

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



Contents

Abstract.....	5
Declaration.....	8
Preface.....	9
Acknowledgments.....	10
List of Figures.....	12
List of Tables.....	13
Abbreviations.....	14
Chapter 1 Introduction.....	16
1.1 Preamble.....	16
1.2 Cancer.....	17
1.2.1 Cancer is a global issue.....	17
1.2.2 Breast cancer.....	19
1.2.3 Benign and malignant tumours.....	20
1.2.4 EMT and MET in development and cancer metastasis.....	20
1.2.5 Transcription factors in cancer and EMT.....	22
1.2.6 ZEB1.....	23
1.3 Epigenetics.....	28
1.3.1 An overview.....	28
1.3.2 Bivalent genes in development and EMT.....	29
1.3.3 Epigenetic dysregulation in cancer.....	30
1.4 Thesis structure.....	31
Chapter 2 Materials and methods.....	33
2.1 General cell culture.....	33
2.1.1 Cell culture.....	33
2.1.2 Freezing cells.....	33
2.1.3 Thawing cells.....	34
2.1.4 Plasmid and siRNA transfection.....	34
2.1.5 Virus production.....	34
2.1.6 Viral transduction.....	35
2.1.7 Kill curve and stable selection.....	35
2.1.8 Incucyte proliferation assay.....	35
2.1.9 Incucyte analysis – proliferation assay.....	35
2.1.10 Incucyte scratch wound assay.....	36
2.1.11 Incucyte analysis – scratch wound assay.....	36
2.2 Molecular techniques.....	37
2.2.1 Lysate preparation.....	37
2.2.2 Bradford and Western blot.....	37
2.2.3 RNA isolation, cDNA synthesis and qRT-PCR.....	38
2.2.4 PCR.....	40
2.2.5 Restriction enzyme based cloning.....	41
2.2.6 Ligation.....	41
2.2.7 DNA Gel Electrophoresis.....	41
2.2.8 Gel Extraction.....	42
2.2.9 Gateway cloning.....	42
2.2.10 Bacterial transformation.....	43
2.2.11 Luciferase Assay.....	43

2.2.12	DNA crosslinking	43
2.2.13	ChIP-qPCR assays	44
2.2.14	Bioinformatic analysis	46
2.2.15	Statistical Analysis	48
Chapter 3	Identification and characterisation of novel bivalent genes in an <i>in vitro</i> model of EMT	49
3.1.1	Chapter outline	49
3.1.2	HMLE cell line model	49
3.1.3	Analysis of HMLE-mesHMLE ChIP-seq data	52
3.1.4	ADM2, PLEKHO1 and RASA3	58
3.2	Results	60
3.2.1	ADM2, PLEKHO1 and RASA3 are predominantly expressed in mesenchymal cell lines	60
3.2.2	ADM2, PLEKHO1 and RASA3 have validated bivalent promoters.....	69
3.2.3	Construction of ADM2, RASA3 and PLEKHO1 expression vectors and selection of stable overexpression cell lines	71
3.2.4	ADM2, PLEKHO1 and RASA3 have no consistent effect on proliferation	73
3.2.5	PLEKHO1 and ADM2 overexpression increases cell migration in MCF10A cells	77
3.2.6	PLEKHO1 and ADM2 knockdown decreases migration in MDA-MB-231 and mesHMLE cells	78
3.2.7	Role of EGF signalling in PLEKHO1 mediated migration	80
3.2.8	Co-culturing ADM2 and PLEKHO1 stable cell lines results in a synergistic increase in migration.....	83
3.2.9	Treating MCF10A_PLEKHO1 cells with MCF10A_ADM2-conditioned media increases migration rate	85
3.3	Discussion	88
3.3.1	Overview	88
3.3.2	Role of bivalent genes in the HMLE EMT model.....	88
3.3.3	Role of ADM2, PLEKHO1 and RASA3 on proliferation and migration.....	89
3.3.4	Final perspective	90
Chapter 4	Dissecting the role of ZEB1 isoforms in EMT	92
4.1	Introduction	92
4.1.1	Splicing and isoforms	92
4.1.2	ZEB1 isoforms	93
4.1.3	Chapter outline.....	94
4.2	Results	95
4.2.1	ZEB1 isoforms - mRNA and protein differences	95
4.2.2	ZEB1 isoforms are not differentially regulated during EMT	96
4.3	Conclusion: ZEB1 isoforms are not specifically regulated during	98
Chapter 5	Identifying novel ZEB1 targets in EMT	99
5.1	Introduction	99
5.2	Results	99
5.2.1	Validation of ZEB1 promoter bivalence.....	99
5.2.2	ZEB1 ChIP-seq attempts.....	101
5.2.3	Analysing ENCODE ZEB1 ChIP-seq data.....	108
5.2.3.1	Quality control and removing adapters	108
5.2.3.2	Mapping reads and removing duplicates.....	111
5.2.3.3	Calling ZEB1 peaks and ZEB1 motif analysis.....	112

5.2.4	Potential ZEB1 target genes.....	115
5.2.5	Validation of potential ZEB1 targets – qRT-PCR.....	118
5.2.6	Validation of potential ZEB1 targets - luciferase	121
5.2.7	Validation of ZEB1 targets – ChIP-qRT-PCR.....	122
5.3	Discussion	125
5.3.1	Interpretation of the ENCODE ZEB1 ChIP-seq.....	126
5.3.2	Validation of ZEB1 targets	126
5.3.3	Characterisation of ZEB1 targets.....	127
Chapter 6	Discussion and Conclusion	129
6.1	Discussion	129
6.1.1	Emerging picture of the role of bivalently regulated genes in development and EMT	129
6.1.2	Characterisation of ADM2 and PLEKHO1 with an EMT focus	131
6.1.3	Characterising expression of ZEB1 isoforms in EMT.....	133
6.1.4	Bioinformatic insights on ZEB1 binding	134
6.1.5	Validation and characterisation of novel ZEB1 targets	135
6.2	Conclusion.....	137
Chapter 7	Supplementary data.....	138
Chapter 8	References.....	171

