University of South Bohemia in České Budějovice

Faculty of Science

Regulation of mammalian cell-fate:

Insights from genomics based investigation of transcriptional regulation and chromatin structure & studies of preimplantation mouse embryogenesis.

HABILITATION THESIS

Alexander William Bruce Ph.D.

2017

CONTENTS

ACKNOWLEDGEMENTS

To my mind the execution of science represents one of the great social endeavours of our age, and is perhaps akin to the collective effort evident in a classical musical orchestra. Indeed, anyone who thinks they can produce a musical masterpiece with a pair of cymbals alone is either deluded or a genius. It won't come as a surprise to many that I do not consider myself a genius, so I will take this opportunity to thank and acknowledge the people who have metaphorically played along with me over the years; we might not have an magnificent musical opus to our name but I think we have certainly been responsible for more than a few pleasant tunes along the way!

I would like to start by acknowledging the input of some key people, who have been central to my post-doctoral and early independent research career: Firstly, David Vetrie Ph.D. for encouraging me, after finishing my Ph.D., to apply for a Sanger Institute Wellcome Trust Fellowship and then accepted me into his group; thanks for providing me with such a good start. Secondly, Prof. Magdalena Zernicka-Goetz for taking the chance of employing me to her mouse embryo group, despite the fact I had no prior experience in her field; I guess she must have seen some potential in me and I thank her for that and for providing me the opportunity to enter the wonderful world of the early embryo, a world from which I do not think I'll ever emerge! Thirdly, I must also thank Prof. Martin H. Johnson (also at the University of Cambridge) for his kind advice and career counsel. Fourthly, prof. RNDr. Libor Grubhoffer CSc., who as Dean of the Faculty of Science at the University of South Bohemia, recruited me to an independent research position and provided all the attendant support I required. I also acknowledge his successor as Dean, prof. RNDr. František Vácha, Ph.D., who has continued this policy of support and I would like in particular to thank him for sanctioning the purchasing of a new Faculty confocal microscope; an instrument that my group heavily relies upon. Finally, thanks and gratitude are due to successive heads of the Department of Molecular Biology (latterly merged with the Department of Genetics) for their active support, valued counsel and general '*good-eggery*'; cheers to prof. RNDr. Julius Lukeš, CSc., prof. RNDr. Ivo Šauman, Ph.D. and doc. Mgr. Tomáš Doležal, Ph.D.; plus prof. RNDr. Marek Jindra CSc.

I would next like to collectively acknowledge and thank all of my co-authors, without whom all the herein listed papers simply wouldn't have been published – so many thanks indeed! I would also like to praise all the past and current members of my own research group for all their hard work and would specifically like to single out; Vasanth Thamodaran Ph.D., Aleksandar Mihajlović Ph.D., Mgr. Lenka Veselovská/Gahurová Ph.D. and Giorgio Virnicchi M.Sc. As life should be fun, and science should not be an exception, I want to thank a few specific colleagues for their friendship and camaraderie (over and above any professional relationships), namely; Pawandeep Dhami Ph.D., Shane Dillon Ph.D. and Jonathan Cooper B.Sc. (from the Sanger Institute); Dr. Helen Bolton DLM MRCOG Ph.D., Samantha Morris Ph.D., Anna Ajduk Ph.D., David-Emlyn Parfitt Ph.D. and Rui Martins Ph.D. and Mr. Alex Sossick (Gurdon Institute); doc. Mgr. Hassan Hashimi Ph.D. and Ryan Rego Ph.D. (University of South Bohemia).

Extended thanks also go to the people who have taught on courses I guarantee here at the University of South Bohemia; Mgr. Aleš Horák Ph.D., Mgr. Lenka Veselovská/Gahurová Ph.D.,

RNDr. Eva Nováková Ph.D., RNDr. Filip Husník, Mgr. Jan Pačes Ph.D. & RNDr. Miluše Hroudová Ph.D.

The indispensible role of the funding agencies of the presented work is also gratefully acknowledged; The Wellcome Trust (United Kingdom), European Union (Marie Curie Actions) and GA ČR & GA JU (Czech Republic).

Lastly, and by no means least, I want to thank and acknowledge the contribution of my most important collaborator, RNDr. Alena Krejčí Ph.D.; a remarkable scientist and women who also just happens to be my remarkable wife and the remarkable mother to our two lovely boys, Kubík and Honzík. Alenka, your constant love and support, as well as your scientific input and discussions, are a priceless source of encouragement to me.

GLOSSARY OF ABBREVIATIONS

AJs	Adherens junctions
Amot	Angiomotin (protein/gene); mouse
bHLH	Basic-helix-loop-helix motif (commonly found in transcription factor proteins)
bp	Base-pair
Bmp	Bone morphogenetic protein 'x' (protein/gene); mouse
Bmpr	Bone morphogenetic protein receptor 'x' (protein/gene); mouse
Carm1	Co-activator associated arginine methyltransferase 1 (protein/gene); mouse
Cdc42	Cell division cycle 42 (protein/gene); mouse
Cdh1	E-cadherin (protein/gene); mouse
Cdx2	Caudal type 2 homeobox 2 (protein/gene); mouse
ChIP	Chromatin immuno-precipitation
ChIP:Chip	Chromatin immuno-precipitation coupled with genomic DNA microarray hybridisation
ChIP:Seq	Chromatin immuno-precipitation coupled with next generation DNA sequencing technologies
CTCF	CCCTC-binding factor/transcriptional repressor (protein/gene); human
Ctnnb1	β catenin (protein/gene); mouse
Dab2	Disabled homolog 2 (protein/gene); mouse
DCG	Dense core secretory granule
DMSO	Dimethyl sulphoxide
DNA	Deoxyribonucleic acid
DNaseI	Deoxyribonuclease I
DsRed	A red fluorescent protein (under ultra-violet light), originally isolated from *Discosoma* coral species
dsRNA	Double stranded ribonucleic acid
EGFP	Enhanced green fluorescent protein (under ultra-violet light), mutated from the original variant isolated from the jellyfish, *Aequorea victoria*
ENCODE	Encyclopaedia of DNA elements project (human genome)
ENC-array	Tile-paved ENCODE project genomic DNA microarray
EPI	Epiblast (pluripotent) blastocyst cell lineage
Erk1/2	Extracellular signal-regulated kinase 1/2 (protein/gene); mouse
ERM/pERM	Ezrin-radaxin-moesin/phospho-ezrin-radaxin-moesin (anti-sera recognised antigens); mouse
ESC	Embryonic stem cell
Ezr/(p)Ezr	Ezrin/phospho-Ezrin (protein/gene); mouse
FAIRE	Formaldehyde assisted isolation of regulatory elements
Fgf	Fibroblast growth factor 'x' (protein/gene); mouse
Fgfr	Fibroblast growth factor receptor 'x' (protein/gene); mouse
FISH	Fluorescence *in situ* hybridisation
GATA1	GATA binding protein 1 (protein/gene); human
Gata3	GATA binding protein 3 (protein/gene); mouse
Gata4	GATA binding protein 4 (protein/gene); mouse
Gata6	GATA binding protein 6 (protein/gene); mouse
GFP	Green fluorescent protein (under ultra-violet light), originally isolated from the jellyfish, *Aequorea victoria*

GTPase	Guanosine triphosphate hydrolase
HMM	Hidden Markov model
HP1	Heterochromatin protein 1 (protein/gene); human
HSC	Haematopoietic stem cell
ICM	Inner cell mass (blastocyst stage embryo)
Kb	Kilo-base-pair
Lats1/2	Large tumour suppressor kinase 1/2 (protein/gene); mouse
LDB1	LIM domain binding transcription factor (protein/gene); human
Llgl1	Lethal giant larvae homolog 1 (protein/gene); mouse
LMO2	LIM domain only transcription factor (protein/gene); human
Lrp2	Low density lipoprotein receptor 2 (protein/gene); mouse
Mapk11	p38-mitogen-activated kinase 11/β (protein/gene); mouse
Mapk12	p38-mitogen-activated kinase 12/γ (protein/gene); mouse
Mapk13	p38-mitogen-activated kinase 13/δ (protein/gene); mouse
Mapk14	p38-mitogen-activated kinase 14/α (protein/gene); mouse
Mark1/2	Microtubule affinity-regulating kinases 1/2 (protein/gene); mouse
Mb	Mega-base-pair
Mek1/2	Mitogen-activated kinase-kinase 1/2 (protein/gene); mouse
mRNA	Messenger ribonucleic acid
Nanog	Nanog homeobox transcription factor (protein/gene); mouse
NRSE	Neuron restrictive silencer element (also known as RE-1)
NRSF	Neuron restrictive silencer factor (also known as REST); human
Oct4	Octamer binding factor 4 (gene/protein; also known as Pou5f1); mouse
OGDBs	Oregon-green conjugated dextran beads
PAR-aPKC	Partitioning defective-atypical protein kinase C complex
Pard3	Partitioning defective 3 homolog (protein gene); mouse
Pard6b	Partitioning defective 6 homolog beta (protein gene); mouse
PB	Second meiotic polar body
PCR	Polymerase chain reaction
p38-Mapks	p38-mitogen-activated kinases (protein/gene); mouse
PI	Preimplantation
Pou5f1	POU class 5 homeobox 1 (gene/protein; also known as Oct4); mouse
aPKCλ	Atypical protein kinase C lambda (protein/gene); mouse
aPKCζ	Atypical protein kinase C zeta (protein/gene); mouse
PRC1	Polycomb repressive complex 1 (protein/gene); human
PrE	Primitive endoderm (sometimes referred to as the hypoblast); differentiating extraembryonic blastocyst cell lineage
Prkci	Atypical protein kinase C –lambda/iota (protein/gene); mouse
Prkcz	Atypical protein kinase C –lambda/zeta (protein/gene); mouse
RDBs	Rhodamine (red) conjugated dextran beads
RE1	Repressor elemet-1 (also known as NRSE)
REST	Repressor element-1 silencing transcription factor (also known as NRSF); human
RhoA	Ras homolog family member A (protein/gene); mouse
RNA	Ribonucleic acid
RNAi	Ribonucleic acid (RNA) interference
RNApol2	Ribonucleic acid (RNA) polymerase II

Rock1/2	Rho-associated protein kinase 1/2 (protein/gene); mouse
RSP	Regulated secretory pathway
RTPCR	Reverse transcription polymerase chain reaction
SCL	Stem cell leukaemia (gene/transcription factor); human
SCL-array	Tile-paved human genomic microarray spanning the *Stem cell leukaemia/SCL* gene locus
Scrib	Scribbled planar cell polarity protein homolog (protein/gene); mouse
shRNA	Short hairpin ribonucleic acid (used to effect RNAi)
siRNA	Small interfering ribonucleic acid (used to effect RNAi)
Smad	Small body size/mothers against decapentaplegic homologs 'x' (protein/gene); mouse
Sox2	SRY-box 2 (protein/gene); mouse
Sox17	SRY-box 17 (protein/gene); mouse
TAFII250	TATA-box binding protein associated factor, 250KDa (protein/gene); human
TAL1	T-cell acute leukaemia (gene/transcription factor); human
Taz	Tafazzin (protein/gene); mouse
TE	Trophectoderm (differentiating extraembryonic) blastocyst cell lineage
Tead4	TEA domain transcription factor 4 (protein/gene); mouse
TEL1	ETS-related transcription factor 1 (protein/gene); human
TEL2	ETS-related transcription factor 2 (protein/gene); human
TFBS	Transcription factor binding site
Tjp1	Tight junction associated protein-1 (protein/gene); mouse
TJs	Tight junctions
Yap1	Yes associated protein 1/Yorkie homolog (protein/gene); mouse

PREFACE

1.1 A summary of ones research career to date.

During the last thirteen years, encompassing ones post-doctorate career, one has charted a course that began with a natural extension to ones Ph.D. thesis, involving moving to the laboratory of ones Ph.D. supervisor's collaborator (at the Wellcome Trust Sanger Institute, Cambridge, United Kingdom; to study the regulation of developmentally important gene expression on the transcriptional level), and has culminated in one establishing ones own research group at the Faculty of Sciences (University of South Bohemia in České Budějovice, Czech Republic), investigating mechanisms of cell-fate derivation during preimplantation mouse embryo development; after receiving post-doctoral training at the Gurdon Institute (University of Cambridge, United Kingdom). Although this may seem to some as an improbable route to have deliberately chosen, a few clarifying sentences may help the reader to appreciate it was perhaps not so esoteric as it may at first appear.

Indeed, just as one was finishing ones Ph.D. related to the transcriptional regulation imparted by what was considered the pan-neuronal specific gene transcriptional repressor REST, the field of chromatin structure and in particular so-called epigenetic post-translational histone modifications, was beginning to fully emerge. One had already secured a post-doctoral fellowship to move to the Sanger Institute (where a significant part of the human genome was sequenced) to continue work on REST and another developmentally important (in the context of the haematopoietic system) human transcription factor called SCL; however in addition to this work one also become involved in mapping protein-genome-chromatin interactions (using various novel techniques) that relied on the hybridisation of fractionated, purified and fluorescently labelled genomic DNA fragments to genomic DNA microarrays. The aim of using such techniques was to provide fundamental understanding of the general relationships between chromatin structure and function (for example, the regulation of gene transcription or co-transcriptional mRNA splicing) and contributed a large part of the data published in the pilot phase of the international ENCODE consortium project (Encyclopaedia of DNA elements; aimed at identifying all the functional DNA sequence elements in the human genome). Whilst, working on these REST/SCL and broader chromatin structural projects represented an extremely exciting, intensive and productive time at a 'first-class' research institute (comprising three years and ultimately yielding six publications), one began to become slightly disillusioned that the work had started to become less and less hypothesis driven and more of a 'factory-based' screening exercise, largely dictated by availability of novel reagents (*e.g.* antibodies) to well developed and optimised techniques. Whilst, this was an effective approach, one began to seek alternative employ in a more traditional research laboratory that could nevertheless benefit from ones research experience.

It was at this time that one read a *Nature* paper from Prof. Magdalena Zernicka-Goetz group, at the Gurdon Institute, detailing how cells of the 4-cell stage mouse embryo exhibit differing levels of a particular histone chromatin modification (di-methylation of arginine-26, histone H3) that correlated with the pluripotency of their progeny cells by the peri-implantation blastocyst stage. Accordingly, I speculatively applied to her laboratory and was accepted to a senior post-doctoral research position and it has been from that point

forward that I have become completely enchanted with preimplantation mouse embryos and the execution of the earliest possible cell-fate decisions of development, *per se* (not just in relation to chromatin). Indeed, after the first week in this new lab I knew this was the field in which I wanted to spend the remainder of my research career! It has therefore transpired that I have been working on preimplantation mouse embryos for the last decade, the latter years of which have been as an independent research group leader here in České Budějovice (after moving here with my Czech wife in 2010).

