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