalent
NR3C2	bivalent	bivalent	N/A	N/A
NRARP	bivalent	bivalent	N/A	N/A
NRCAM	bivalent	bivalent	bivalent	active only
NRIP3	bivalent	bivalent	N/A	N/A
NRN1	bivalent	bivalent	bivalent	bivalent
NRXN2	bivalent	bivalent	N/A	N/A
NSUN7	bivalent	bivalent	bivalent	bivalent
NT5C1A	bivalent	bivalent	bivalent	bivalent
NTN1	bivalent	bivalent	bivalent	bivalent
NTNG2	bivalent	bivalent	bivalent	bivalent
NTRK1	bivalent	bivalent	bivalent	bivalent
NTRK2	bivalent	bivalent	bivalent	bivalent
NUAK1	bivalent	bivalent	N/A	N/A
NUDT10	bivalent	bivalent	bivalent	bivalent
NUP210	bivalent	bivalent	N/A	N/A
NXPH2	bivalent	repressive only	bivalent	bivalent
NXPH3	bivalent	bivalent	N/A	N/A
NXPH4	bivalent	bivalent	bivalent	bivalent
OGDHL	bivalent	bivalent	bivalent	bivalent
ONECUT1	bivalent	bivalent	bivalent	bivalent
ONECUT2	bivalent	bivalent	bivalent	bivalent
OVOL1	bivalent	repressive only	bivalent	active only
OXCT2	bivalent	bivalent	bivalent	bivalent
OXGR1	bivalent	bivalent	bivalent	active only
PADI2	bivalent	bivalent	N/A	N/A
PAK7	bivalent	bivalent	N/A	N/A
PAPLN	bivalent	bivalent	bivalent	bivalent
PAQR9	bivalent	bivalent	bivalent	bivalent
PARM1	bivalent	bivalent	N/A	N/A
PAX6	bivalent	bivalent	bivalent	active only
PCDH10	bivalent	bivalent	bivalent	bivalent
PCDH17	bivalent	bivalent	bivalent	bivalent
PCDH20	bivalent	bivalent	bivalent	bivalent
PCDHB1	bivalent	repressive only	N/A	N/A
PCDP1	bivalent	bivalent	bivalent	bivalent
PCP4L1	bivalent	bivalent	bivalent	bivalent

PCSK5	bivalent	bivalent	N/A	N/A
PCSK9	bivalent	bivalent	N/A	N/A
PDE11A	bivalent	bivalent	bivalent	bivalent
PDE1B	bivalent	bivalent	N/A	N/A
PDE3B	bivalent	bivalent	N/A	N/A
PDE8B	bivalent	bivalent	bivalent	bivalent
PDE9A	bivalent	bivalent	N/A	N/A
PDGFA	bivalent	bivalent	bivalent	bivalent
PDGFRA	bivalent	bivalent	bivalent	bivalent
PDK4	bivalent	bivalent	N/A	N/A
PDLIM3	bivalent	bivalent	bivalent	bivalent
PDPN	bivalent	bivalent	bivalent	bivalent
PDX1	bivalent	bivalent	bivalent	bivalent
PDZRN3	bivalent	bivalent	bivalent	bivalent
PELI2	bivalent	bivalent	N/A	N/A
PEX5L	bivalent	bivalent	N/A	N/A
PGM5	bivalent	bivalent	bivalent	bivalent
PGM5P2	bivalent	bivalent	bivalent	bivalent
PHF21B	bivalent	bivalent	bivalent	bivalent
PHPT1	bivalent	bivalent	N/A	N/A
PHYHIPL	bivalent	bivalent	bivalent	active only
PIK3CD	bivalent	bivalent	bivalent	bivalent
PIK3R5	bivalent	repressive only	bivalent	bivalent
PIP5K1B	bivalent	bivalent	bivalent	bivalent
PKNOX2	bivalent	bivalent	N/A	N/A
PLAC2	bivalent	bivalent	N/A	N/A
PLAG1	bivalent	bivalent	bivalent	bivalent
PLCL1	bivalent	bivalent	N/A	N/A
PLCL2	bivalent	bivalent	N/A	N/A
PLEKHG1	bivalent	bivalent	N/A	N/A
PLXDC1	bivalent	bivalent	N/A	N/A
PLXDC2	bivalent	bivalent	N/A	N/A
PLXNC1	bivalent	bivalent	bivalent	bivalent
PLXND1	bivalent	bivalent	bivalent	bivalent
PNMA2	bivalent	bivalent	bivalent	bivalent
PODN	bivalent	bivalent	N/A	N/A
PODXL2	bivalent	bivalent	bivalent	bivalent
POU3F1	bivalent	bivalent	bivalent	bivalent
POU3F2	bivalent	bivalent	bivalent	bivalent
POU4F1	bivalent	bivalent	bivalent	bivalent
PP7080	bivalent	bivalent	N/A	N/A
PPAP2C	bivalent	bivalent	bivalent	active only
PPFIA3	bivalent	bivalent	N/A	N/A
PPFIBP2	bivalent	bivalent	N/A	N/A
PPM1E	bivalent	bivalent	bivalent	bivalent
PPP1R16B	bivalent	bivalent	bivalent	bivalent

PPP1R9A	bivalent	bivalent	N/A	N/A
PRDM6	bivalent	bivalent	bivalent	bivalent
PRIMA1	bivalent	bivalent	N/A	N/A
PRKAA2	bivalent	bivalent	N/A	N/A
PRKAG2	bivalent	bivalent	N/A	N/A
PRKAR2B	bivalent	bivalent	N/A	N/A
PRKG1	bivalent	bivalent	N/A	N/A
PRMT8	bivalent	bivalent	N/A	N/A
PRODH	bivalent	bivalent	bivalent	bivalent
PROM1	bivalent	bivalent	bivalent	active only
PROX1	bivalent	bivalent	bivalent	active only
PRR18	bivalent	bivalent	bivalent	bivalent
PRR25	bivalent	bivalent	N/A	N/A
PRSS33	bivalent	bivalent	N/A	N/A
PSD	bivalent	bivalent	bivalent	bivalent
PSD2	bivalent	bivalent	bivalent	bivalent
PSMA1	bivalent	bivalent	N/A	N/A
PTCH1	bivalent	bivalent	N/A	N/A
PTCHD2	bivalent	bivalent	bivalent	bivalent
PTGER3	bivalent	bivalent	bivalent	bivalent
PTPRD	bivalent	bivalent	N/A	N/A
PTPRO	bivalent	bivalent	bivalent	active only
PTPRS	bivalent	bivalent	bivalent	bivalent
PURG	bivalent	bivalent	bivalent	bivalent
PUS3	bivalent	bivalent	bivalent	active only
RAB39A	bivalent	bivalent	N/A	N/A
RAB3C	bivalent	bivalent	bivalent	bivalent
RAB3IL1	bivalent	bivalent	bivalent	bivalent
RAB6B	bivalent	bivalent	N/A	N/A
RAB6C	bivalent	bivalent	bivalent	bivalent
RAD21L1	bivalent	bivalent	N/A	N/A
RAMP2	bivalent	bivalent	bivalent	bivalent
RAPGEF4	bivalent	bivalent	bivalent	bivalent
RARB	bivalent	bivalent	N/A	N/A
RASD1	bivalent	bivalent	N/A	N/A
RASGEF1B	bivalent	bivalent	N/A	N/A
RASGRF2	bivalent	bivalent	bivalent	bivalent
RASGRP1	bivalent	bivalent	bivalent	bivalent
RASSF2	bivalent	bivalent	N/A	N/A
RBM20	bivalent	bivalent	bivalent	bivalent
RBP7	bivalent	bivalent	bivalent	bivalent
RBPMS2	bivalent	bivalent	N/A	N/A
REEP1	bivalent	bivalent	bivalent	bivalent
RFPL2	bivalent	bivalent	N/A	N/A
RFX6	bivalent	repressive only	bivalent	bivalent
RGAG4	bivalent	bivalent	bivalent	bivalent