1.2 Explanation of the presentation of two research themes within the thesis.

As outlined above, my post-doctoral research has ranged from working on the transcriptional regulation of and by, key human cell-fate related transcription factor genes *REST* and *SCL*, through fundamental investigations of post-translational histone modification and chromatin structure/function relationships and ending with core research into how sub-populations of cells of the preimplantation stage mouse embryo regulate the balance between the initiation of differentiation to extraembryonic tissues or the retention of pluripotency required later in development to form all the cells of the foetus (the principle area that is now pursued as ones independent research programme).

Consequently the habilitation thesis is divided into two main *themes*, each of which are summarised in greater detail in the sections immediately following:

- ***Theme I*: Regulation of specific gene transcription & chromatin structure.**
- ***Theme II*: Pluripotency & the acquisition of cell-fate in the preimplantation mouse embryo.**

Firstly, please note that whilst such themes appear superficially unrelated, they do nevertheless centre on understanding how cells acquire specific and specialised cell-fates and traits, but by employing different experimental models and paradigms, available at the time of research. Secondly, and notwithstanding this qualification, one has consciously decided to weight the attention given in thesis habilitation thesis in favour of ones more contemporary and indeed current work, as described in relation to *theme II* (conducted both here in České Budějovice and at the Gurdon Institute, University of Cambridge) versus that of *theme I* (performed exclusively whilst at the Sanger Institute, Cambridge). Consequently, the summary of work performed under *theme I* will comprise of relatively short introductions and experimental descriptions/data interpretation of some of ones more significant research outputs and will contrast with a more traditional style consisting of a comprehensive introduction and placement of ones research achievements within the general field, in relation to theme II. A principle that is also reflected in the written abstract and content of ones presented habilitation lecture.

THEME I: REGULATION OF SPECIFIC GENE TRANSCRIPTION & CHROMATIN STRUCTURE.

2.1 Introduction to working on the 'Repressor element 1 (RE1)' and 'Repressor element 1 silencing transcription factor (REST)' gene expression regulatory system.*

Following the award of ones Ph.D., in the spring of 2004, one moved from the School of Molecular Biology and Biochemistry at the University of Leeds (United Kingdom) to take up a Wellcome Trust sponsored, and competitively awarded, post-doctoral fellowship at the Sanger Institute, Cambridge, (United Kingdom); working in the laboratory of an existing collaborator, David Vetrie Ph.D.

During ones Ph.D. research, one had contributed to the bioinformatic identification, categorisation (as a relational on-line data base) and functional verification, of both canonical and variant DNA sequence binding motifs (also known as transcription factor binding sites/TFBSs), known as RE1s, for a transcriptional repressor protein called REST[1], using the newly available human and mouse genome sequence being published at that time as references[2,3]. The original intention of studying such REST/RE1 interactions was drawn from the fact they were first characterised as being responsible for ensuring neuronal specific expression of initially a limited number of independently confirmed target genes, by transcriptionally repressing their expression outside the nervous system (where REST protein is abundantly expressed and is known to recruit at least two distinct and post-translational histone/chromatin modifying transcriptional co-repressor complexes; represented by mSIN3A and COREST[4-7]). Moreover, as ones Ph.D. hosting research laboratory was indeed focussed on developmental molecular neurobiology, REST/RE1 interactions became of preeminent interest. However, it also quickly become clear that the REST/RE1 system has some uniquely desirable features for the direction of meaningful genomics-based research, *per se*, in what was then the emergence of a 'post-genome' sequence era; indeed, it is noteworthy that the originally presented and overly simplistic role of the REST/RE1 system as a simple mechanism to prevent neuronal specific gene expression outside the nervous system (as introduced above) was ultimately revised and developed, partly due to some latter genomics-based works described below.

Principle amongst such desirable features is the length of the hitherto identified canonical RE1 DNA sequence element, that at 21 base-pairs (bp) is uniquely long[8,9] when compared to virtually all other specific transcription factor binding sites (TFBSs); that are commonly much smaller (for example E-box motifs responsible for binding members of the bHLH family of transcription factors are typically 6 bp in length). Consequently, the probability with which such RE1 sequences arise in genomic DNA sequence by chance alone is statistically and significantly smaller than that of other much smaller TFBSs; meaning the presence of RE1 (or RE1-like) sequences are more likely to reflect functionality, particularly if clustered with other known transcriptional regulatory elements and other genomic features, such as promoters or CpG motifs. It was therefore, to follow up this work in identifying RE1 sequences and validating their functionality in relation to REST interaction and the transcriptional regulation of putative target genes, one was awarded the fellowship to work with David Vetrie Ph.D. at the Sanger Institute, utilising what were contemporary and

** Also known as the 'Neuron restrictive silencer element (NRSE)' and 'Neuron restrictive silencer factor (NRSF)' gene expression regulatory system.*

cutting edge genomics tools and techniques, as described (although ones involvement in other non-REST related projects is also described/summarised below).

2.2 Identification of related and REST regulated genes required for establishing the neuroendocrine secretory pathway.

First, amongst this works[10] was a refinement of the bioinformatic analyses to identify RE1 sequences in genomic DNA sequence data (human and mouse) *in silico,* using refined search algorithms based on expanded experimental and verified REST/RE1 interaction data.

The second work output[11] was a demonstration that multiple genes involved in the stimulated/regulatory secretory pathway (RSP; a hallmark of neuroendocrine-secretory cells) have RE1 sequences in their promoter proximal regions, at frequencies far above that which could be explained by random probability; suggesting the whole RSP could be under coordinated and REST-mediated transcription control. Indeed, negative expression of the *Rest* gene in the rat neuroendocrine model cell line PC12, was supportive of this model and we therefore generated PC12 cell lines with ectopic *REST* expression and were able to show using microarray-based transcriptome profiling that the vast majority of the identified RSP-related candidate genes exhibited substantially reduced or ablated mRNA expression. Moreover, using a technique called chromatin immuno-precipitation (ChIP) to isolated genomic DNA fragments in association with specific proteins (utilising antibodies that recognise specific proteins, in this case REST) we showed that the transcriptionally repressed RSP-related gene candidate RE1 sequence were occupied by bound REST and that over-expression of a recombinant dominant negative form of REST was sufficient to drive transcriptional re-expression/activation of these genes. Importantly, we also demonstrated a functional and RSP-related consequence for neuroendocrine PC12 cells expressing REST; specifically that they exhibited significantly reduced numbers of cytoplasmic dense core granules (DCGs; the organelles responsible for storing secretory cargos that will then be extra-cellularly released in a controlled manner to specific stimuli) and were less able to store/retain and release up-taken radio-labelled noradrenaline (a classical DCG cargo); Fig. 1.

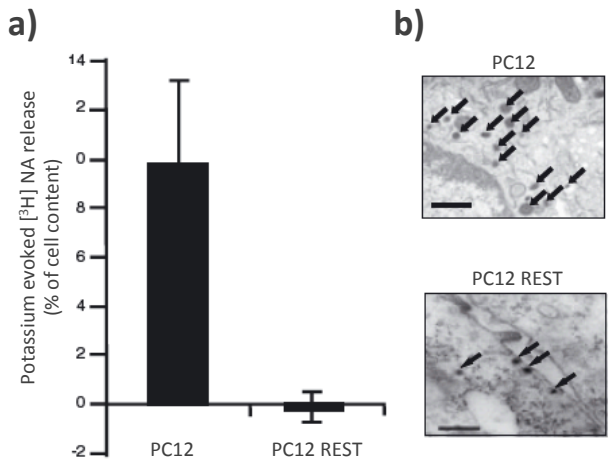

Figure 1: REST acts as a master transcriptional regulator of the neuroendocrine secretory pathway: a) The rat neuroendocrine cell line model PC12 and a stable derived line ectopically expressing REST (PC12 REST) were incubated with tritiated nor-adrenaline ([^{3}H] NA) to promote uptake and the regulated release in response to a depolarising concentration of potassium recorded. PC12 REST cells are severely impaired in regulated release **b)** Electron micrographs of PC12 and PC12 REST cells, showing dense core secretory granules (black arrows). PC12 REST cells have quantifiably fewer per unit of cytoplasm – see original publication. Scale bar = 1μm

Hence, this work demonstrate that the REST/RE1 system is fundamental in permitting neuroendocrine cells to adopt their correct developmental potential and acquire their

fundamental capacity to secrete various vesicle bound cargos to the extra-cellular milieu in a regulated and coordinated manner as a response to specific extra-cellular stimuli, by itself transcriptionally regulating the coordinated expression of the genetic components of that pathway. Interestingly, since this work was published numerous other studies[12-14] have also confirmed, what was at the time, this novel finding.

2.3 Utilising genomics tools identify functional diversity between REST/RE1 interactions across the human genome, defined at the DNA sequence level by *in vivo* binding affinity hierarchies.

As described above, the fundamental nature of the REST/RE1 system is well suited to genomics based research. At the time of ones joining David Vetrie's research group, the involvement of the Sanger Institute in an international consortium of groups aimed at identifying all the functional DNA sequence elements in the human genome (referred to as the (Encyclopaedia of DNA elements/ENCODE project;[15,16]) was confirmed. Accordingly, the project was to be first run in a pilot phase, interrogating 1% of the human genome sequence (spread across 44 individual yet contiguous regions throughout the entire genome), using various bioinformatic and more traditional experimental, yet genomically scaled, techniques, such as ChIP, transcribed RNA sequence tag mapping, DNaseI hypersensitive mapping, chromatin accessibility and interaction assays, *etc*.; before being rolled out to tackle the human genome in its entirety. As part of this pilot project phase, the Sanger Institute had been charged with constructing a highly sensitive genomic DNA microarray resource, that consisted of overlapping single stranded arrayed probe DNA molecules (each 1 -1.5 Kb in length and derived using PCR;[17]), that together provided a relative tile-paved coverage of the 44 identified contiguous pilot/test human genomic regions. As reflected above, such tile-path ENCODE microarrays (ENC-arrays) were intended to be used in hybridisation experiments, using fluorescently-labelled nucleic acid species, segregated and purified on the basis of their genome biology using some of the above mentioned techniques, thus identifying human genomic regions with specific and/or inferred functions. Indeed, one was involved in a number of these experiments, the results of which were published in the ENCODE projects preliminary findings ([15]; and will be expanded upon in a brief summary of ones own involvement in subsequent sections). However, with the development of such a resource, it was immediately apparent that REST/RE1 occupancy across 1% of the human genome could be readily assayed using the same ChIP optimised protocols already in use in the lab ([11]; see above section 2.2), only modified to assay the purified DNA fragments, after their fluorescent labelling, by hybridising to the ENC-arrays (using a so called ChIP:Chip approach), rather than by more traditional PCR and quantitative PCR based methods. Indeed, it was also recognised that such experiments would represent a far larger potential experimental sample size, than had ever been achieved before for any animal transcription factor genome interaction study (*n.b.* at this time next generation DNA sequencing techniques had not been developed, so the capability to perform analogous yet higher throughput ChIP:Seq procedures, that involved the direct sequencing of all ChIP isolated DNA fragments, was not yet possible; however, incidentally when such techniques were first reported, the prove of concept was confirmed using ChIP against the human REST protein[18]).

Therefore, an experimental design was conceived in which REST/RE1 (or to be more exact REST/genome) occupancy was assayed using ENC-arrays in eight distinct and diverse human cell lines; four each of non-neural and neural origin and of which seven were positive for REST protein expression (*n.b.* the expression of *REST* in neural tissues was still at this time a highly debated issue, and in this instance the studies findings provided important novel insight; however one of the neural cell lines expectedly did not expressing *REST* and proved the valuable negative control originally anticipated and sought) using the previously optimised anti-REST ChIP protocol[19]. After normalisation of the data (using parallel ChIP:Chip datasets, generated using non-specific antibodies) we used Hidden Markov Models/HMMs to identify a total of 250 independent REST/genome interactions across the seven *REST* expressing cells lines (no hits were confirmed in the REST-negative cell line); Fig.2 provides examples of normalised HMM identified REST/genome interactions, as viewed in the UCSC genome browser – note the relative fold enrichment of ENC-array probes associated with normalised HMM REST interaction hits versus the majority of other probes represented in the tile path of the region.

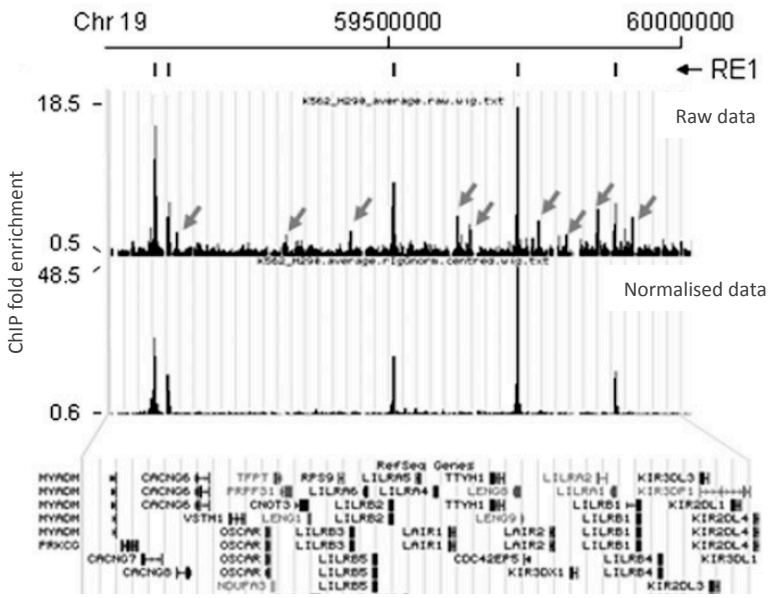

Figure 2: Exemplar normalised identified REST/genome interactions: Upper panel - UCSC genome browser view of raw anti-REST ChIP:Chip enrichments (y-axis) detected across 1Mb ENCODE region (ENm002) in K562 cell line. Lower panel – as above, after data was normalised using parallel ChIP:Chip enrichment data values, derived using non-specific rabbit antibody. Note, the identification of five *bona fide* REST interactions (labelled as RE1) in this chromosomal region by Hidden Markov Models (HMM) and the filtering out of potential false negative interactions in the raw data (grey arrows – upper panel).