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 TGF β 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 where 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, Jiyeon 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

Figure 1 The hallmarks of cancer including the six original, and four additional, hallmarks.	18
Figure 2 Morphological changes in HMLE versus mesHMLE cells.....	50
Figure 3 Epigenetic changes in bivalent promoters after EMT in HMLE cells treated with TGF β for 18 days.	51
Figure 4. Selected pairwise comparisons between marker and candidate genes.	65
Figure 5 Expression levels of ADM2, PLEKHO1, RASA3 and the three EMT markers in nine breast cell lines.....	68
Figure 6 Enrichment levels of bivalent epigenetic marks for PLEKHO1, ADM2 and RASA3.	71
Figure 7 Expression level of ADM2, PLEKHO1 and RASA3 in MCF10A stable cell lines..	72
Figure 8 Proliferation rate of MCF10A stable cell lines overexpressing RASA3, PLEKHO1 or ADM2.....	74
Figure 9 Expression level of ADM2, PLEKHO1 and RASA3 in mesHMLE and or MDA-MB-231 cells after transient siRNA knockdown.....	75
Figure 10 Effect on proliferation of transient siRNA transfection of RASA3, PLEKHO1 or ADM2 in MDA-MB-231 and mesHMLE cells	76
Figure 11 Effect on migration of constitutive overexpression of RASA3, PLEKHO1 and ADM2 in MCF10A cells	78
Figure 12 Effect of transient siRNA knockdown of RASA3, PLEKHO1 and ADM2 on migration in MDA-MB-231 and mesHMLE cells.....	79
Figure 13 Effect of EGF induction on migration of MCF10A stable cell lines overexpressing ADM2 or PLEKHO1	81
Figure 14 Effect of co-culturing MCF10A_EV, MCF10A_ADM2 and MCF10A_PLEKHO1 under different media conditions	84
Figure 15 Effect of ADM2 media on individually or co-cultured MCF10A_EV, MCF10A_ADM2 and MCF10A_PLEKHO1.....	86
Figure 16 Schematic of the different ZEB1 isoforms.....	95
Figure 17 Relative expression of the four ZEB1 isoforms across an EMT time course in HMLE cells and in MDA-MB-231 cells.	97
Figure 18 Enrichment levels of bivalent epigenetic marks for ZEB1 in HMLE and mesHMLE cells.	100
Figure 19 Comparison of percent input for H3 ChIP and ZEB1 ChIP.....	102
Figure 20 Western blot showing ZEB1 expression in 293t cells used to prepare virus compared to the EV control.	104
Figure 21 Characterisation of ZEB1 expression in 29 stable clones after doxycycline induction.	106
Figure 22 Changes in clone 22 expression and morphology over time.....	107
Figure 23. Summary of observations on ZEB1 binding from ChIP-seq analysis.....	114
Figure 24 Effect of ZEB1 knockdown on candidate gene expression in MDA-MB-231 and mesHMLE cells.	119
Figure 25 Effect of ZEB1 overexpression on candidate gene expression in HMLE and MCF10A cells.....	120
Figure 26 Effect of ZEB1 overexpression on candidate gene luciferase activity in MCF10A and HMLE cells.	122
Figure 27 ChIP-qRT-PCR of ZEB1 at tiled INADL and F11R promoters in MDA-MB-231 cells.	124

List of Tables

Table 1 Summary of ZEB1 target genes and their role in cancer and maintenance of the epithelial phenotype.	27
Table 2 Media, supplements and disassociation reagents used for cell culture.	33
Table 3 Composition of DNA and siRNA transfection for a 6-well plate.	34
Table 4 Antibody details for Western Blot and ChIP where applicable.	38
Table 5 Primer sequences for qRT-PCR.	39
Table 6 Primer sequences used for luciferase cloning.	40
Table 7 Reaction mixture for restriction digest.	41
Table 8 Ligation reaction mixture.	Error! Bookmark not defined.
Table 9 BP reaction components.	42
Table 10 LR reaction components.	42
Table 11 Primer sequences for ChIP-qRT-PCR.	46
Table 12 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.	55
Table 13 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.	57
Table 14. Z-scores for the epithelial marker CDH1 and the two mesenchymal markers, VIM and ZEB1.	62
Table 15 Pair-wise associations between the three control and three candidate bivalent genes based on a subset of the NCI-60.	66
Table 16 Summary of expression levels from the nine cell line panel assays expressed as the average epithelial or mesenchymal expression and the ratio.	66
Table 17 Summary of the processing and FastQC analysis for the HEPG2 runs.	108
Table 18 Summary of the processing and FastQC analysis for the GM12878 runs.	110
Table 19 Summary of mapped reads that met the quality threshold for each run for both cell lines.	111
Table 20 Summary of duplicates in mapped reads identified and removed using PICARD.	112
Table 21 Summary of ChIP-seq data for candidate ZEB1 target genes separated into high probability candidates and high interest candidates, and positive controls.	116
Table 22 Summary of gene expression and epigenetic state of novel ZEB1 target genes, F11R and INADL.	125

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
CO ₂	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