RGMA	bivalent	bivalent	bivalent	bivalent
RGS22	bivalent	repressive only	bivalent	bivalent
RGS7BP	bivalent	repressive only	bivalent	bivalent
RGS9BP	bivalent	bivalent	N/A	N/A
RHBDL1	bivalent	repressive only	N/A	N/A
RHBDL3	bivalent	bivalent	bivalent	bivalent
RHOU	bivalent	bivalent	N/A	N/A
RIC3	bivalent	bivalent	bivalent	bivalent
RIMBP3B	bivalent	bivalent	N/A	N/A
RIMBP3C	bivalent	bivalent	N/A	N/A
RIMKLA	bivalent	bivalent	N/A	N/A
RND1	bivalent	bivalent	N/A	N/A
RNF150	bivalent	bivalent	N/A	N/A
RNF157	bivalent	bivalent	N/A	N/A
RNF165	bivalent	bivalent	bivalent	bivalent
RNF180	bivalent	bivalent	bivalent	bivalent
RNF182	bivalent	bivalent	bivalent	active only
ROBO2	bivalent	bivalent	bivalent	bivalent
ROPN1	bivalent	bivalent	bivalent	bivalent
ROPN1B	bivalent	bivalent	bivalent	bivalent
ROR2	bivalent	bivalent	bivalent	bivalent
RORB	bivalent	bivalent	bivalent	bivalent
RPL38	bivalent	bivalent	bivalent	bivalent
RPP25	bivalent	bivalent	N/A	N/A
RPRML	bivalent	bivalent	bivalent	active only
RRAGD	bivalent	bivalent	bivalent	bivalent
RSPO3	bivalent	bivalent	bivalent	bivalent
RTBDN	bivalent	bivalent	N/A	N/A
RTN1	bivalent	bivalent	bivalent	bivalent
RTN4RL1	bivalent	bivalent	bivalent	bivalent
RTN4RL2	bivalent	bivalent	bivalent	bivalent
RUNDC3A	bivalent	bivalent	N/A	N/A
RXFP3	bivalent	bivalent	N/A	N/A
RXRG	bivalent	repressive only	N/A	N/A
RYR3	bivalent	repressive only	N/A	N/A
SALL4	bivalent	bivalent	N/A	N/A
SAMD14	bivalent	bivalent	N/A	N/A
SAMD5	bivalent	bivalent	bivalent	bivalent
SATB1	bivalent	bivalent	N/A	N/A
SBK1	bivalent	bivalent	N/A	N/A
SCAMP5	bivalent	bivalent	bivalent	bivalent
SCML2	bivalent	bivalent	bivalent	bivalent
SCN1B	bivalent	bivalent	bivalent	bivalent
SCN3B	bivalent	bivalent	bivalent	bivalent
SCN5A	bivalent	bivalent	bivalent	bivalent
SCNN1B	bivalent	bivalent	bivalent	bivalent

SCNN1G	bivalent	bivalent	bivalent	bivalent
SCRT1	bivalent	bivalent	bivalent	bivalent
SCUBE2	bivalent	bivalent	N/A	N/A
SCUBE3	bivalent	bivalent	bivalent	bivalent
SDC2	bivalent	bivalent	N/A	N/A
SDK1	bivalent	bivalent	bivalent	bivalent
SDK2	bivalent	bivalent	bivalent	bivalent
SELV	bivalent	bivalent	bivalent	bivalent
SEMA6A	bivalent	bivalent	N/A	N/A
SEMA6D	bivalent	bivalent	N/A	N/A
SETD7	bivalent	bivalent	N/A	N/A
SEZ6L2	bivalent	bivalent	bivalent	active only
SFMBT2	bivalent	bivalent	bivalent	bivalent
SGCZ	bivalent	bivalent	N/A	N/A
SGPP2	bivalent	repressive only	bivalent	active only
SGSM1	bivalent	bivalent	N/A	N/A
SH3GL2	bivalent	bivalent	bivalent	bivalent
SHC3	bivalent	bivalent	bivalent	bivalent
SHE	bivalent	bivalent	bivalent	bivalent
SHISA2	bivalent	bivalent	bivalent	bivalent
SHROOM2	bivalent	bivalent	bivalent	bivalent
SIM2	bivalent	bivalent	bivalent	active only
SIX1	bivalent	bivalent	bivalent	active only
SLAIN1	bivalent	bivalent	bivalent	bivalent
SLC12A7	bivalent	bivalent	N/A	N/A
SLC16A10	bivalent	bivalent	bivalent	bivalent
SLC16A6	bivalent	bivalent	bivalent	bivalent
SLC16A9	bivalent	bivalent	N/A	N/A
SLC17A7	bivalent	bivalent	bivalent	bivalent
SLC1A1	bivalent	bivalent	N/A	N/A
SLC22A17	bivalent	bivalent	N/A	N/A
SLC22A3	bivalent	bivalent	bivalent	bivalent
SLC25A48	bivalent	bivalent	N/A	N/A
SLC26A4	bivalent	bivalent	bivalent	active only
SLC27A6	bivalent	bivalent	N/A	N/A
SLC30A10	bivalent	bivalent	bivalent	bivalent
SLC30A3	bivalent	bivalent	bivalent	bivalent
SLC35D3	bivalent	bivalent	bivalent	bivalent
SLC35F1	bivalent	bivalent	bivalent	bivalent
SLC40A1	bivalent	bivalent	N/A	N/A
SLC4A4	bivalent	bivalent	bivalent	bivalent
SLC5A8	bivalent	repressive only	N/A	N/A
SLC6A2	bivalent	repressive only	N/A	N/A
SLC6A20	bivalent	bivalent	bivalent	bivalent
SLC8A3	bivalent	bivalent	bivalent	bivalent
SLC9A2	bivalent	bivalent	bivalent	bivalent