In contrast to what was the prevailing view of the function of REST at the time (*i.e.* to merely prevent inappropriate neuronal specific gene expression in non-neuronal tissues) it was discovered that the vast majority of the 250 identified REST-genome interactions (89.2%) showed a degree of *'restricted'* cell line specificity, with two-thirds of these sites actually reporting '*unique'* occupancy in a single cell line. Only 27 of 250 identified individual interactions were *'common'* to all REST expressing cell lines (Fig. 3); the type of occupancy paradigm that the previous simple model of REST/RE1 interaction/function would predict. Such relationships also stood when only the non-neural datasets were considered. Hence these data, for the first time, indicated that *in vivo* REST/genome interactions are predominantly cell-/tissue-specific, implying hitherto unrecognised complexity in the REST regulome (see below). Moreover, when the average level of ChIP:Chip probe enrichment for each HMM identified interaction was plotted as a function of it either belonging to the unique, restricted or common groups (described above), it was found that the common group reported the highest enrichment, followed by the restricted group, with the lowest enrichment observed within the group of unique interactions; implying decreasing binding

affinity for REST protein for these sequences (Fig. 3). However, and rather counter-intuitively when the overall inter-species DNA sequence conservation of the ENC-array probes containing HMM identified hits was investigated, it was found that that the more

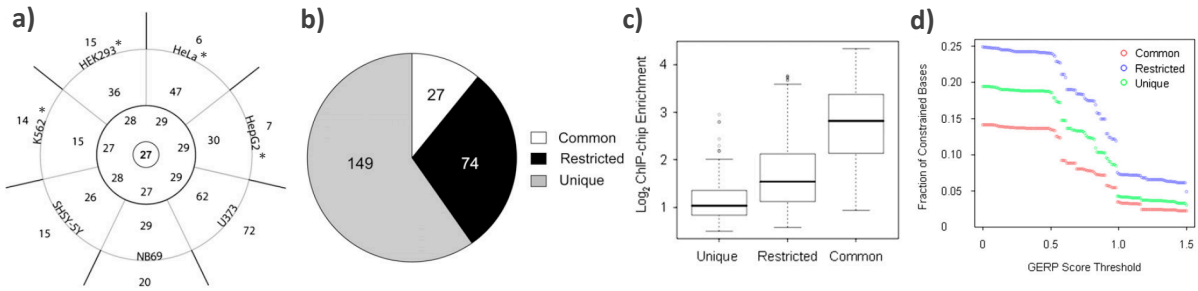

Figure 3: The majority of REST/genome interactions are unique or restricted across multiple cell lines and reflect DNA sequences with lower REST binding affinities that are nonetheless more evolutionary conserved versus sites identified in common across all cell lines: a) Number identified REST/genome interactions in three neural and four non-neural cells (*); 'uniquely' – outside circle, 'restricted' to a max. of 5 other cells lines – middle section, 'common' to all cell lines – hub. **b)** Overall fraction of all identified REST/genome interactions, either in common, restricted or unique to cell lines studied. **c)** Box plots describing the level of REST/genome ChIP:Chip enrichments, classified by unique, restricted or common occupancy in multiple cell lines studied. **d)** Evolutionary DNA sequence conservation constraints on unique, restricted and common classes of identified REST/genome interactions in respect to ENC-array probe sequence in which they reside.

weakly interacting restricted and unique REST/genome interactions resided in regions of the human genome more evolutionarily conserved that the strongest interacting regions found in common to all cell lines (Fig. 3). Moreover, a search of enriched gene ontological terms associated with genes in proximity to the identified REST/genome interactions revealed the more conserved yet weaker affinity sites were associated with genes of specific tissue function, whereas the strongest affinity sites in common to all the cell lines examined where proximal to genes reflecting intrinsic cellular and neuronal function.

Therefore the analysis revealed the evolution of an apparent affinity based hierarchy of REST/genome interactions in the human genome that was reflected in either specific or pan cell type functions of REST and its target genes. To explore this further as series of similar REST ChIP:Chip experiments were conducted in one non-neural cell line (K562) after siRNA mediated knockdown of REST protein expression (72 and 96 hour post transfection time-points; Fig. 4). It was discovered that despite effective removal of overall REST protein (detected by western blotting), there were varying levels of REST protein clearance from previously identified genomic interaction regions (that also positively correlated with some enhanced transcription of proximal target genes), with some sites not exhibiting any clearance (versus control siRNA conditions), some partial clearance, whilst still others failed to retain any REST. When the relative ChIP:Chip enrichments of the verified REST/genome interactions in the various conditions were subject to statistical hierarchal clustering analysis, three clusters of interaction were identified akin to the 'unique', 'restricted' and 'common' classes described above. Moreover, when the individual RE1 sequences present in each REST/genome interaction class were used to determine a class/cluster specific consensus RE1 sequence, distinct sequence variation patterns emerged, confirming a DNA sequenced based mechanism of establishing functionally relevant *in vivo* REST/genome interaction binding affinity hierarchies. Hence, the direct consequence of this study was to demonstrate the hitherto unrecognised complexity in the human REST regulome and that the majority of REST/RE1 interactions actually govern highly specialised tissue specific

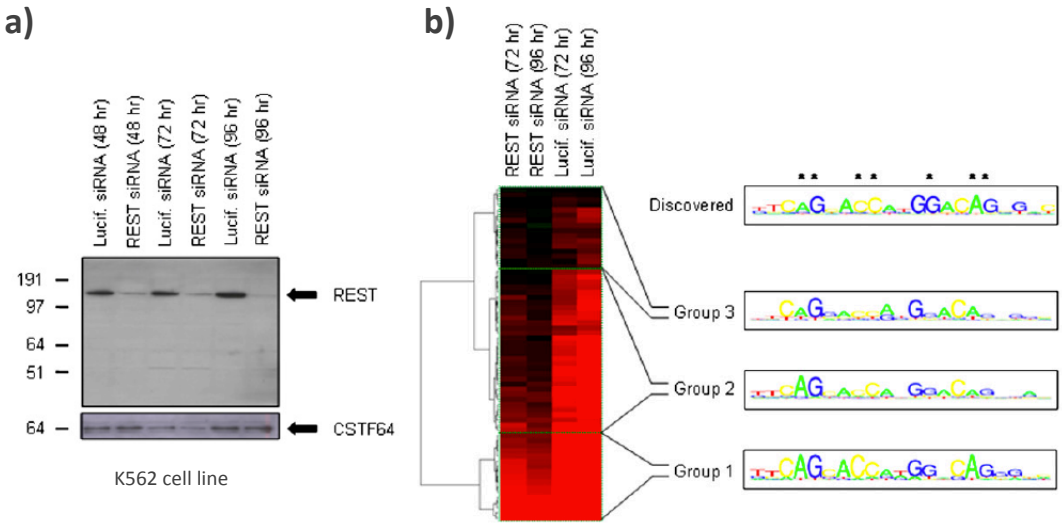

Figure 4: The varying identified REST/genome binding site affinities associated with tissue specificity and target gene transcription is dictated by DNA sequence: a) Western blotting of K562 cells transfected with either REST specific or control luciferase siRNA, detecting REST protein expression and knockdown (plus loading control detection of splice factor CSTF64 – lower panel) 48, 72 and 96 hour post-transfection. **b)** Treeview dendogram of hierarchical clustering of HMM identified REST/genome interacting ENC-array probe tile enrichments following REST specific and luciferase control siRNA knockdowns. Dendogram tile red-green shading reflects REST enrichment at individual REST/genome interacting regions (red indicating positive enrichment). To the right, consensus RE1 sequences derived from the enriched regions in each of the three identified affinity clusters/groups; a compound RE1 consensus, 'Discovered', is provide above.

genes/functions and not just neuronal gene suppression outside the nervous system, as previously proposed; Indeed REST has confirmed tissue specific roles ranging from cardiac development and specific colon and small cell lung carcinomas outside the nervous system[20-22] and Huntington disease, Downs syndrome, epileptic seizure and ischaemic shock in neural tissues[7,23-27].

2.4 Utilising genomics tools to comprehensively interrogate the human *SCL* gene loci for novel regulatory elements.

As referred to above, ones recruitment to the Sanger Institute also involved working on another project relating to the human haematopoietic cell lineage transcription factor gene *SCL* (*Stem cell leukaemia,* also known as *TAL1*/*T-cell acute leukaemia*). It is known that *SCL* is essential for the development of the entire haematopoietic system (Fig. 5) by promoting the differentiation of multipotent haemangioblasts into haematopoietic stem cells (HSCs) and hence all the subsequent haematopoietic cell lines. Its expression is maintained in the myeloid haematopoietic branch and in erythroid and megakaryocytic progenitor cells to promote the normal development of these lineages. However, its inappropriate expression during T-cell differentiation in the lymphoid haematopoietic branch leads to T-cell acute lymphoblastic leukaemia[28,29]. Accordingly, expression of the *SCL* gene is thought to be under tight transcriptional control and it was to a project interrogating the entire human *SCL* gene locus for the characterisation of existing and identification of potentially transcriptional regulatory elements that one was engaged.

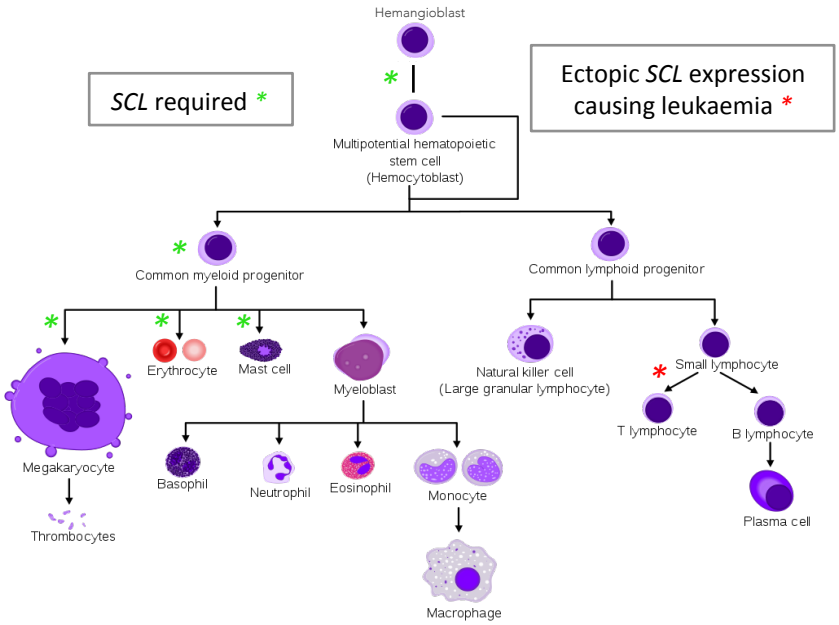

Figure 5: A summary of mammalian haematopoiesis: *SCL* expression is required for haematopoietic stem cell formation and in the differentiation of myeloid linages but its aberrant expression in lymphoid lineages can be pathogenic.

As described above for the REST related research, this study[30] also relied on the construction of a PCR product derived tile paved genomic DNA microarray. However, rather than comprising 1% of the human genome (ENC-array), this microarray covered a contiguous 256 Kb genomic locus surrounding the human *SCL* gene (and including neighbouring flanking genes; 5' - *CYP4AZ1*, *CYP2A22* and *MAP17* and 3' *SIL* and *KCY*) at an average resolution of 459bp per probe/tile, and is referred to as the *SCL*-array. Such *SCL*-arrays were intended for use in ChIP:Chip experiments, technically analogous to those described above for REST, whereby they were hybridised with fluorescently labelled genomic DNA fragments obtained using a broad spectrum of antibodies raised to relevant transcriptionally related proteins; it was anticipated this would help the high resolution identification of discreet regions, known and novel, within the *SCL* locus that contribute to its tightly controlled transcriptional regulation. Accordingly, ChIP:Chip experiments were conducted in the well-studied erythroid cell line model K562, using antibodies that recognised core components of the transcriptional machinery (RNApol2 and TAFII250; a component of the general transcription factor TFIID responsible for assembling RNApol2 initiation complex at gene promoters), insulator elements bind proteins, that segmentally define chromosomal transcriptional regulatory regions (CTCF) and core histones (H2B and H3) to give a readout of nucleosome density (as nucleosomes are typically depleted around active DNA sequence elements); a second, ChIP independent, technique called FAIRE (formaldehyde assisted isolation of regulatory elements) was also used to assay nucleosome density*. The results of these experiments are summarised in figure 6; whereby the ChIP:Chip/FAIRE enrichments (Log_2 scale converted) for each tile/probe on the SCL-array is plotted against its genomic coordinate, with gene annotation.