SLCO4A1	bivalent	bivalent	bivalent	active only
SLCO5A1	bivalent	bivalent	bivalent	active only
SLITRK1	bivalent	bivalent	N/A	N/A
SMAD6	bivalent	bivalent	N/A	N/A
SMAD9	bivalent	bivalent	N/A	N/A
SMO	bivalent	bivalent	N/A	N/A
SMPD3	bivalent	bivalent	N/A	N/A
SNAI3	bivalent	bivalent	bivalent	bivalent
SNAP25	bivalent	bivalent	N/A	N/A
SNAP91	bivalent	bivalent	N/A	N/A
SNCAIP	bivalent	bivalent	N/A	N/A
SNTB1	bivalent	bivalent	bivalent	active only
SNTG1	bivalent	bivalent	N/A	N/A
SNX22	bivalent	bivalent	bivalent	bivalent
SNX32	bivalent	bivalent	N/A	N/A
SOBP	bivalent	bivalent	bivalent	bivalent
SOX21	bivalent	bivalent	bivalent	bivalent
SOX5	bivalent	bivalent	bivalent	bivalent
SP5	bivalent	bivalent	bivalent	bivalent
SP8	bivalent	bivalent	bivalent	bivalent
SP9	bivalent	bivalent	bivalent	bivalent
SPATA13	bivalent	bivalent	N/A	N/A
SPOCK2	bivalent	bivalent	bivalent	bivalent
SPOCK3	bivalent	bivalent	N/A	N/A
SPTSSB	bivalent	bivalent	N/A	N/A
SRCIN1	bivalent	bivalent	N/A	N/A
SRP68	bivalent	bivalent	bivalent	bivalent
SSBP2	bivalent	bivalent	N/A	N/A
ST14	bivalent	bivalent	bivalent	active only
ST6GALNAC5	bivalent	bivalent	bivalent	bivalent
ST8SIA1	bivalent	bivalent	N/A	N/A
ST8SIA4	bivalent	bivalent	bivalent	bivalent
STK32A	bivalent	bivalent	N/A	N/A
STK32C	bivalent	bivalent	N/A	N/A
STMN3	bivalent	bivalent	N/A	N/A
STOX1	bivalent	bivalent	N/A	N/A
STOX2	bivalent	bivalent	bivalent	bivalent
STXBP5L	bivalent	bivalent	N/A	N/A
SULT4A1	bivalent	bivalent	bivalent	bivalent
SUSD4	bivalent	bivalent	bivalent	bivalent
SUZ12P	bivalent	bivalent	N/A	N/A
SV2B	bivalent	bivalent	N/A	N/A
SVEP1	bivalent	bivalent	bivalent	bivalent
SYMPK	bivalent	bivalent	N/A	N/A
SYN2	bivalent	bivalent	bivalent	bivalent
SYN3	bivalent	repressive only	bivalent	bivalent

SYNE1	bivalent	bivalent	N/A	N/A
SYNGR3	bivalent	bivalent	bivalent	bivalent
SYPL2	bivalent	bivalent	N/A	N/A
SYT10	bivalent	bivalent	bivalent	bivalent
SYT6	bivalent	bivalent	N/A	N/A
SYT7	bivalent	bivalent	bivalent	active only
TACR1	bivalent	bivalent	bivalent	bivalent
TBC1D1	bivalent	bivalent	N/A	N/A
TBC1D30	bivalent	bivalent	N/A	N/A
TBKBP1	bivalent	bivalent	N/A	N/A
TBR1	bivalent	bivalent	bivalent	bivalent
TBX2	bivalent	bivalent	bivalent	bivalent
TCEAL2	bivalent	bivalent	bivalent	bivalent
TCF24	bivalent	bivalent	N/A	N/A
TCHH	bivalent	bivalent	N/A	N/A
TCTE1	bivalent	bivalent	bivalent	bivalent
TDRD10	bivalent	bivalent	bivalent	bivalent
TEKT3	bivalent	bivalent	N/A	N/A
TESC	bivalent	bivalent	bivalent	active only
THSD7A	bivalent	bivalent	N/A	N/A
TIAM1	bivalent	bivalent	bivalent	bivalent
TIGD3	bivalent	bivalent	N/A	N/A
TLL2	bivalent	bivalent	bivalent	bivalent
TM6SF1	bivalent	bivalent	bivalent	bivalent
TMEFF2	bivalent	bivalent	bivalent	bivalent
TMEM108	bivalent	bivalent	bivalent	bivalent
TMEM121	bivalent	bivalent	N/A	N/A
TMEM150C	bivalent	bivalent	bivalent	bivalent
TMEM151B	bivalent	bivalent	bivalent	bivalent
TMEM163	bivalent	bivalent	bivalent	bivalent
TMEM170B	bivalent	bivalent	N/A	N/A
TMEM178	bivalent	bivalent	N/A	N/A
TMEM181	bivalent	bivalent	bivalent	bivalent
TMEM196	bivalent	repressive only	bivalent	bivalent
TMEM56-RWDD3	bivalent	bivalent	N/A	N/A
TMOD2	bivalent	bivalent	N/A	N/A
TNFRSF11A	bivalent	bivalent	bivalent	active only
TNFRSF11B	bivalent	bivalent	N/A	N/A
TOX2	bivalent	bivalent	bivalent	bivalent
TP53I11	bivalent	bivalent	N/A	N/A
TPPP	bivalent	bivalent	bivalent	bivalent
TRAPPC5	bivalent	bivalent	bivalent	bivalent
TRIL	bivalent	bivalent	bivalent	bivalent
TRIM36	bivalent	bivalent	bivalent	bivalent
TRIM73	bivalent	bivalent	N/A	N/A
TRIM9	bivalent	bivalent	N/A	N/A

TRPC3	bivalent	bivalent	bivalent	active only
TRPC4	bivalent	bivalent	N/A	N/A
TRPV4	bivalent	bivalent	bivalent	bivalent
TSHR	bivalent	bivalent	N/A	N/A
TSHZ2	bivalent	repressive only	N/A	N/A
TSHZ3	bivalent	bivalent	N/A	N/A
TSPAN11	bivalent	bivalent	bivalent	bivalent
TSPAN12	bivalent	bivalent	N/A	N/A
TSPAN33	bivalent	bivalent	bivalent	bivalent
TSPAN7	bivalent	bivalent	bivalent	bivalent
TTC40	bivalent	bivalent	N/A	N/A
TTC9	bivalent	bivalent	bivalent	bivalent
TTLL6	bivalent	bivalent	bivalent	bivalent
TTLL9	bivalent	bivalent	bivalent	bivalent
TTYH2	bivalent	bivalent	N/A	N/A
TUBB2B	bivalent	bivalent	bivalent	bivalent
UCP1	bivalent	bivalent	bivalent	bivalent
UGT8	bivalent	bivalent	bivalent	active only
ULBP1	bivalent	bivalent	bivalent	bivalent
ULK4P2	bivalent	bivalent	N/A	N/A
UNC5B	bivalent	bivalent	bivalent	active only
UNC5D	bivalent	bivalent	bivalent	bivalent
VASH1	bivalent	bivalent	bivalent	bivalent
VASH2	bivalent	bivalent	bivalent	bivalent
VAX2	bivalent	bivalent	bivalent	bivalent
VCAN	bivalent	bivalent	bivalent	active only
VGF	bivalent	bivalent	N/A	N/A
VGLL2	bivalent	bivalent	bivalent	bivalent
VLDLR	bivalent	bivalent	N/A	N/A
VSTM2L	bivalent	bivalent	N/A	N/A
VWA5B2	bivalent	bivalent	bivalent	bivalent
WASH1	bivalent	bivalent	N/A	N/A
WASH2P	bivalent	bivalent	N/A	N/A
WFDC2	bivalent	bivalent	N/A	N/A
WIPF3	bivalent	bivalent	N/A	N/A
WNK3	bivalent	bivalent	bivalent	bivalent
WNT10B	bivalent	bivalent	N/A	N/A
WNT11	bivalent	bivalent	bivalent	bivalent
WNT16	bivalent	bivalent	bivalent	bivalent
WNT2	bivalent	bivalent	bivalent	bivalent
WNT2B	bivalent	bivalent	bivalent	bivalent
WNT4	bivalent	bivalent	bivalent	bivalent
WNT5B	bivalent	no mark	N/A	N/A
WNT7B	bivalent	bivalent	bivalent	bivalent
WNT9B	bivalent	bivalent	bivalent	bivalent
WRAP73	bivalent	bivalent	N/A	N/A

WRN	bivalent	bivalent	bivalent	active only
WT1	bivalent	bivalent	N/A	N/A
WT1-AS	bivalent	bivalent	N/A	N/A
WTH3DI	bivalent	bivalent	N/A	N/A
XKR5	bivalent	bivalent	bivalent	bivalent
ZAR1	bivalent	bivalent	bivalent	bivalent
ZBTB10	bivalent	bivalent	N/A	N/A
ZBTB42	bivalent	bivalent	N/A	N/A
ZBTB7C	bivalent	bivalent	N/A	N/A
ZBTB8B	bivalent	bivalent	bivalent	bivalent
ZDHHC24	bivalent	bivalent	N/A	N/A
ZFPM2	bivalent	bivalent	bivalent	bivalent
ZIC2	bivalent	bivalent	bivalent	bivalent
ZIC5	bivalent	bivalent	N/A	N/A
ZMYND15	bivalent	bivalent	N/A	N/A
ZNF232	bivalent	bivalent	bivalent	bivalent
ZNF385B	bivalent	bivalent	bivalent	bivalent
ZNF467	bivalent	bivalent	N/A	N/A
ZNF503	bivalent	bivalent	N/A	N/A
ZNF503-AS2	bivalent	bivalent	N/A	N/A
ZNF521	bivalent	bivalent	N/A	N/A
ZNF7	bivalent	bivalent	N/A	N/A
ZNF704	bivalent	bivalent	N/A	N/A
ZNF804A	bivalent	bivalent	bivalent	bivalent
ZNF853	bivalent	bivalent	N/A	N/A
ZRANB2-AS1	bivalent	bivalent	N/A	N/A
ZSWIM5	bivalent	bivalent	N/A	N/A

Chapter 8 References

1. Bernstein, B.E., et al., *A bivalent chromatin structure marks key developmental genes in embryonic stem cells.* Cell, 2006. **125**(2): p. 315-26.
2. Sharma, S., T.K. Kelly, and P.A. Jones, *Epigenetics in cancer.* Carcinogenesis, 2010. **31**(1): p. 27-36.
3. Egger, G., et al., *Epigenetics in human disease and prospects for epigenetic therapy.* Nature, 2004. **429**(6990): p. 457-63.
4. Audia, J.E. and R.M. Campbell, *Histone Modifications and Cancer.* Cold Spring Harb Perspect Biol, 2016. **8**(4).
5. Organisation, W.H., *World Cancer Report 2014*, ed. W.H. Stewart and P.C. Wild. 2014: WHO Press.
6. Gupta, G.P. and J. Massague, *Cancer metastasis: building a framework.* Cell, 2006. **127**(4): p. 679-95.
7. AustralianGovernment. [cited 2016 08/06/2016]; Available from: http://www.cancer.org.au/about-cancer/what-is-cancer/facts-and-figures.html.
8. Council, N.H.a.M.R. *Research funding statistics and data*. 2015 28-04-2016 [cited 2016 22/06/16]; Available from: https://www.nhmrc.gov.au/grants-funding/research-funding-statistics-and-data.
9. Hanahan, D. and R.A. Weinberg, *Hallmarks of cancer: the next generation.* Cell, 2011. **144**(5): p. 646-74.
10. Hanahan, D. and R.A. Weinberg, *The hallmarks of cancer.* Cell, 2000. **100**(1): p. 57-70.
11. Berger, M.F., et al., *Integrative analysis of the melanoma transcriptome.* Genome Res, 2010. **20**(4): p. 413-27.
12. Knudson, A.G., *Cancer genetics.* Am J Med Genet, 2002. **111**(1): p. 96-102.
13. Pollack, V.A., et al., *Inhibition of epidermal growth factor receptor-associated tyrosine phosphorylation in human carcinomas with CP-358,774: dynamics of receptor inhibition in situ and antitumor effects in athymic mice.* J Pharmacol Exp Ther, 1999. **291**(2): p. 739-48.
14. Hollstein, M., et al., *P53 MUTATIONS IN HUMAN CANCERS.* Science, 1991. **253**(5015): p. 49-53.
15. Ferlay, J., et al. *GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11* 2013 [cited 2016 08/04/2016]; Available from: http://globocan.iarc.fr.
16. Sørlie, T., et al., *Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications.* Proceedings of the National Academy of Sciences, 2001. **98**(19): p. 10869-10874.
17. Kennecke, H., et al., *Metastatic Behavior of Breast Cancer Subtypes.* Journal of Clinical Oncology, 2010. **28**(20): p. 3271-3277.
18. Parker, J.S., et al., *Supervised Risk Predictor of Breast Cancer Based on Intrinsic Subtypes.* Journal of Clinical Oncology, 2009. **27**(8): p. 1160-1167.
19. Steeg, P.S., *Tumor metastasis: mechanistic insights and clinical challenges.* Nat Med, 2006. **12**(8): p. 895-904.
20. Luzzi, K.J., et al., *Multistep nature of metastatic inefficiency: dormancy of solitary cells after successful extravasation and limited survival of early micrometastases.* Am J Pathol, 1998. **153**(3): p. 865-73.