As a direct consequence of these experiments, all previously characterised *SCL*-gene specific transcriptional regulatory elements were identified. Additionally, a further six novel elements, both upstream and downstream of the *SCL* gene were identified. These were

** Note that parallel ChIP:Chip experiments using antibodies raised against specific epigenetic post-translation histone modifications, thought to correlate with transcriptionally permissive and repressive chromatin states were also performed, but similar experiments in terms of another more comprehensive project are summarised below.*

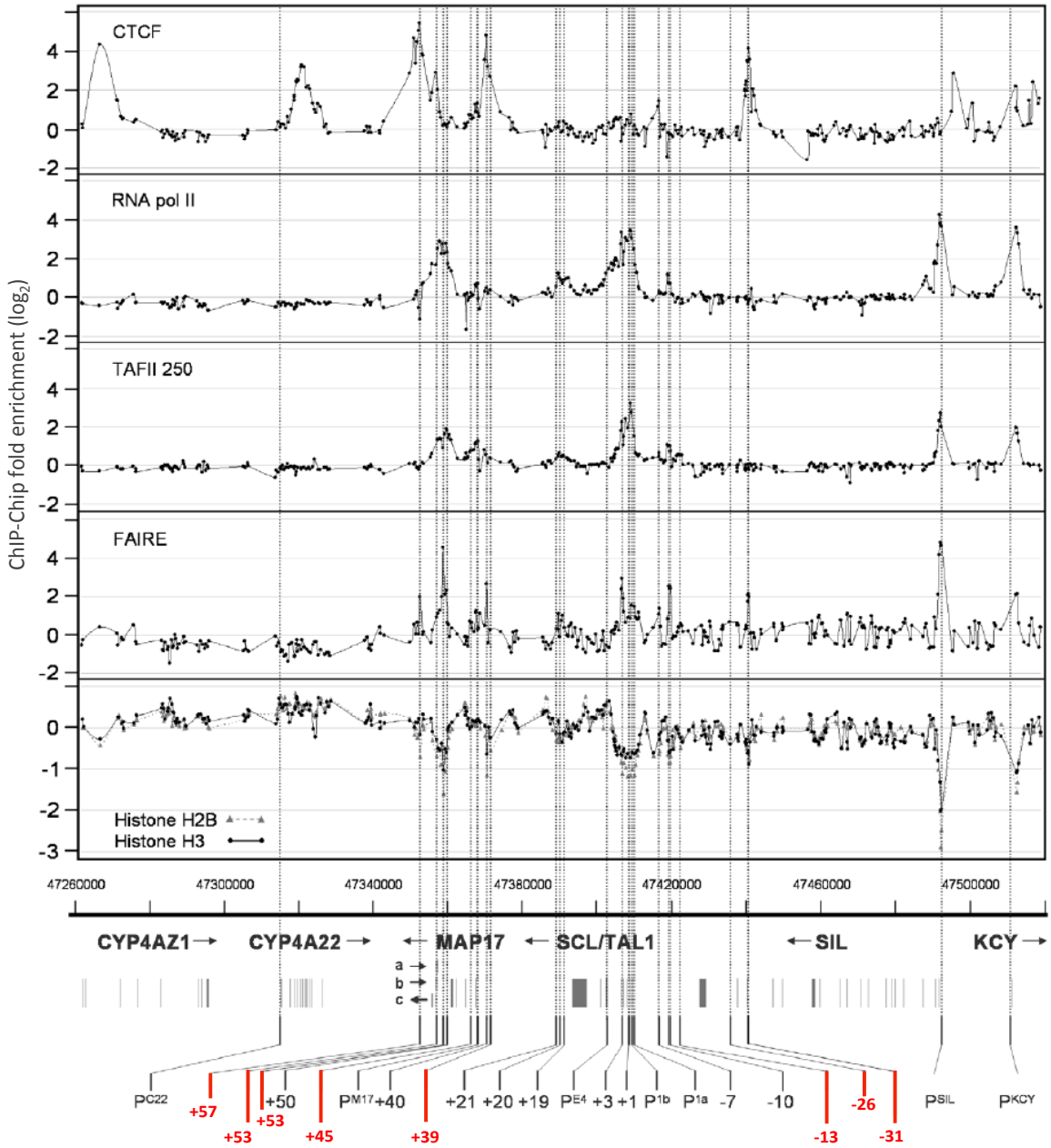

Figure 6: Interrogating the human *SCL* gene locus for novel transcriptional regulatory elements using ChIP:Chip (and FAIRE): Shown on the y-axes are ChIP:Chip enrichments (log_2 transformed) for various stated protein factors and histones (plus FAIRE accessibility enrichments) across the human *SCL* locus (gene annotation on chromosome 1 shown on x-axis; vertical bars denoting gene exons; a b and c represent identified transcripts of unknown function). Vertical lines passing through data panels to the gene annotations below represent the genomic location of previously known (grey) and novel (red) DNA regulatory elements identified in this study (gene promoters are denoted by P); six of the novel elements proving to be functional in regulating the *SCL* gene promoter 1a and interacting with known haematopoietic transcription factors. Note the DNA regulatory elements are named with the nomenclature of their location in the locus relative to *SCL* gene promoter 1a, in Kb.

shown to have either intrinsic enhancer and repressor function when cloned in luciferase reporter constructs under control of *SCL* promoter 1a in erthyroid but not lymphoid cell line models (K562 and HPB-ALL, respectively). Lastly, ChIP analysis confirmed the recruitment of a number of key haematopoietic transcription factors (including GATA1, LMO2, LDB1, TEL1/2 and SCL itself) to the novel enhancers that were also enriched in multi-species conserved DNA sequences known to interact with these factors. Overall, these data

substantially expanded the general knowledge regarding the transcriptional regulation of the key *SCL* transcription factor gene; a critically important gene during both the initiation of blood cell developmental and latterly into selected myeloid lineages. Such enhanced understanding was also augmented by the results of another project, that for reasons of space constraint one cannot substantially expand upon, that utilised a modified version of the DNaseI hypersensitive site mapping protocol (a gold standard method of functional DNA element identification, albeit low throughput) to hybridise SCL-arrays (and their counterparts derived from mouse genomic DNA) and subsets of the ENC-array to also successfully identify novel DNA elements relevant to haematopoietic development[31].

2.5 Assaying specific post-translational histone modifications across the genome; uncovering chromatin structure function relationships.

As has been stated above, ones time at the Sanger Institute also coincided with the institute's involvement in the initial pilot phases of the human ENCODE project consortiums efforts to map all the regulatory elements in the human genome. Principle amongst the Sanger's contribution was the manufacture (drawing on expertise gleaned in the production of the *SCL*-array) and exploitation of the ENC-array, comprising 44 individual tile-paved regions constituting 1% of the human genome. As has been summarised, one did use this resource to reveal fundamental insights into the REST regulome[19], however it was originally intended to use it in ChIP:Chip based assays to probe the distribution of a comprehensive selection of post-translational histone modifications (that are extraordinarily well conserved across eukaryotic species) across the genome and to functionally relate these data to underlying genetic features (*e.g.* gene promoters, enhancers, intron-exon boundaries, expression domains *etc.*). It is important to note, that at this time the role of such modifications was poorly understood. Although it was recognised they could be inherited from one cell generation to the next (so called 'epigenetic' inheritance; that is independent that of the DNA sequence itself) their exact roles, other than some reported associations of specific modifications in either transcriptionally permissive or restrictive chromatin, was far from clear and only beginning to emerge; although the idea that specific combinations of modifications may reflect some instructive potential was postulated as the 'Histone Code Hypothesis'[32]. Figure 7 summarises a still contemporary view of post-translational histone modification distribution across actively transcribed and repressed genes[33].

Therefore, the ENCODE experiments, as proposed, were very much at the vanguard of epigenetic chromatin research. Therefore, owing to ones involvement with the similar *SCL*-array and REST based projects, one inevitably became integrated into the collective effort that was the ENCODE project; most typically testing various reagents for an ever-expanding selection of specific post-translational histone modifications (usually using the SCL-array as a test bed, hence also contributing to the identification of the novel *SCL* gene regulating DNA elements, as described above; for example demonstrating that active gene enhancers are selectively enriched in mono-methylated-lysine-4 histone H3/K4H3me1 species compared with active promoters that are enriched in K4H3me2 and K4H3me3 modified histones) and optimising techniques to directly hybridised fluorescently labelled cDNA to the genomic ENC-arrays; thus identifying actively transcribed regions of the genome. Consequently, these and similar experiments were published in the preliminary findings of the ENCODE pilot project; both as an overall summarising paper in the journal *Nature*[15] and also as series

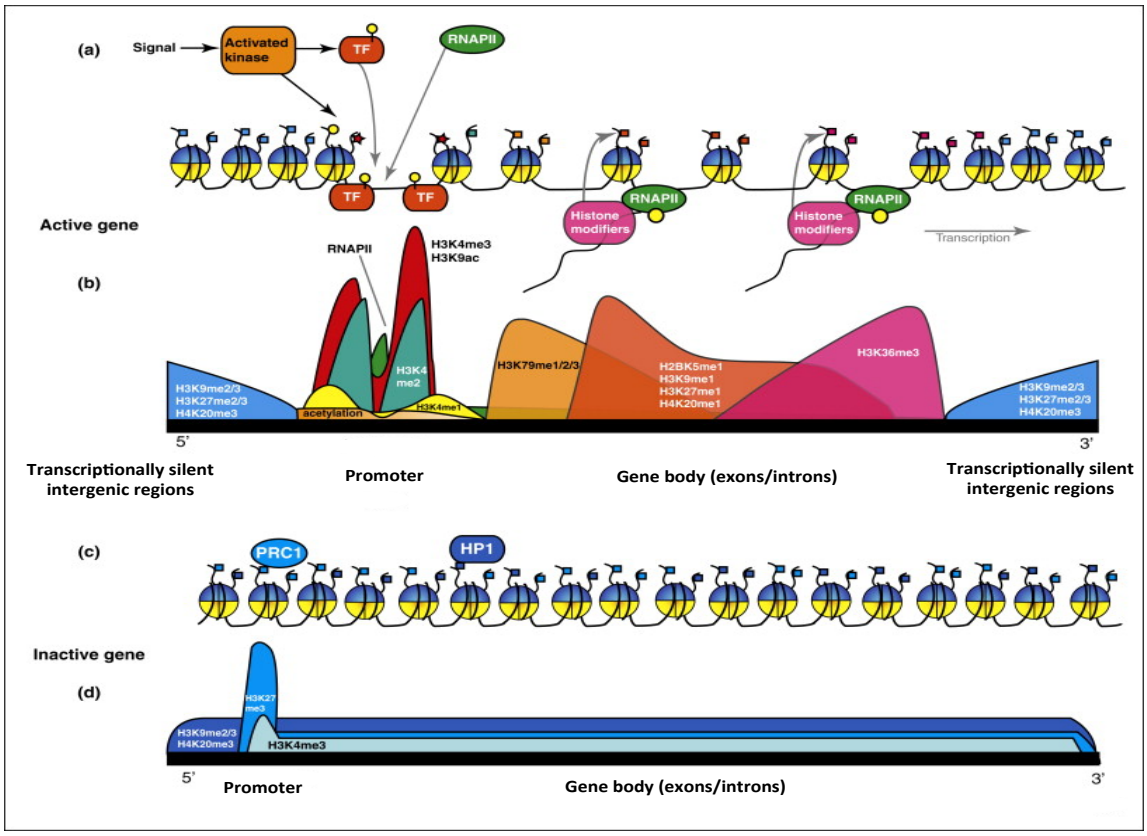

Figure 7: A summary of individual post-translational histone modification distribution across gene loci according to transcriptional status: a&b) Specific transcription factors modify histones at promoter regions with low nucleosome density, permitting RNApol2 binding. Initiated transcription is first maintained and then prolonged by the recruitment of other histone modifying activities, by RNApol2, targeting nucleosomes in the gene body. Non-transcribed intergenic regions are marked by transcriptionally non-permissive modifications. **c&d)** Transcriptionally repressed gene loci are marked throughout by similarly transcriptionally non-permissive modifications that recruit heterochromatin forming proteins (*e.g.* HP1). Other enzymatic complexes (*e.g.* PRC1) mark the promoter region making it inaccessible to RNApol2. Post-translational histone modification nomenclature: *e.g.* H3K4me3 – tri-methylation of lysine-4 of histone H3, *etc.*

of manuscripts from the consortium's participating laboratories in a special edition of *Genome Research*, that included the Sanger Institute output referred to above[17]. The space constraints placed upon this habilitation thesis preclude similar in depth discussion of the results contained within these publications (moreover, they are less suited to a habilitation thesis than the other described publications owing to the highly collaborative nature of work), however it is prescient to highlight the main findings of the pilot ENCODE project; i) the human genome is pervasively transcribed (the majority of DNA bases are associated with at least one primary transcript) and many transcripts link distal genomic regions (with protein coding potential); ii) a proliferation in the number of novel non-protein-coding transcripts was identified; iii) chromatin accessibility and histone modification patterns are indicative of the presence and activity of transcriptional start sites; iv) specific combinations of chromatin/histone modification patterns can distinguish gene promoters and distal transcription factor specific enhancers/repressor or insulator elements (initially identified by DNaseI hypersensitivity assay mapping); iv) transcriptional regulatory sequence elements evenly cluster both upstream and downstream of transcription start sites; v) DNA replication timing is correlated with chromatin structure; vi) of the 5% of human genome bases known to be under evolutionary constraint, 60% identified with function based on

ENCODE data; vii) however, a surprisingly large number of identified functional elements are not constrained throughout mammalian evolution. It is important to note that the presented findings in relation to chromatin structure of histone modifications, largely conducted at the Sanger Institute, was a primer of our current knowledge (summarised Fig. 7). Moreover, such findings were also supported by our own concurrent/subsequent observations assaying the human *SCL* gene locus in previously unprecedented detail[34].

2.6 A summary of theme I research.

The work described above centres on the time in ones research career immediately following ones Ph.D. award as a Sanger Institute Welcome Trust fellow. It describes how one extended ones Ph.D. studies on the developmentally important and relevant transcription factor REST. First, identifying it as the master transcriptional regulator of the RSP and hence neuroendocrine cell-fate[10,11]. Secondly, and by utilising novel and sophisticated genomics techniques, uncovering hitherto unrecognised complexity in the types of genes regulated by REST and moreover identifying *in vivo* REST DNA binding affinity hierarchies that are manifest in the evolutionary sequence conservation of its cognate DNA binding element (the RE1); culminating in a greater understanding of what had been previously the under-recognised tissue specific functions/roles of REST, in addition to its traditionally recognised role as a neuronal gene repressor outside the nervous system[19]. Also described is ones involvement in using similar techniques to comprehensively survey the genomic locus of the key human haematopoietic transcription factor *SCL* gene. Specifically, using antibodies raised against specific core components of the RNA polymerase machinery, relevant individual transcription factors, insulator binding proteins and distinct post-translational modified histones in high-resolution ChIP:Chip assays, coupled to various chromatin accessibility assays to identify, confirm and validate the existence of 6 novel regulatory elements[30,31,34]. Lastly, how this experience was also applied to the Sanger Institute's involvement of the international ENCODE consortiums efforts to identify all functionally relevant DNA sequence elements in the human genome; pilot project phase[15,17]. Hence, providing a genome centric insight into ones early research career into the regulation of mammalian cell-fate.

2.7 References relevant to theme I.

1 Bruce, A. W. *et al.* Genome-wide analysis of repressor element 1 silencing transcription factor/neuron-restrictive silencing factor (REST/NRSF) target genes. *Proc Natl Acad Sci U S A* **101**, 10458-10463,(2004).

2 Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921 (2001).

3 Mouse Genome Sequencing, C. *et al.* Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520-562 (2002).

4 Roopra, A. *et al.* Transcriptional repression by neuron-restrictive silencer factor is mediated via the Sin3-histone deacetylase complex. *Mol Cell Biol* **20**, 2147-2157 (2000).