21. Cajal, R.y., *Manual de anatomía patológica general : seguida de un resumen de microscopia aplicada a la histología y bacteriología patológicas* 2ed. 1896, Madrid: Moya. 514.
22. Hay, E.D., *Organisation and fine structure of epithelium and mesenchyme in the developing chick embryo, in: Epithelial-Mesenchymal Interactions*, ed. R. Fleischmajer and R.E. Billingham. 1968, Baltimore: Williams & Wilkins.
23. Burrows, M.T., *Studies on Wound Healing : I. "First Intention" Healing of Open Wounds and the Nature of the Growth Stimulus in the Wound and Cancer.* J Med Res, 1924. **44**(5): p. 615-644.1.
24. Clark, R.A., et al., *Fibronectin and fibrin provide a provisional matrix for epidermal cell migration during wound reepithelialization.* J Invest Dermatol, 1982. **79**(5): p. 264-9.
25. Dvorak, H.F., *Tumors: wounds that do not heal. Similarities between tumor stroma generation and wound healing.* N Engl J Med, 1986. **315**(26): p. 1650-9.
26. Hay, E.D., *An overview of epithelio-mesenchymal transformation.* Acta Anat (Basel), 1995. **154**(1): p. 8-20.
27. Muller, T., et al., *Regulation of epithelial cell migration and tumor formation by beta-catenin signaling.* Exp Cell Res, 2002. **280**(1): p. 119-33.
28. Gavert, N. and A. Ben-Ze'ev, *Epithelial-mesenchymal transition and the invasive potential of tumors.* Trends Mol Med, 2008. **14**(5): p. 199-209.
29. Aigner, K., et al., *The transcription factor ZEB1 (deltaEF1) promotes tumour cell dedifferentiation by repressing master regulators of epithelial polarity.* Oncogene, 2007. **26**(49): p. 6979-88.
30. Burk, U., et al., *A reciprocal repression between ZEB1 and members of the miR-200 family promotes EMT and invasion in cancer cells.* EMBO Rep, 2008. **9**(6): p. 582-9.
31. Bracken, C.P., et al., *A Double-Negative Feedback Loop between ZEB1-SIP1 and the microRNA-200 Family Regulates Epithelial-Mesenchymal Transition.* Cancer Research, 2008. **68**(19): p. 7846-7854.
32. Park, S.M., et al., *The miR-200 family determines the epithelial phenotype of cancer cells by targeting the E-cadherin repressors ZEB1 and ZEB2.* Genes Dev, 2008. **22**(7): p. 894-907.
33. Gheldof, A., et al., *Evolutionary functional analysis and molecular regulation of the ZEB transcription factors.* Cell Mol Life Sci, 2012. **69**(15): p. 2527-41.
34. Funahashi, J., et al., *Delta-crystallin enhancer binding protein delta EF1 is a zinc finger-homeodomain protein implicated in postgastrulation embryogenesis.* Development, 1993. **119**(2): p. 433-46.
35. Williams, T.M., et al., *Identification of a zinc finger protein that inhibits IL-2 gene expression.* Science, 1991. **254**(5039): p. 1791-4.
36. Gregory, P.A., et al., *The miR-200 family and miR-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1.* Nat Cell Biol, 2008. **10**(5): p. 593-601.
37. Korpal, M., et al., *The miR-200 family inhibits epithelial-mesenchymal transition and cancer cell migration by direct targeting of E-cadherin transcriptional repressors ZEB1 and ZEB2.* J Biol Chem, 2008. **283**(22): p. 14910-4.
38. Hurteau, G.J., et al., *Overexpression of the microRNA hsa-miR-200c leads to reduced expression of transcription factor 8 and increased expression of E-cadherin.* Cancer Res, 2007. **67**(17): p. 7972-6.
39. Christoffersen, N.R., et al., *miR-200b mediates post-transcriptional repression of ZFHX1B.* RNA, 2007. **13**(8): p. 1172-8.

40. Aceto, N., et al., *Tyrosine phosphatase SHP2 promotes breast cancer progression and maintains tumor-initiating cells via activation of key transcription factors and a positive feedback signaling loop.* Nat Med, 2012. **18**(4): p. 529-37.
41. Wellner, U., et al., *The EMT-activator ZEB1 promotes tumorigenicity by repressing stemness-inhibiting microRNAs.* Nat Cell Biol, 2009. **11**(12): p. 1487-95.
42. Wu, K., et al., *PI3K/Akt to GSK3beta/beta-catenin signaling cascade coordinates cell colonization for bladder cancer bone metastasis through regulating ZEB1 transcription.* Cell Signal, 2012. **24**(12): p. 2273-82.
43. Spaderna, S., et al., *The transcriptional repressor ZEB1 promotes metastasis and loss of cell polarity in cancer.* Cancer Res, 2008. **68**(2): p. 537-44.
44. Polytarchou, C., D. Iliopoulos, and K. Struhl, *An integrated transcriptional regulatory circuit that reinforces the breast cancer stem cell state.* Proc Natl Acad Sci U S A, 2012. **109**(36): p. 14470-5.
45. Chu, P.Y., et al., *Epithelial-mesenchymal transition transcription factor ZEB1/ZEB2 co-expression predicts poor prognosis and maintains tumor-initiating properties in head and neck cancer.* Oral Oncol, 2013. **49**(1): p. 34-41.
46. Merikallio, H., et al., *Expression of snail, twist, and Zeb1 in malignant mesothelioma.* APMIS, 2013. **121**(1): p. 1-10.
47. Ying, L., et al., *Upregulated MALAT-1 contributes to bladder cancer cell migration by inducing epithelial-to-mesenchymal transition.* Mol Biosyst, 2012. **8**(9): p. 2289-94.
48. Jia, B., et al., *Overexpression of ZEB1 associated with metastasis and invasion in patients with gastric carcinoma.* Mol Cell Biochem, 2012. **366**(1-2): p. 223-9.
49. Alkatout, I., et al., *Transcription factors associated with epithelial-mesenchymal transition and cancer stem cells in the tumor centre and margin of invasive breast cancer.* Exp Mol Pathol, 2013. **94**(1): p. 168-73.
50. Chamberlain, E.M. and M.M. Sanders, *Identification of the novel player deltaEF1 in estrogen transcriptional cascades.* Mol Cell Biol, 1999. **19**(5): p. 3600-6.
51. Takagi, T., et al., *DeltaEF1, a zinc finger and homeodomain transcription factor, is required for skeleton patterning in multiple lineages.* Development, 1998. **125**(1): p. 21-31.
52. Liu, Y., et al., *Zeb1 mutant mice as a model of posterior corneal dystrophy.* Invest Ophthalmol Vis Sci, 2008. **49**(5): p. 1843-9.
53. Hurt, E.M., et al., *Expression of the ZEB1 (deltaEF1) transcription factor in human: additional insights.* Mol Cell Biochem, 2008. **318**(1-2): p. 89-99.
54. Higashi, Y., et al., *Impairment of T cell development in deltaEF1 mutant mice.* J Exp Med, 1997. **185**(8): p. 1467-79.
55. Ikeda, K. and K. Kawakami, *DNA binding through distinct domains of zinc-finger-homeodomain protein AREB6 has different effects on gene transcription.* Eur J Biochem, 1995. **233**(1): p. 73-82.
56. Smith, G.E. and D.S. Darling, *Combination of a zinc finger and homeodomain required for protein-interaction.* Mol Biol Rep, 2003. **30**(4): p. 199-206.
57. Remacle, J.E., et al., *New mode of DNA binding of multi-zinc finger transcription factors: deltaEF1 family members bind with two hands to two target sites.* EMBO J, 1999. **18**(18): p. 5073-84.
58. Franklin, A.J., et al., *BZP, a novel serum-responsive zinc finger protein that inhibits gene transcription.* Mol Cell Biol, 1994. **14**(10): p. 6773-88.
59. DeCastro, A.J., et al., *MiR203 mediates subversion of stem cell properties during mammary epithelial differentiation via repression of DeltaNP63alpha and promotes mesenchymal-to-epithelial transition.* Cell Death Dis, 2013. **4**: p. e514.