5 Andres, M. E. *et al.* CoREST: a functional corepressor required for regulation of neural-specific gene expression. *Proc Natl Acad Sci U S A* **96**, 9873-9878 (1999).

6 Naruse, Y., Aoki, T., Kojima, T. & Mori, N. Neural restrictive silencer factor recruits mSin3 and histone deacetylase complex to repress neuron-specific target genes. *Proc Natl Acad Sci U S A* **96**, 13691-13696 (1999).

7 Lepagnol-Bestel, A. M. *et al.* DYRK1A interacts with the REST/NRSF-SWI/SNF chromatin remodelling complex to deregulate gene clusters involved in the neuronal phenotypic traits of Down syndrome. *Hum Mol Genet* **18**, 1405-1414 (2009).

8 Chong, J. A. *et al.* REST: a mammalian silencer protein that restricts sodium channel gene expression to neurons. *Cell* **80**, 949-957 (1995).

9 Schoenherr, C. J. & Anderson, D. J. The neuron-restrictive silencer factor (NRSF): a coordinate repressor of multiple neuron-specific genes. *Science* **267**, 1360-1363 (1995).

10 Johnson, R. *et al.* Identification of the REST regulon reveals extensive transposable element-mediated binding site duplication. *Nucleic Acids Res* **34**, 3862-3877 (2006).

11 Bruce, A. W. *et al.* The transcriptional repressor REST is a critical regulator of the neurosecretory phenotype. *J Neurochem* **98**, 1828-1840 (2006).

12 Prada, I. *et al.* REST/NRSF governs the expression of dense-core vesicle gliosecretion in astrocytes. *J Cell Biol* **193**, 537-549 (2011).

13 D'Alessandro, R., Klajn, A. & Meldolesi, J. Expression of dense-core vesicles and of their exocytosis are governed by the repressive transcription factor NRSF/REST. *Ann N Y Acad Sci* **1152**, 194-200 (2009).

14 Pance, A., Livesey, F. J. & Jackson, A. P. A role for the transcriptional repressor REST in maintaining the phenotype of neurosecretory-deficient PC12 cells. *J Neurochem* **99**, 1435-1444 (2006).

15 Birney, E. *et al.* Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**, 799-816 (2007).

16 Consortium, E. P. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74 (2012).

17 Koch, C. M. *et al.* The landscape of histone modifications across 1% of the human genome in five human cell lines. *Genome research* **17**, 691-707 (2007).

18 Johnson, D. S., Mortazavi, A., Myers, R. M. & Wold, B. Genome-wide mapping of in vivo protein-DNA interactions. *Science* **316**, 1497-1502 (2007).

19 Bruce, A. W. *et al.* Functional diversity for REST (NRSF) is defined by in vivo binding affinity hierarchies at the DNA sequence level. *Genome research* **19**, 994-1005 (2009).

20 Kreisler, A. *et al.* Regulation of the NRSF/REST gene by methylation and CREB affects the cellular phenotype of small-cell lung cancer. *Oncogene* **29**, 5828-5838 (2010).

21 Huang, Z. & Bao, S. Ubiquitination and deubiquitination of REST and its roles in cancers. *FEBS Lett* **586**, 1602-1605 (2012).

22 Zhang, D. *et al.* Non-CpG methylation by DNMT3B facilitates REST binding and gene silencing in developing mouse hearts. *Nucleic Acids Res*, (2016); in press.

23 Garriga-Canut, M. *et al.* 2-Deoxy-D-glucose reduces epilepsy progression by NRSF-CtBP-dependent metabolic regulation of chromatin structure. *Nat Neurosci* **9**, 1382-1387 (2006).

24 Kaneko, N., Hwang, J. Y., Gertner, M., Pontarelli, F. & Zukin, R. S. Casein kinase 1 suppresses activation of REST in insulted hippocampal neurons and halts ischemia-induced neuronal death. *J Neurosci* **34**, 6030-6039 (2014).

25 Zuccato, C. *et al.* Widespread disruption of repressor element-1 silencing transcription factor/neuron-restrictive silencer factor occupancy at its target genes in Huntington's disease. *J Neurosci* **27**, 6972-6983 (2007).
26 Conforti, P. *et al.* Binding of the repressor complex REST-mSIN3b by small molecules restores neuronal gene transcription in Huntington's disease models. *J Neurochem* **127**, 22-35 (2013).
27 McClelland, S. *et al.* The transcription factor NRSF contributes to epileptogenesis by selective repression of a subset of target genes. *Elife* **3**, e01267 (2014).
28 Correia, N. C., Arcangeli, M. L., Pflumio, F. & Barata, J. T. Stem Cell Leukemia: how a TALented actor can go awry on the hematopoietic stage. *Leukemia* **30**, 1968-1978 (2016).
29 Porcher, C., Chagraoui, H. & Kristiansen, M. S. SCL/TAL1: a multi-faceted regulator from blood development to disease. *Blood* (2017).
30 Dhami, P. *et al.* Genomic approaches uncover increasing complexities in the regulatory landscape at the human SCL (TAL1) locus. *PLoS One* **5**, e9059 (2010).
31 Follows, G. A. *et al.* Identifying gene regulatory elements by genomic microarray mapping of DNaseI hypersensitive sites. *Genome research* **16**, 1310-1319 (2006).
32 Strahl, B. D. & Allis, C. D. The language of covalent histone modifications. *Nature* **403**, 41-45 (2000).
33 Barth, T. K. & Imhof, A. Fast signals and slow marks: the dynamics of histone modifications. *Trends Biochem Sci* **35**, 618-626 (2010).
34 Dhami, P. *et al.* Complex exon-intron marking by histone modifications is not determined solely by nucleosome distribution. *PLoS One* **5**, e12339 (2010).

THEME II: PLURIPOTENCY & THE ACQUISITION OF CELL-FATE IN THE PREIMPLANTATION MOUSE EMBRYO.

3.1 How one met the early mouse embryo.

As explained in the preface of this thesis, one regards ones time at the Sanger Institute extremely fondly but acknowledges that towards the end of that period one had started to yearn for a return to some more hypothesis, rather than primarily technology, driven research. This culminated in ones appointment to the group of Prof. Magdalena Zernicka-Goetz at the University of Cambridge's Gurdon Institute, to study mechanisms underpinning the derivation of cell-fate during preimplantation mouse embryo development. The initial synergy behind this transition was the then novel observation that the majority of 4-cell stage mouse embryos exhibit inter-cell heterogeneity in the overall levels of a particular chromatin post-translational histone modification; asymmetric-dimethylation of arginine-17 & -26 of histone H3 (R17/26H3me2a), that is associated with transcriptionally active chromatin[1]. Moreover, that enhanced levels of R17/26H3me2a, catalysed by the arginine methyl-transferase Carm1, were associated with progeny cells ultimately adopting a pluripotent cell-fate (eventually colonising the epiblast/EPI – see below) rather than differentiating and forming one of the two extraembryonic tissues (trophoectoderm/TE or primitive endoderm/PrE – also see below) in the implantation competent late blastocyst (E4.5) stage embryo. Therefore, it was to work on a project relating to Carm1, chromatin and its relationship to pluripotency that one started (initially in mouse embryonic stem/ES cells) but inevitably expanded to include research into other projects, notably on the TE-specific transcription factor Cdx2; indeed, one became so enthralled with this the earliest window of mammalian embryo development that one also contributed to well-cited reviews and theoretical publications, co-authored with Prof. Zernicka-Goetz[2,3] and eventually established ones own research group, here at the Faculty of Science, University of South Bohemia, to develop this interest further.

3.2 An introduction to mouse preimplantation embryo development.

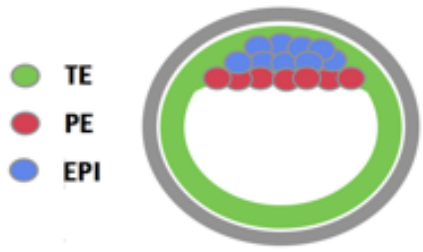

Figure. 8: Cell lineage segregation in the mouse late blastocyst stage embryo.

The mouse represents the most convenient and utilised model for the study of mammalian preimplantation stage embryonic development; the earliest period of development, characterised by the emergence from a fertilised oocyte (*i.e.* the totipotent zygote), via rounds of consecutive cleavage cell divisions, of a highly structured/ordered blastocyst stage embryo (after 4.5 days/E4.5 – reviewed in[2,4,5]), capable of uterine implantation and continued post-implantation development. The blastocyst, comprises three distinct cell lineage populations (Fig. 8), two that have initiated differentiation and will ultimately give rise to supporting extraembryonic tissues after implantation (*i.e.* trophectoderm/TE and primitive endoderm/PrE), and a third epiblast (EPI) lineage that remains pluripotent and serves as a progenitor pool for all subsequent embryonic cells. The TE comprises an epithelium of cells on the outer surface of the blastocyst, encapsulating both a second cell population called the inner cell mass (ICM) and

the fluid filled blastocyst cavity and are polarised along their apical-basolateral axis (the radial axis of the embryo); a process key to specifying and maintaining their fate[2-5]; expanded below. The PrE comprises another monolayer of polarised cells at the interface between the ICM and the cavity, whereas the non-differentiating pluripotent EPI resides deep within the ICM[2,4,5]. Precisely how the blastocyst, arguably the first recognisable tissue of mammalian development, appropriately develops, and the exact identity of the regulatory mechanisms at work, remains one of biology's most fundamental and enduring questions (especially when considered against the extensive spatio-temporal constraints and the remarkable capacity of the mammalian embryo to successfully regulate its developmental programme, unlike other model metazoan species such as the nematode worm *Caenorhabditis Elegans*, in response to perturbations, such as addition or removal of cells, *i.e.* 'regulative' development[6]). It is against this fundamental backdrop ones research group, using experiments similar to those described herein, seek to identify and characterise the genes/proteins and the associated mechanisms that impact on cell-fate during this time. Indeed, in a general context, an increased understanding of such early developmental events retains the potential to impact human reproductive health, as the developmental period in question is paralogous to that mimicked in human infertility treatments, such *as in vitro* fertilisation (IVF). Accordingly, a more fundamental appreciation of the mechanisms that naturally govern development during this time, is a necessary prerequisite to furthering our understanding about how such early development can also become perturbed, with often drastic and terminal consequences for the human embryo; moreover, such insights are potentially informative to wider mammalian development, including that of commercially relevant species such as cattle. Moreover, novel insights gained from researching the preimplantation developmental period have the important potential to further our fundamental understanding of the mechanisms governing cellular potency and the constraints and abilities of developing cells to differentiate towards specialised functions, that in the clinic may be put to meaningful therapeutic use; *e.g.* tissue repair/replacement. Thus, active research into the mammalian preimplantation developmental period is of great fundamental and potentially applied interest and the mouse model is ideally positioned as the most retractable system in which to conduct such important investigations. Accordingly, the following sections will provide a brief summary of our current knowledge regarding the derivation of the TE and ICM lineages, prior to blastocyst formation, and how the ICM then becomes further refined to yield the EPI and PrE lineages during the maturation of the blastocyst, prior to uterine implantation.

3.3 'The first cell-fate decision'; specification of the TE and ICM.

Central to the acquisition of appropriate cell-fate, during preimplantation mouse embryo development, is the spatial segregation of cells to either inner (encapsulated) embryo positions, where they ultimately contribute to the either of the ICM lineages (*i.e.* EPI or PrE), or relative outer positions, where they will form TE[7]. Such spatial segregation arises as a result of the orientation of the cell divisions occurring at the 8- to 16- and 16- to 32-cell stage transitions[8]. Accordingly, apical-basolaterally polarised 8-cell (and outer 16-cell) stage blastomeres (an archaic term for cleavage derived cells of the preimplantation stage embryo) can either divide to yield one outer residing and polar TE-progenitor daughter cell and a second encapsulated and apolar ICM cell (classically referred to as 'differentiative' or 'asymmetric' divisions), or they can divide to result in both daughter cells occupying outer

positions ('conservative' or 'symmetric' divisions)[8]; both of which remain, theoretically, polarised along their apical-basolateral axis . Thus, such divisions provide progenitor cells specific to each blastocyst lineage with an appropriate spatial identity, also reflected in their polarisation status. Although the above nomenclature relating to division orientation and the generation of spatially distinct cell populations remains within the parlance of the field at large, recent evidence suggest most inner cells are in-fact generated, post-cell division, by neighbouring cell engulfment and/or active internalisation processes, rather than perfectly aligned asymmetric divisions[9]; in a process driven by heterogeneities in cortical actomyosin contractility and the extent of intra-cellular apical-basolateral polarity[9-12]. It is important to note that the generation of such heterogeneities are themselves are nevertheless related to the orientation angle of the preceding divisions, as reflected by the amount of the apical domain, and hence apical-basolateral polarity, the relative daughter cells will inherit. However, the resulting spatial delineation (*i.e.* separation of TE versus ICM progenitors) is often referred to as the 'first cell-fate decision' and is distinct to the 'second cell-fate decision' whereby ICM cells segregate to either PrE or EPI populations (see below). Such spatial segregation of prospective TE and ICM linages is also accompanied by the establishment of specific and distinct transcription factor driven gene expression programmes required to successfully specify and maintain TE differentiation or retain ICM pluripotency, respectively. For example, the transcription factors Tead4[13] (discussed below, in the context of apical-basolateral polarity and hippo-signalling), Cdx2[14], Gata3[15] and Tfap2c[16] participate in appropriate TE formation, whereas in the ICM induction of the *Nanog*, and *Sox2*[17] transcription factor genes regulate the retention of pluripotency, in concert with Oct4/Pou5f1; that forms a mutually antagonistic gene expression regulatory loop with the TE-required transcription factor gene *Cdx2*, to further refine, drive and separate the fate of inner and outer cells[18]. Thus, TE- and ICM-lineage specific transcription factors and the transcriptional circuitries they direct are important but how their expression in the spatially distinct cells of the developing preimplantation mouse embryo is regulated is only beginning to emerge; moreover the crucial importance of the mechanistic role played by intra-cellular apical-basolateral polarity in this process has been brought to the forefront of our collective understanding[19].

Indeed, cell polarity, as observed in all epithelia (reviewed in[20]), can be defined as an asymmetric organisation of intra-cellular structural and effector molecules that together impart specific individual cellular functionalities to an individual cell and that of its descendants. Thus, the onset of intra-cellular polarity results in the creation of two highly distinct plasma membrane domains, enriched in specific and distinct multi-protein polarity complexes; termed apical (not in cell-cell contact) or basolateral domains (at cell-cell contact regions). These domains are separated by strong cell adhesion junction complexes (*e.g.* adherens junctions/AJs or tight junctions/TJs), located at the most apical extremity of lateral cell surfaces and provide effective diffusion barriers that ensure structural and functional domain integrity. In the mouse preimplantation embryo model, prior to the first possible round of differentiative/conservative cell divisions, 8-cell stage embryos undergo a process of compaction (*i.e.* an increase in inter-cell adhesions and flattened cell morphology[4]). Compaction is mediated by the adhesion molecule E-cadherin/Cdh1, leading to the formation of AJs[21], restricting Cdh1 (plus other AJ components *e.g.* β-catenin/Ctnnb1) to cell-cell contact/basolateral regions. Another key compaction determinant, at this time, is the ERM-domain containing protein Ezrin/Ezr, whose intra-cellular distribution also

becomes polarised after redistribution from the entire cell cortex to only the apical region (where it exists in its activated phospho-form, (p)Ezr[22]). Ezr is a mediator between the plasma membrane and underlying F-actin network, promoting the formation of the microvilli typical of the apical domain[23]; additionally its redistribution also permits Cdh1 mediated basolateral cell-adhesion via the formation of AJs between neighbouring, microvilli free, blastomere plasma membranes. Concomitant with compaction, 8-cell stage blastomeres also undergo *de novo* apical-basolateral polarisation (see reviews[4,24,25]) and form TJs, typified by accumulation of the TJ protein Tjp1 at apical-lateral interfaces[26]. Such polarisation is characterised by the accumulation, and then spreading, of polarity factors of the Par-aPKC complex at the extremity of the apical pole; specifically, the mouse homologs of PAR3 (Pard3), PAR6 (represented by expression of the *Pard6b* isoform of the gene) and atypical protein kinase C isoforms/aPKCλ/ζ (respectively expressed from the *Prkci* and *Prkcz* genes). Additionally, both, Pard6b and aPKCλ/ζ are localised to forming TJs with Pard3 recruited later, after vacating the apical pole[27,28]. Coincidently, components of the Scribble complex, represented by the murine homologs of Scribble (Scrib), PAR1 (Mark2/3) and LGL (Lgl1), are enriched on basolateral membranes and are excluded from the apical pole[29,30]. These patterns of intra-cellular apical-basolateral polarity persist through the 16-cell stage to the late blastocyst (E4.5), yet are only evident in outer cells, with inner-positioned cells only inheriting basolateral polarity markers and thus becoming apolar (Fig. 9 provides a visual summary[24]). Therefore, following conservative divisions, both outer TE-progenitor daughter cells have the potential to retain apical-basolateral polarity (depending on the angle of cell division, that will dictate how much of the apical pole they inherit and their potential to become internalised) but inner ICM-progenitor daughters, derived via/after differentiative divisions become apolar (lacking an inherited apical domain[8]).

Whilst, the significance of such asymmetric inheritance (regarding ultimate cell-fate) has long been appreciated, only relatively recently has any in-depth mechanistic insight emerged. For example, whilst it was known for some time that direct functional down-regulation of apical-polarity factors (Pard3/Pard6b/aPKCλ) is associated with cells adopting inner positions and ICM-fates[27,31,32] the precise mechanisms were unknown. However, it is now known that apical polarity factors play a central role in regulating the activity of the hippo-signalling pathway (classically known for its role in regulating organ size/growth[33]) and that this pathway in turn directly regulates the transcription of TE-related (and ICM-related) genes (reviewed in[25]). Briefly (and summarised in Fig. 10.), the hippo-signalling transcriptional co-activator effectors, Yap1 and Taz (herein simply referred to as Yap1), are absolutely required for the TE specifying transcription factor Tead4[34] to activate TE-related

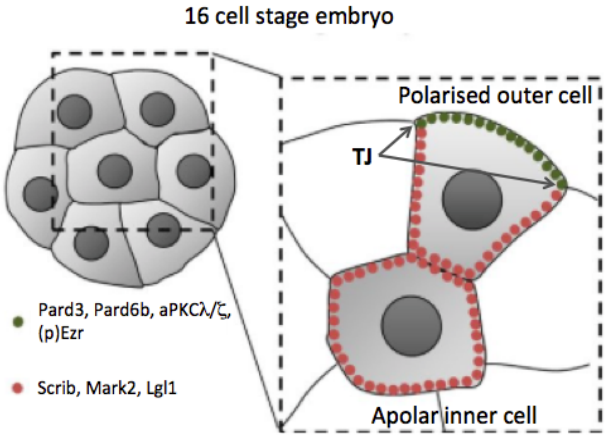

Figure. 9: Differential intra-cellular localisation of apical-basolateral polarity proteins in the outer- and inner-cells of the preimplantation mouse embryo (16-cell stage depicted); note, outer-polarised and inner-apolar cells; Localisation of apical factors depicted in green and basolateral factors in red with tight junctions are highlighted (TJ). Adapted from Ajduk & Zernicka-Goetz (2015)[24].

gene transcription (*e.g. Cdx2*). However, in apolar inner cells, Yap1 is phosphorylated by the active hippo-pathway related kinases Lats1/2 and as a consequence is sequestered to the cytoplasm (by interaction with 14-3-3 proteins) and is eventually degraded[34]. Therefore, despite itself being nuclear localised, Tead4 is appropriately unable to activate TE genes (although interestingly, such activated hippo-signalling is nonetheless required to activate transcription of the pluripotency related TF, *Sox2*, via an as yet unknown mechanism[17]; potentially reflecting a 'relief from transcriptional repression' principle). However, in polarised outer cells, hippo-signalling is supressed, allowing non-phosphorylated Yap1 to enter the nucleus, bind to Tead4 and direct apposite TE-related gene transcription. The key determinant of such differential hippo-signalling activation/suppression is the presence, only in polarised outer-cells, of an apical domain (enriched in apical polarity factors) that directly sequesters the hippo-pathway activator Amot[35-37]; therefore preventing it from associating with the other hippo-related proteins at basolaterally located AJs, required to activate the terminal Lats1/2 kinases[38]. Indeed, intra-cellular hippo-signalling activity is directly dependent on individual apical (*e.g.* aPKCλ/ζ and Pard6b) and basolateral (Mark2/3) polarity factors[36], as their functional down-regulation culminates in ectopic outer-cell basolateral Amot localisation (plus ectopic localisation of other non-targeted polarity factors), enrichment of phosphorylated cytoplasmic Yap1 and ablated Cdx2 (TE marker) protein expression. Moreover, and consistent with pioneering earlier work on Rho-family GTPases (RhoA and Cdc42), ourselves and others (discussed in more detail in section 3.9, below) have recently shown that chemical inhibition of RhoA effector kinases, Rock1/2, also causes outer-cell mis-localisation of apical/basolateral factors, restriction of (p)Ezr to apical pole extremities, defective TJ formation and aberrant hippo-activation (*i.e.* cytoplasmic Yap1 localisation, resulting in blocked TE specification), in an Amot-dependent manner[30,39].

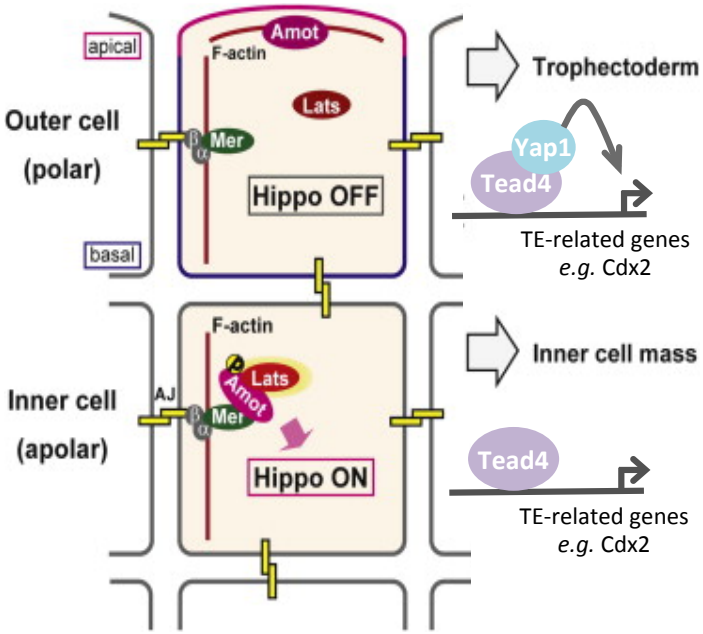

Figure 10: Spatial regulation of hippo-signalling and cell-fate by apical-basolateral polarisation in preimplantation stage mouse embryos. Note, the hippo-pathway activator Amot localises with other components of the pathway at adherens junctions (AJ) in apolar inner-cells, to activate Lats1/2 kinases that phosphorylate the Tead4 TF co-activator Yap1 to prevent its nuclear import; thus TE specific genes are not transcribed. Conversely, in outer polarised cells, Amot is sequestered to the apical domain and thus, the hippo-pathway is not activated, permitting unphosphorylated Yap1 to associate with Tead4 in the nucleus, to activate TE-specific gene expression. Adapted from Hirate *at al.,* (2013)[36].

Therefore to summarise, the first cell-fate decision of preimplantation mouse embryo development involves the spatial segregation of blastomeres to either the relative inner- or

outer- compartments of the embryo; starting from the 8-cell stage. Such spatial allocation is influenced by the orientation of the (outer) cell divisions during the 8- to 16-cell and 16- to 32-cell transitions that can result in the generation of inter-blastomere heterogeneities, reflected in the extent of apical-basolateral polarity and actomyosin contractility, that dictate a cells final spatial position. The specification of the appropriate cell-fate, relative to the allocated spatial position, is also dictated by the presence or absence of apical-basolateral polarity, in outer- and inner-residing cells, respectively. Accordingly, TE cell-fate, and germane gene expression, is driven in polarised outer-cells by the transcriptional complex of Tead4 and Yap1, thanks to actively supressed hippo-signalling, whereas the lack of apical-basolateral polarity in inner-cells permits hippo-pathway activation and the phosphorylation dependent exclusion of Yap1 from the nuclei, resulting in the appropriate failure of Tead4 to activate TE-specific genes (and also potentiating conditions in which pluripotency-related genes can be expressed).

3.4 'The second cell-fate decision'; EPI and PrE formation within the ICM.

At the early blastocyst stage, all ICM cells appear uncommitted to either an EPI or PrE cell-fate as they co-express the transcription factors Nanog (plus Sox2) and Gata6; the respective expression of which ultimately defines the specified /segregated pluripotent EPI and differentiating PrE populations, by the late (E4.5) blastocyst stage (reviewed in[4,5] and summarised in Fig. 11[5,40]). Hence, the establishment of ICM pluripotency, in the form of specified EPI cells, occurs from a cell population in which differentiation has already been initiated (towards PrE; Gata6 expression). However, during blastocyst maturation, the initially uncommitted ICM cells transit through a developmental stage by which they are transformed to express either Nanog or Gata6 (plus other EPI; Sox2, and PrE; Sox17 specific genes) in a mutually exclusive mosaic, termed the 'salt and pepper' pattern[41] (Fig. 11[5,40]); from which lineage specified cells are spatially segregated into the recognisable late blastocyst (E4.5) tissue layers, via active cell-sorting mechanisms (Fig. 8 and Fig. 11[5,40]).

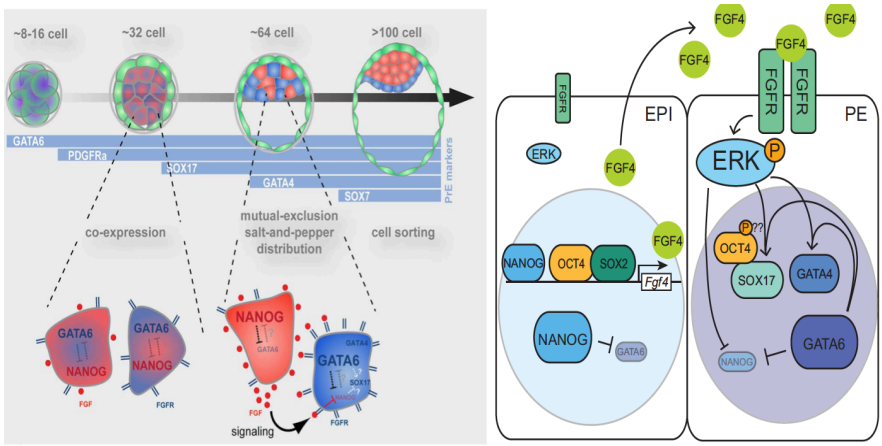

Figure 11: *left*; Mouse blastocyst ICM maturation, detailing the progressive temporal activation of PrE-related transcription factor genes, plus Nanog expression, in the respectively emerging PrE (blue) and EPI (red) lineages; starting from initially uncommitted ICM founding cells co-expressing Gata6 and Nanog, progressing through a random and mutually exclusive cell expression pattern (*i.e.* 'salt and pepper') that culminates, after cell sorting, in refined EPI and PrE tissue layers (TE shown in green). A process driven by Fgf4-signalling from EPI to PrE. *right*; A contemporary mechanistic summary of PrE and EPI cell-fate derivation in the mouse blastocyst - taken from Kang *et al.,* (2013)[40] and Chazaud & Yamanaka (2016)[5], respectively.