60. Ahn, Y.H., et al., *ZEB1 drives prometastatic actin cytoskeletal remodeling by downregulating miR-34a expression.* J Clin Invest, 2012. **122**(9): p. 3170-83.
61. Sanchez-Tillo, E., et al., *ZEB1 Promotes invasiveness of colorectal carcinoma cells through the opposing regulation of uPA and PAI-1.* Clin Cancer Res, 2013. **19**(5): p. 1071-82.
62. Fontemaggi, G., et al., *The transcriptional repressor ZEB regulates p73 expression at the crossroad between proliferation and differentiation.* Mol Cell Biol. 2001 Dec;21(24):8461-70., 2001.
63. van Grunsven, L.A., et al., *Interaction between Smad-interacting protein-1 and the corepressor C-terminal binding protein is dispensable for transcriptional repression of E-cadherin.* J Biol Chem, 2003. **278**(28): p. 26135-45.
64. Grooteclaes, M.L. and S.M. Frisch, *Evidence for a function of CtBP in epithelial gene regulation and anoikis.* Oncogene, 2000. **19**(33): p. 3823-8.
65. Vannier, C., et al., *Zeb1 regulates E-cadherin and Epcam expression to control cell behavior in early zebrafish development.* J Biol Chem, 2013.
66. Roche, J., et al., *Global Decrease of Histone H3K27 Acetylation in ZEB1-Induced Epithelial to Mesenchymal Transition in Lung Cancer Cells.* Cancers (Basel), 2013. **5**(2): p. 334-56.
67. Aigner, K., et al., *The transcription factor ZEB1 (deltaEF1) represses Plakophilin 3 during human cancer progression.* FEBS Lett, 2007. **581**(8): p. 1617-24.
68. Guo, S., et al., *deltaEF1 down-regulates ER-alpha expression and confers tamoxifen resistance in breast cancer.* PLoS One, 2012. **7**(12): p. e52380.
69. Williams, K.C., et al., *The microRNA (miR)-199a/214 cluster mediates opposing effects of progesterone and estrogen on uterine contractility during pregnancy and labor.* Mol Endocrinol, 2012. **26**(11): p. 1857-67.
70. Lazarova, D.L., M. Bordonaro, and A.C. Sartorelli, *Transcriptional regulation of the vitamin D(3) receptor gene by ZEB.* Cell Growth Differ, 2001. **12**(6): p. 319-26.
71. Bae, G.Y., et al., *Loss of E-cadherin activates EGFR-MEK/ERK signaling, which promotes invasion via the ZEB1/MMP2 axis in non-small cell lung cancer.* Oncotarget, 2013. **4**(12): p. 2512-22.
72. Bird, A., *Perceptions of epigenetics.* Nature, 2007. **447**(7143): p. 396-8.
73. Richmond, T.J. and C.A. Davey, *The structure of DNA in the nucleosome core.* Nature, 2003. **423**(6936): p. 145-50.
74. Arents, G., et al., *The nucleosomal core histone octamer at 3.1 A resolution: a tripartite protein assembly and a left-handed superhelix.* Proc Natl Acad Sci U S A, 1991. **88**(22): p. 10148-52.
75. Bannister, A.J. and T. Kouzarides, *Regulation of chromatin by histone modifications.* Cell Research, 2011. **21**(3): p. 381-395.
76. Kiefer, C.M., et al., *Epigenetics of beta-globin gene regulation.* Mutat Res, 2008. **647**(1-2): p. 68-76.
77. Sims, R.J., 3rd, et al., *Human but not yeast CHD1 binds directly and selectively to histone H3 methylated at lysine 4 via its tandem chromodomains.* J Biol Chem, 2005. **280**(51): p. 41789-92.
78. Champagne, K.S. and T.G. Kutateladze, *Structural insight into histone recognition by the ING PHD fingers.* Curr Drug Targets, 2009. **10**(5): p. 432-41.
79. Hansen, K.H., et al., *A model for transmission of the H3K27me3 epigenetic mark.* Nat Cell Biol, 2008. **10**(11): p. 1291-300.
80. Li, C.H., et al., *Enhancer of zeste homolog 2 silences microRNA-218 in human pancreatic ductal adenocarcinoma cells by inducing formation of heterochromatin.* Gastroenterology, 2013. **144**(5): p. 1086-1097.e9.