It has been proposed that the emergence of such ICM inter-cell heterogeneity is related to previous cell history (whereby the relative developmental timing of TE and ICM founder cell spatial segregation, at either the 8- to 16- or 16- to 32-cell transition, can bias resulting inner cells and their progeny to ultimately

contribute to the EPI or PrE lineages, respectively[42-45]; experimentally developed below in section 3.8, relating to interlinked/integrated cell-fate decisions during preimplantation mouse embryo development), and self-organising gene expression networks, that arise out of essentially stochastic gene expression[46]; *n.b.* these potential mechanisms are often presented as being in conflict, though fundamentally there is no reason why they cannot mutually exist. However it is clearly indisputable that the initiation of ICM heterogeneity is indeed reinforced by the establishment of defining and differential transcriptional regulatory loops (as in the case in the separation of TE and ICM – see above, section 3.3); for example, genetic loss of either the *Nanog* or *Gata6* genes leads to homogenous expression of the opposing transcription factor across all cells of the ICM, implying (in)direct and mutual transcriptional repression mechanisms that reinforce specified EPI/PrE fate (Fig. 11[5,40]) Similarly, it is known that *Gata6* expression is required to set in train the sequential activation of the PrE- specific and required transcription factor genes, *Sox17*, *Gata4* and *Sox7*, [47] and that Oct4, despite being expressed in both EPI and PrE destined cells during blastocyst maturation, can bind either Sox2 (selectively expressed in the EPI specified cells) or Sox17 (expressed in differentiating PrE) transcription factor proteins to effect lineage appropriate gene transcription, and thus reinforce the already specified fate[48]. Interestingly the pan-ICM expression of Gata6 protein seen in *Nanog*$^{-/-}$ null blastocysts is not accompanied by the sequential expression of the stated later PrE markers (*e.g.* Gata4), indicative of a missing differentiation component, but it can be rescued by exogenous addition of recombinant Fibroblast growth factor-4 (Fgf4) protein; the endogenous expression of which is absent in such embryos[49]. Indeed, Fgf4 signalling has a central role in ICM cell-fate derivation given that the emergence of the salt and pepper pattern of EPI and PrE specified cells is blocked in maternal/zygotic *Fgf4*$^{-/-}$ double null embryos and is associated with pan-ICM EPI marker protein (Nanog) expression (although initial *Gata6* activation is not impaired[40]). Consistently, pharmacological inactivation of Fibroblast growth factor receptors (Fgfrs) and the known functionally downstream Erk1/2 mitogen activated kinases, from a point prior to blastocyst formation, are also associated with a similar cell-fate specification failure phenotypes in the ICM; moreover the exogenous addition of Fgf4 ligand causes the inverse pan-ICM conversion to PrE marker gene expression[50-52]. Indeed, given the facts that bimodal expression of the *Fgf4* gene represents the earliest know heterogeneity between cells of the ICM, that *Fgf4* expression specifically marks EPI cells later in development and that *Nanog* mutants do not express Fgf4[46,49], the current model of blastocyst ICM cell-fate derivation is as follows (summarised in Fig. 11 – see right panel for detailed mechanistic summary[5,40]); although please note PrE differentiation has also been shown to be influenced by Bone morphogenetic protein/Bmp and Platelet-derived growth factor alpha/Pdgfα related signalling pathways[5]. Small heterogeneities within the apparently uncommitted ICM cells of early blastocysts (arising as a consequence of cell history and/or stochastic gene expression; referred to above) result in the emergence of two populations of cells. One in which comparatively enhanced Nanog levels antagonise *Gata6* expression (thus preserving pluripotency and demonstrated by the expression of *Sox2*, and potentially also reflected in reduced Fgfr levels to prevent autocrine signalling) and potentiates the production of secreted Fgf4 ligands that signal to the second population of Fgfr expressing uncommitted cells (in which Gata6 levels may have started to dominate). This causes the activation of the downstream Erk1/2 pathway, in this second population of cells, that then functions to supress any Nanog induced pluripotency and facilitates sequential PrE-required gene expression (starting with *Sox17*). As such the two ICM cell-fates become specified and

are detectable as the two mutually exclusive populations represented in the salt and pepper pattern towards the mid-blastocyst (~E3.75 – E4.0) stage; moreover the specification of both PrE and EPI cell-fate is further reinforced by Oct4 in both lineages, by its association with the appropriately expressed Sox2 or Sox17 binding partner, and in the case of the PrE leads to the sequential activation of later markers (*e.g.* Gata4 and Sox7). Furthermore, although ICM cells are fate specified by the emergence of the salt and pepper pattern, experiments pharmacologically manipulating Fgf-based signalling after this time, show that cells remain developmentally plastic and can change fate (up to a point; ~E4.0 – E4.25). Therefore, demonstrating Fgf-based signalling is needed for both the initiation of ICM-cell fate and to maintain it during blastocyst maturation.

Consistent with this notion, we ourselves have very recently publish novel findings demonstrating a distinct role for p38 mitogen-activated-kinases (p38-Mapks) in regulating ICM cell-fate blastocyst maturation[50] (*i.e.* distinct from that already characterised for the related Erk1/2 pathway[23,52]); further developed below, see section 3.10. Specifically, that p38-Mapk activity is required during the early stages of blastocyst maturation (prior to the emergence of the salt and pepper pattern at ~E3.75) to enable uncommitted cells to specify the PrE lineage; moreover, that this activity is functionally downstream of activated Fgfr and potentially non-canonical Bmpr signalling mediated by the mitogen-activated-kinase-kinase-kinase, Tak1. Accordingly, we have shown p38-Mapk specific pharmacological inhibition of *in vitro* cultured (E3.5 - E4.5) maturing mouse blastocysts is associated with a profound deficit in committed (*i.e.* Gata4 positive) PrE cells and increased numbers of uncommitted cells expressing both Nanog and Gata6 protein. Crucially, this phenotype is distinct from that induced by inhibiting the Erk1/2 pathway (by pharmacologically targeting Mek1/2 – as referenced above), as the window of p38-Mapk inhibition sensitivity only extends to E3.75 versus the ~E4.0 - E4.25 limit observed under Erk1/2 inhibited conditions. Therefore, implicating p38-Mapk and Erk1/2, both activated by Fgf-based signalling mechanisms, as the initial enablers (*i.e.* permitting initiation of PrE cell-fate specification, evidenced in the salt and pepper pattern) and drivers (*i.e.* ensuring maintained PrE-related gene expression) of ICM cell-fate derivation, respectively[50]; each responding in a temporally defined way to Fgf-based signals secreted by EPI progenitors to ultimately derive PrE.

In summary, what is classically referred to as the 'second cell-fate decision' of preimplantation mouse embryo development involves the specification of initially uncommitted early blastocyst (E3.5) ICM cells to ultimately derive the segregated PrE or EPI lineages of the late blastocyst (E4.5) stage embryo; whereby Fgf-based signalling (plus potentially other signalling ligands) originating from the emerging EPI lineage population of cells acts to first initiate and then maintain PrE differentiation in a second population of receptive cells, reinforced by sequential activation of PrE-required transcription factors. Such PrE differentiation is under the control of active p38-Mapks, thus enabling PrE specification (evidenced by the salt and pepper pattern at ~E3.75) but is latterly driven by activation of the Erk1/2 pathway; thus separating the extraembryonic PrE lineage from the pluripotent EPI in which differentiation is resisted (most probably by a mechanism involving a failure to receive the extra-cellular ligand-based cues).

3.5 Carm1 acts at the fulcrum of mouse ES pluripotency retention and initiated differentiation.

As described in the introductory section to this, ones second, research theme (see 3.1 and also the overall preface), ones arrival at the Gurdon Institute coincided with ones new group recent publication of a manuscript, in *Nature*, detailing quantifiable differences in global chromatin levels of the post-translational histone modifications R17/26H3me2a between blastomeres in the majority of 4-cell stage embryos (*i.e.* those exhibiting so-celled tetrahedral cleavage patterns that typically exhibit the most viable development when transferred to pseudo-pregnant foster mother mice). Moreover, that such relative elevated levels correlated with blastomeres yielding progeny biased to populate the pluripotent ICM blastocyst lineage, rather than the differentiating and outer-residing extra-embryonic TE; as substantiated by the fact that clonal over-expression of the enzyme responsible for the deposition of the R17/26H3me2a epigenetic chromatin marks, Carm1 (achieved by microinjecting one blastomere at the 2-cell stage with recombinant derived Carm1 mRNA and a fluorescent lineage tracer; DsRed mRNA), potentiated pluripotency and elevated expression of *Nanog* and *Sox2*[1]. Hence, ones first set project was to exploit mouse ES cell cultures systems to perform biochemical and genomics scale experiments, not feasibly suited to the small amounts of starting material offered by preimplantation mammalian/mouse embryos, to better understand the mechanisms by which R17/26H3me2a levels and Carm1 function to influence cell-fate; accordingly this work yielded a joint first author publication in *Stem Cells*[53].

Specifically, these experiments demonstrated that the promoter regions of both the *Oct4* and *Sox2* genes are enriched for Carm1 protein binding and the presence of R17/26H3me2a modified chromatin, when assayed in ES cells using conventional ChIP assays; signals that could be ablated if the same parental ES cells lines were transfected with Carm1 sequence specific RNAi constructs (*i.e.* shRNA producing plasmids), to effect efficient knockdown in endogenous *Carm1* gene expression. Although, curiously the *Nanog* gene promoter did not display any ChIP derived signals consistent with Carm1 occupancy. Crucially and importantly, the RNAi mediated knockdown of endogenous *Carm1* expression also caused morphological changes in ES cell colony appearance that are consistent with the initiation of spontaneous differentiation, plus the loss of alkaline phosphatase staining (a classical marker of all pluripotent stem cells); features that could be rescued by co-transfection of recombinant derived Carm1 expression 'rescue' constructs (specifically mutated to render them insensitive to the shRNA used but that also exploit redundancies in the genetic code to maintain the Carm1 primary amino acid sequence). Real-time quantitative PCR and western blot analyses revealed that knockdown of endogenous *Carm1* expression was associated with profoundly reduced expression of the pluripotent triad of *Oct4*, *Sox2* and *Nanog* transcription factor gene expression at both the transcript and protein level, respectively; data supported by similar microarray based assays of global transcript levels. Interestingly, such microarray data not only confirmed the down-regulation of pluripotency-related genes but also demonstrated clusters of commonly up-regulated genes enriched in gene ontological/GO-terms associated with 'development' and 'differentiation' that are considered markers of all tissue germ layers. Lastly, experiments conducted on the same parental ES cell lines in which endogenous *Carm1* expression was augmented by the over-expression of recombinant Carm1, showed opposing effects; namely such cells resisted

differentiation, under *in vitro* culture conditions sufficient to induce well characterised differentiation in the control ES cells.

In summary, this work demonstrates that Carm1 exerts its pluripotency promoting effects in ES cells by directly associating with the promoters of the *Oct4* and *Sox2* genes, depositing transcriptionally activating R17/26H3me2a chromatin marks; Sox2 and Oct4 proteins most probably activate, as a secondary consequence, the transcriptional expression of the *Nanog* gene, thereby ensuring the ES cell appropriate expression of the pluripotency-related triad of transcription factors. Equally, loss of *Carm1* expression (be it consequent to experimental intervention or differentiative inductive cues) causes the breakdown of this regulatory network and initiates conditions permissive to the induction of differentiation towards all germ layers (Fig. 12); *i.e.* as could be predicted by the aggregated phenotypes reported for the genetic loss of each of the pluripotent triad genes alone[54-58]. These concepts have been later developed in further preimplantation mouse embryo studies from the Zernicka-Goetz group[59-61].

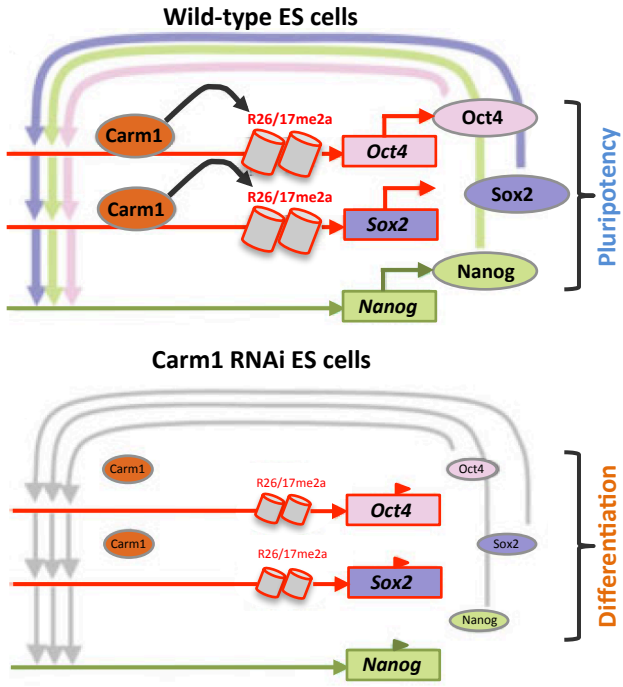

Figure 12: A model for the modulation of pluripotency by Carm1: Carm1 activity is required for embryonic stem (ES) cell pluripotency because cells differentiate after its down regulation (compare upper and lower panels) Pluripotency is maintained in wild-type ES cells in part as a result of asymmetric di-arginine methylation of histone H3 (R17/26me2a) by Carm1 at the *Oct4* and *Sox2* promoters. The Oct4 and Sox2 proteins act on their own gene enhancers and the *Nanog* gene enhancer (upper panel). The resultant expression of the triad of pluripotent – related transcription factors (Nanog, Sox2 and Oct4) ensures a pluripotent ES cell state. However, if Carm1 levels (or activity) is reduced, the direct R17/26me2a mediated activation of the *Oct4* and *Sox2* genes is impaired, leading to reduced *Nanog* expression and conditions permissive to spontaneous differentiation, representative of all germ layers.

3.6 Evidence for a functional role for maternally inherited Cdx2 during preimplantation mouse embryo development.

Although ones transition from the Sanger to Gurdon Institutes initially involved working on what might be described as a 'chromatin-related bridging' project, as described in the preceding section (*i.e.* R17/26H3me2a/Carm1 related), one quickly became assimilated into the wider objectives of the preimplantation embryo development field represented by Prof. Zernicka-Goetz's research group. As such, one began collaborating on a project related to the TE specific transcription factor Cdx2 (for general contextual background on the relevance of the *Cdx2* gene, see section 3.2). The role of the *Cdx2* gene in preimplantation development originates from observations that outer, but not inner, cells of the blastocyst display anti-Cdx2 immuno-reactivity[62] and that two independent genetic knockout studies

report peri-implantation lethality of *Cdx2*$^{-/-}$ zygotically null mouse embryos[63,64]. Subsequent, and comprehensive characterisation of such zygotic null *Cdx2*$^{-/-}$ phenotypes has revealed defective epithelial integrity within the blastocyst TE, that is unable to support the necessary expansion of the cavity and hence embryo hatching from the *zona pellucida* and eventual implantation into the uterine endometrium[14]. Still later investigations, originating in ES cells, have demonstrated that induced expression of recombinant Cdx2 is sufficient to drive TE differentiation and that a transcriptional complex of Cdx2 and Oct4 acts to reciprocally repress transcription of each other genetic loci[56], thus contributing to the resolution of the Cdx2 blastocyst protein expression domain (described above), from a initial population of morula stage blastomeres that can express both Oct4 and Cdx2 protein[65,66]; although current opinion suggests this resolution is primarily driven by differential suppression of the hippo-signalling pathway between derived outer- and inner-cell embryo populations, respectively[36] (introduced in section 3.2 above and expanded upon in sections 3.5 and 3.6 below).