81. Hahn, M.A., et al., *Loss of the Polycomb Mark from Bivalent Promoters Leads to Activation of Cancer-Promoting Genes in Colorectal Tumors.* Cancer Research, 2014. **74**(13): p. 3617-3629.
82. Sachs, M., et al., *Bivalent Chromatin Marks Developmental Regulatory Genes in the Mouse Embryonic Germline in Vivo.* Cell reports, 2013. **3**(6): p. 1777-1784.
83. Tam, W.L. and R.A. Weinberg, *The epigenetics of epithelial-mesenchymal plasticity in cancer.* Nat Med, 2013. **19**(11): p. 1438-49.
84. O'Sullivan, D.E., et al., *Epigenetic and genetic burden measures are associated with tumor characteristics in invasive breast carcinoma.* Epigenetics, 2016. **11**(5): p. 344-53.
85. Valk-Lingbeek, M.E., S.W. Bruggeman, and M. van Lohuizen, *Stem cells and cancer; the polycomb connection.* Cell, 2004. **118**(4): p. 409-18.
86. Zhang, Y., et al., *Model-based analysis of ChIP-Seq (MACS).* Genome Biol, 2008. **9**(9): p. R137.
87. Ashburner, M., et al., *Gene Ontology: tool for the unification of biology.* Nat Genet, 2000. **25**(1): p. 25-29.
88. Consortium, T.G.O., *Gene Ontology Consortium: going forward.* Nucleic Acids Research, 2015. **43**(D1): p. D1049-D1056.
89. Lim, Y.Y., et al., *Epigenetic modulation of the miR-200 family is associated with transition to a breast cancer stem-cell-like state.* J Cell Sci, 2013. **126**(Pt 10): p. 2256-66.
90. Conn, Simon J., et al., *The RNA Binding Protein Quaking Regulates Formation of circRNAs.* Cell. **160**(6): p. 1125-1134.
91. Elenbaas, B., et al., *Human breast cancer cells generated by oncogenic transformation of primary mammary epithelial cells.* Genes Dev, 2001. **15**(1): p. 50-65.
92. Thiery, J.P., *Epithelial-mesenchymal transitions in tumour progression.* Nat Rev Cancer, 2002. **2**(6): p. 442-54.
93. Warzecha, C.C., et al., *An ESRP-regulated splicing programme is abrogated during the epithelial-mesenchymal transition.* Embo j, 2010. **29**(19): p. 3286-300.
94. *Gene Ontology Consortium: going forward.* Nucleic Acids Res, 2015. **43**(Database issue): p. D1049-56.
95. Roh, J., et al., *Intermedin is a calcitonin/calcitonin gene-related peptide family peptide acting through the calcitonin receptor-like receptor/receptor activity-modifying protein receptor complexes.* J Biol Chem, 2004. **279**(8): p. 7264-74.
96. Chauhan, M., et al., *Expression of adrenomedullin 2 (ADM2)/intermedin (IMD) in human placenta: role in trophoblast invasion and migration.* Biol Reprod, 2009. **81**(4): p. 777-83.
97. Havemann, D., et al., *Intermedin/adrenomedullin 2 is associated with implantation and placentation via trophoblast invasion in human pregnancy.* J Clin Endocrinol Metab, 2013. **98**(2): p. 695-703.
98. Chauhan, M., et al., *Adrenomedullin 2/intermedin regulates HLA-G in human trophoblasts.* Biol Reprod, 2011. **85**(6): p. 1232-9.
99. Zhang, W., et al., *Intermedin: a novel regulator for vascular remodeling and tumor vessel normalization by regulating vascular endothelial-cadherin and extracellular signal-regulated kinase.* Arterioscler Thromb Vasc Biol, 2012. **32**(11): p. 2721-32.
100. Smith, R.S., Jr., et al., *Intermedin is a new angiogenic growth factor.* Am J Physiol Heart Circ Physiol, 2009. **297**(3): p. H1040-7.

101. Yamac, A.H., et al., *Implication of plasma intermedin levels in patients who underwent first-time diagnostic coronary angiography: a single centre, cross-sectional study.* BMC Cardiovasc Disord, 2014. **14**: p. 182.
102. Ni, X., et al., *Intermedin/adrenomedullin2: an autocrine/paracrine factor in vascular homeostasis and disease.* Sci China Life Sci, 2014. **57**(8): p. 781-9.
103. Lu, Y.M., et al., *The prognostic value of intermedin in patients with breast cancer.* Dis Markers, 2015. **2015**: p. 862158.
104. Hollander, L.L., et al., *The novel tumor angiogenic factor, adrenomedullin-2, predicts survival in pancreatic adenocarcinoma.* J Surg Res, 2015. **197**(2): p. 219-24.
105. Kim, Y.B., et al., *The Role of the Pleckstrin Homology Domain-Containing Protein CKIP-1 in Activation of p21-activated Kinase 1 (PAK1).* J Biol Chem, 2015.
106. Bos, J.L., *RAS ONCOGENES IN HUMAN CANCER - A REVIEW.* Cancer Research, 1989. **49**(17): p. 4682-4689.
107. Huber, M.A., N. Kraut, and H. Beug, *Molecular requirements for epithelial-mesenchymal transition during tumor progression.* Current Opinion in Cell Biology, 2005. **17**(5): p. 548-558.
108. Waring, P., et al., *RAS Mutations as Predictive Biomarkers in Clinical Management of Metastatic Colorectal Cancer.* Clin Colorectal Cancer, 2016. **15**(2): p. 95-103.
109. McCubrey, J.A., et al., *Roles of the Raf/MEK/ERK pathway in cell growth, malignant transformation and drug resistance.* Biochim Biophys Acta, 2007. **1773**(8): p. 1263-84.
110. Schurmans, S., et al., *The Ras/Rap GTPase activating protein RASA3: from gene structure to in vivo functions.* Adv Biol Regul, 2015. **57**: p. 153-61.
111. Reinhold, W.C., et al., *CellMiner: A Web-Based Suite of Genomic and Pharmacologic Tools to Explore Transcript and Drug Patterns in the NCI-60 Cell Line Set.* Cancer Research, 2012. **72**(14): p. 3499-3511.
112. Tian, B., et al., *Analysis of the TGFbeta-induced program in primary airway epithelial cells shows essential role of NF-kappaB/RelA signaling network in type II epithelial mesenchymal transition.* BMC Genomics, 2015. **16**: p. 529.
113. Slabakova, E., et al., *TGF-beta1-induced EMT of non-transformed prostate hyperplasia cells is characterized by early induction of SNAI2/Slug.* Prostate, 2011. **71**(12): p. 1332-43.
114. Lindley, L.E. and K.J. Briegel, *Molecular characterization of TGFbeta-induced epithelial-mesenchymal transition in normal finite lifespan human mammary epithelial cells.* Biochem Biophys Res Commun, 2010. **399**(4): p. 659-64.
115. Chaffer, C.L., et al., *Poised chromatin at the ZEB1 promoter enables breast cancer cell plasticity and enhances tumorigenicity.* Cell, 2013. **154**(1): p. 61-74.
116. Malouf, G.G., et al., *Architecture of epigenetic reprogramming following Twist1-mediated epithelial-mesenchymal transition.* Genome Biol, 2013. **14**(12): p. R144.
117. Teng, X., et al., *Inhibition of endoplasmic reticulum stress by intermedin(1-53) protects against myocardial injury through a PI3 kinase-Akt signaling pathway.* J Mol Med (Berl), 2011. **89**(12): p. 1195-205.
118. Chen, H., et al., *Intermedin suppresses pressure overload cardiac hypertrophy through activation of autophagy.* PLoS One, 2013. **8**(5): p. e64757.
119. Crick, F.H., *On protein synthesis.* Symp Soc Exp Biol, 1958. **12**: p. 138-63.
120. Han, J., et al., *Pre-mRNA splicing: where and when in the nucleus.* Trends Cell Biol, 2011. **21**(6): p. 336-43.
121. Chen, M. and J.L. Manley, *Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches.* Nat Rev Mol Cell Biol, 2009. **10**(11): p. 741-754.

122. Hartmann, B. and J. Valcarcel, *Decrypting the genome's alternative messages.* Curr Opin Cell Biol, 2009. **21**(3): p. 377-86.
123. Pan, Q., et al., *Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing.* Nat Genet, 2008. **40**(12): p. 1413-5.
124. Wang, E.T., et al., *Alternative isoform regulation in human tissue transcriptomes.* Nature, 2008. **456**(7221): p. 470-6.
125. Ring, H.Z. and J.T. Lis, *The SR protein B52/SRp55 is essential for Drosophila development.* Mol Cell Biol, 1994. **14**(11): p. 7499-506.
126. Wang, J., Y. Takagaki, and J.L. Manley, *Targeted disruption of an essential vertebrate gene: ASF/SF2 is required for cell viability.* Genes Dev, 1996. **10**(20): p. 2588-99.
127. Jumaa, H., G. Wei, and P.J. Nielsen, *Blastocyst formation is blocked in mouse embryos lacking the splicing factor SRp20.* Curr Biol, 1999. **9**(16): p. 899-902.
128. Conn, S.J., et al., *The RNA binding protein quaking regulates formation of circRNAs.* Cell, 2015. **160**(6): p. 1125-34.
129. Srebrow, A. and A.R. Kornblihtt, *The connection between splicing and cancer.* J Cell Sci, 2006. **119**(Pt 13): p. 2635-41.
130. Venables, J.P., *Aberrant and alternative splicing in cancer.* Cancer Res, 2004. **64**(21): p. 7647-54.
131. Brown, R.L., et al., *CD44 splice isoform switching in human and mouse epithelium is essential for epithelial-mesenchymal transition and breast cancer progression.* J Clin Invest, 2011. **121**(3): p. 1064-74.
132. Manavella, P.A., et al., *The ZFHX1A gene is differentially autoregulated by its isoforms.* Biochem Biophys Res Commun, 2007. **360**(3): p. 621-6.
133. Maldonado, A., F. Vara, and A. Sillero, *Improved application of the oscillating method for the isoelectric point determination of protein: potential connection with protein data banks.* Comput Biol Med, 2008. **38**(1): p. 23-30.
134. Peshkin, L., et al., *On the Relationship of Protein and mRNA Dynamics in Vertebrate Embryonic Development.* Dev Cell, 2015. **35**(3): p. 383-94.
135. Maier, T., M. Guell, and L. Serrano, *Correlation of mRNA and protein in complex biological samples.* FEBS Lett, 2009. **583**(24): p. 3966-73.
136. Rockland. *Chromatin Immunoprecipitation: 10 Technical Tips for Success*. 2016 22/08/2016]; Available from: http://www.rockland-inc.com/chip-technical-tips.aspx.
137. *The ENCODE (ENCyclopedia Of DNA Elements) Project.* Science, 2004. **306**(5696): p. 636-40.
138. Siles, L., et al., *ZEB1 imposes a temporary stage-dependent inhibition of muscle gene expression and differentiation via CtBP-mediated transcriptional repression.* Mol Cell Biol, 2013. **33**(7): p. 1368-82.
139. Gubelmann, C., et al., *Identification of the transcription factor ZEB1 as a central component of the adipogenic gene regulatory network.* Elife, 2014. **3**: p. e03346.
140. Machanick, P. and T.L. Bailey, *MEME-ChIP: motif analysis of large DNA datasets.* Bioinformatics, 2011. **27**(12): p. 1696-7.
141. Alam, M., et al., *MUC1-C Represses the Crumbs Complex Polarity Factor CRB3 and Downregulates the Hippo Pathway.* Mol Cancer Res, 2016.
142. Vlachos, I.S., et al., *DIANA miRPath v.2.0: investigating the combinatorial effect of microRNAs in pathways.* Nucleic Acids Research, 2012. **40**(W1): p. W498-W504.
143. Kim, Y.B., et al., *The Role of the Pleckstrin Homology Domain-containing Protein CKIP-1 in Activation of p21-activated Kinase 1 (PAK1).* J Biol Chem, 2015. **290**(34): p. 21076-85.

144. Safi, A., et al., *Role for the pleckstrin homology domain-containing protein CKIP-1 in phosphatidylinositol 3-kinase-regulated muscle differentiation.* Mol Cell Biol, 2004. **24**(3): p. 1245-55.
145. Wang, X.L., et al., *Association of plasma intermedin levels with progression and metastasis in men after radical prostatectomy for localized prostatic cancer.* Cancer Biomark, 2015. **15**(6): p. 799-805.
146. Hikosaka, T., et al., *Adrenomedullin production is increased in colorectal adenocarcinomas; its relation to matrix metalloproteinase-9.* Peptides, 2011. **32**(9): p. 1825-31.
147. Ma, H.-w., et al., *The pseudogene derived long noncoding RNA DUXAP8 promotes gastric cancer cell proliferation and migration via epigenetically silencing PLEKHO1 expression*. 2016. 2016.
148. Nie, J., et al., *CKIP-1 acts as a colonic tumor suppressor by repressing oncogenic Smurf1 synthesis and promoting Smurf1 autodegradation.* Oncogene, 2014. **33**(28): p. 3677-87.
149. Canton, D.A., et al., *The pleckstrin homology domain-containing protein CKIP-1 is involved in regulation of cell morphology and the actin cytoskeleton and interaction with actin capping protein.* Mol Cell Biol, 2005. **25**(9): p. 3519-34.
150. Nie, J., et al., *CKIP-1: a scaffold protein and potential therapeutic target integrating multiple signaling pathways and physiological functions.* Ageing Res Rev, 2013. **12**(1): p. 276-81.
151. Arnott, D., et al., *ABRF-PRG04: Differentiation of Protein Isoforms.* Journal of Biomolecular Techniques : JBT, 2007. **18**(2): p. 124-134.
152. Yu, Y. and R.C. Elble, *Homeostatic Signaling by Cell-Cell Junctions and Its Dysregulation during Cancer Progression.* J Clin Med, 2016. **5**(2).
153. Bracken, C.P., H.S. Scott, and G.J. Goodall, *A network-biology perspective of microRNA function and dysfunction in cancer.* Nat Rev Genet, 2016. **17**(12): p. 719-732.
154. Pena, C., et al., *E-cadherin and vitamin D receptor regulation by SNAIL and ZEB1 in colon cancer: clinicopathological correlations.* Hum Mol Genet, 2005. **14**(22): p. 3361-70.
155. Lee, E., et al., *DNMT1 Regulates Epithelial-Mesenchymal Transition and Cancer Stem Cells, Which Promotes Prostate Cancer Metastasis.* Neoplasia, 2016. **18**(9): p. 553-66.

W