In regard to the establishment of intra-blastomere apical-basolateral polarity, it had been reported that zygotic null *Cdx2*$^{-/-}$ embryos are phenotypically indistinguishable from control embryos, suggesting zygotically derived Cdx2 protein functions at a later developmental stage[66]. However, shortly before ones joining, the Zernicka-Goetz group had published a paper reporting phenotypic effects associated with *Cdx2* over-expression and down-regulation (using both microinjected recombinant Cdx2 mRNA and transcript specific RNAi constructs, respectively; to generate clones of *Cdx2* dysregulated cells) that developmentally preceded those previously described when both zygotic Cdx2 alleles were deleted; manifest in increased apical-basolateral polarisation and increased inner cell contribution, respectively[67]. It was therefore hypothesised that these data may indicate the presence of functionally relevant Cdx2 mRNA that was maternal in origin (*i.e.* had been provided from the oocyte during its maturation). This was because such maternally derived Cdx2 transcripts would be similarly targeted for destruction as any derived by transcription of either of the two zygotic *Cdx2* alleles when using an RNAi-mediated approach. However, such maternal Cdx2 transcripts would persist (at least partially) in zygotic *Cdx2*$^{-/-}$ embryos, given such embryos can only be derived by crossing heterozygous *Cdx2*$^{+/-}$ animals (as the homozygous gene knock out is embryonic lethal by the peri-implantation stage). These data therefore afforded the possibility that the developmentally earlier phenotypes associated with the RNAi-mediated approach to dysregulated *Cdx2* expression were due to the additional destruction of functionally relevant and maternally provided Cdx2 transcripts.

Accordingly, it was against this backdrop that a project was undertaken to directly assay for maternally provided Cdx2 mRNA and any associated function[68]. Using a combination of pre-existing mRNA microarray expression data[69], quantitative RTPCR and Cdx2 mRNA specific fluorescent *in situ* hybridisation/FISH, the presence of maternally provided Cdx2 transcripts was confirmed; moreover, that these could be depleted by microinjecting Cdx2-specific RNAi constructs into recovered early zygote stage embryos, prior to the major onset of zygotic genome activation (*i.e.* the derivation of transcripts by the direct transcription of embryonic alleles) that normally occurs during the late 2-cell stage (reviewed in[70]). Furthermore, it was also confirmed, by the selective utilisation of alpha-amanitin (an inhibitor of global transcription), that such transcripts are translated into detectable Cdx2 protein. To assay whether such maternal transcripts harbour any function, early zygote

stage embryos were microinjected with Cdx2-specific RNAi constructs (as temporally proximal to point of fertilisation as was experimentally feasible) to eliminate both maternally provided and any subsequently zygotically derived Cdx2 transcripts; control groups comprising microinjection of DsRed mRNA alone or RNAi constructs without any mouse genome targets were also included. Developmental progression of all embryo groups was then monitored by time-lapse fluorescent microscopy (as transgenic embryos expressing EGFP-tagged histone H2B were used to aid subsequent analysis) from the 2-cell to mid-blastocyst stage, and the number, lineage (*i.e.* pedigree) relative positioning (*i.e.* outer-TE or inner-ICM) and viability of individual cells recorded (Fig. 13). The results of these experiments clearly demonstrated arrested and defective developmental phenotypes in 88.6% of Cdx2 RNAi microinjected embryos, occurring before the onset of blastocyst maturation and associated cavity expansion (the previously described phenotype observed in zygotic null *Cdx2*$^{-/-}$ mutant embryos[14]). Moreover such delayed development (compare timings in Fig. 13b) was also accompanied by atypical apoptotic cell death, not observed in DsRed mRNA alone microinjected control groups that successfully developed to the blastocyst stage on an appropriate timescale, without any incidence of arrest. The removal of both maternal and zygotically derived Cdx2 transcripts therefore resulted in the segregation of two phenotypic distinct groups, comprising embryos arresting and exhibiting cell death in the 8- to 16- cell stage transitions (morula; group 1) and others characterised by a similar phenotype in the transit from the 16- to 32 cell stages (pre-blastocyst; group 2). Despite some heterogeneity in the exact timing of arrested development, each of these phenotypes was more severe in nature than those previously observed in zygotic null *Cdx2*$^{-/-}$ mutant embryos[14]; moreover they were specific to the removal of Cdx2 transcripts, as co-microinjection of Cdx2 RNAi constructs with excess amounts of recombinant Cdx2 mRNA was able to rescue the blastocyst developmental program. Hence, demonstrating an additional and previously unrecognised role for maternally provided Cdx2 transcripts, that precedes that characterised for zygotically derived Cdx2. Investigating these temporally earlier developmental phenotypes revealed a breakdown in intracellular apical-basolateral polarity from the 8-cell stage (plus accompanying reductions in polarity factor mRNA expression and atypically absent Nanog transcripts by the 16-cell stage); suggesting maternal Cdx2 mRNA has a role in ensuring appropriate establishment of apical-basolateral polarity, that in turn influences the specification/segregation of outer-TE and inner-ICM cells from the 16-cell stage onwards, that is then maintained by the derivation of zygotically derived Cdx2 transcripts. Crucially, very similar results were achieved when both maternal and zygotic transcripts were also functionally targeted, as part of this study, using alternative and specific siRNA and morpholino microinjection strategies. Notwithstanding these controls, another study, also using a siRNA mediated approach to remove maternal and zygotic Cdx2 transcripts, claimed their observations were consistent with no significant functional role for the maternal Cdx2 transcripts they had detected and reported; insisting the phenotypes uncovered were indistinguishable from those observed in zygotic null *Cdx2*$^{-/-}$ mutant embryos (*i.e.* a failure to maintain an expanding blastocyst cavity[71]). Accordingly, a series of commentary correspondences were exchanged between the authors of both studies, discussing the merits and drawbacks of each others technique and offering potential solutions for the apparent discrepancies[72,73]; such explanations ranged from the effectiveness of the RNAi constructs employed and the adequacy of microinjection protocols and controls used (*i.e.* being able to confirm construct delivery and integrity using co-injected mRNAs for fluorescent proteins), through the potential for differential induction

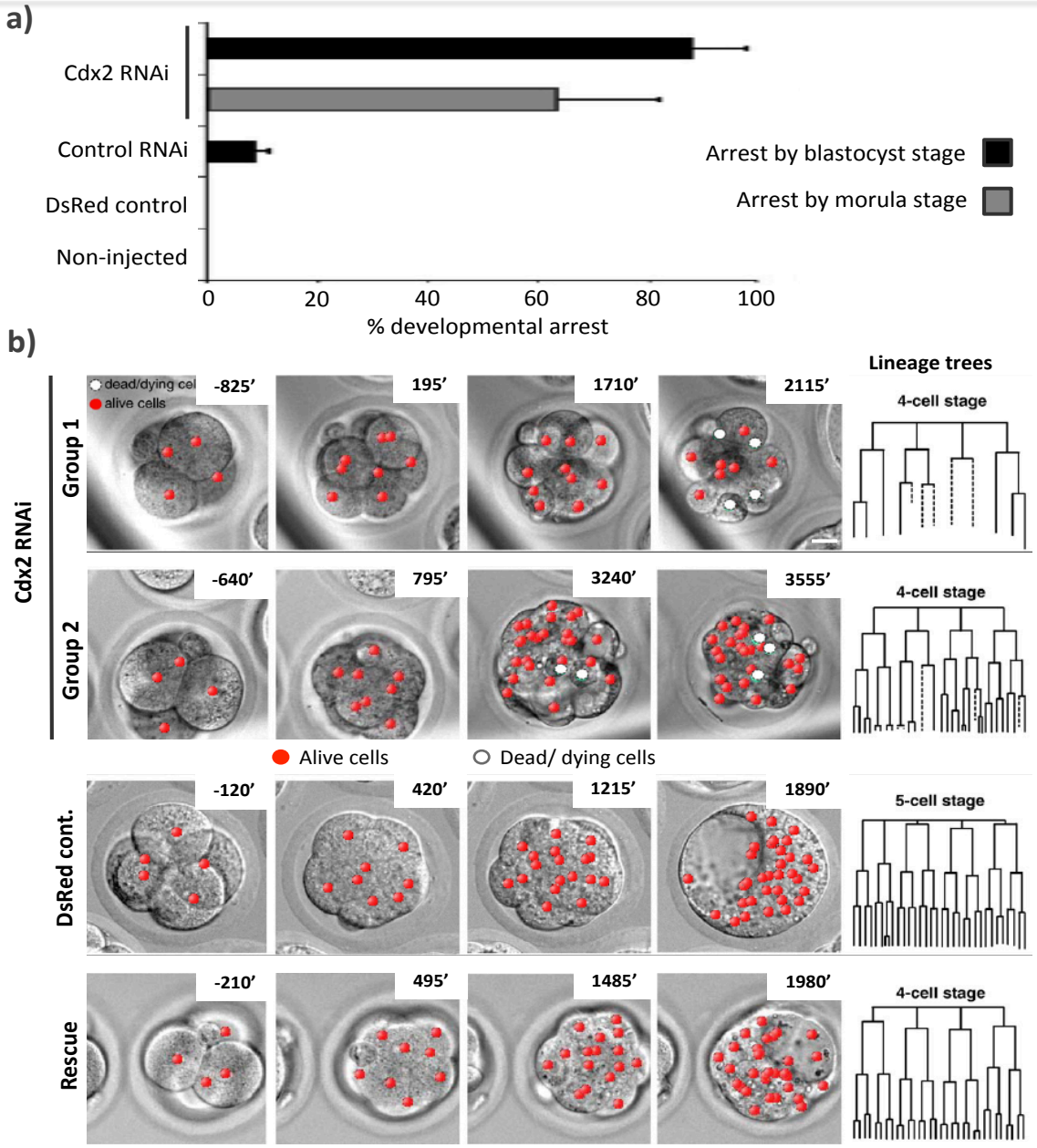

Figure 13: Depletion of maternal and zygotically derived Cdx2 causes developmental arrest and cell death prior to blastocyst stage: a) Rates of developmental arrest in groups of early zygotes microinjected with 'Cdx2 RNAi' constructs by either the morula (16-cell) or blastocyst (~32-cells) stages versus, control microinjected groups (using GFP RNAi construct – 'control RNAi', or DsRed mRNA – 'DsRed control') or 'non-injected controls'. **b)** Time-lapse DIC micrographs of individual Cdx2 RNAi, control DsRed mRNA or Cdx2 RNAi + Cdx2 mRNA (molar excess – 'rescue') microinjected early zygotes in culture. Red circles represent location of GFP-H2B marked nuclei; white circles apoptotic cell nuclei. Stated times (minutes) relative to 8-cell stage entry. Scale bar = μm. Lineage/pedigree trees for every cell from the 4-cell stage, for each embryo example, are given to the right (dashed lines denote dead cells); *n.b.* Cdx2 RNAi microinjected embryos segregated into two groups with arrested development at the 8- to 16-cell stage or at the 16- to 32-cell transition (compare timings with 'DsRed cont.'). Equally, note co-injection of Cdx2 RNAi + excess Cdx2 mRNA ('rescue') is sufficient to restore the developmental programme to equivalency with control ('DsRed' cont.') embryos.

of regulatory mechanisms (possibly dependent on relative knockdown efficacy) that could mask the effect of maternal Cdx2 transcript loss, to the notion that both manuscripts actually represent different points on the same maternal Cdx2 transcript knockdown

phenotypic continuum (*i.e.* they actually both described phenotypes that temporally precede failures in blastocyst cavity expansion, that in some cases overlapped). Consequently, one is content to stand by the data and conclusions of the manuscript described in this habilitation thesis. Indeed, it is noteworthy that in the intervening years, the Zernicka-Goetz group have published a second maternal Cdx2-related study, employing a conditional knock-out approach to specifically remove maternal Cdx2 transcripts from the oocyte[74]. In essence, this newer study reports the same findings as those described above; namely the removal of both maternal and zygotically derived Cdx2 transcripts results in arrest and cell death, beginning from the morula stage, that is also associated with a failure in TE specification (indeed this was true in the majority of cases when only the maternal allele was deleted); whereas loss of only the zygotic *Cdx2* alleles was associated with the classically described blastocyst cavity expansion defect. Thus, the collective evidence of these papers is in support of a prominent functional role for maternally provided Cdx2, during preimplantation mouse embryo development. However, one must also acknowledge the existence of yet another more contemporary study, again utilising a conditional maternal genetic knockout approach, that reports no functional requirement for maternally provided Cdx2[75]. Although not openly discussed within the literature, one suspects the reason for the apparent contradiction between the two conditional knockout strategies may reside in the design of the *Cdx2* gene deletion strategies; whereby the later study relies on conditional alleles in which the *Cdx2* transcriptional start site is deleted after exposure to Cre-recombinase, leading to a frame shift mutation in the remainder of *Cdx2* open reading frame[75] (that nonetheless still contains sequence encoding the majority of the Cdx2 gene, including the DNA binding domain), whilst the former utilises a strategy that deletes the last two of three protein-coding exons of the gene[76] (thus permanently removing protein coding sequence including the DNA binding domain). It is therefore possible that intergenic transcription (known to be pervasive in mammalian genomes[77,78]) may be able to reconstitute, even partially, the remnant of the oocyte-specific conditionally deleted allele utilised in the alternative study, in a manner impossible for that emanating from the Zernicka-Goetz lab (due to absence of the majority of *Cdx2* DNA sequence); thus, restoring some maternal Cdx2 protein function that in turn serves to mask the reported pre-blastocyst phenotypes reported in the other studies[68,71,74]. As previously stated, one is confident in the validity of the maternal Cdx2-related data presented in this habilitation thesis, despite the existence of conflicting reports in the literature.

3.7 Establishing preimplantation mouse embryo research at the Faculty of Science, University of South Bohemia in České Budějovice.

The following described projects/manuscripts were conducted here, within the Faculty of Science, University of South Bohemia in České Budějovice, and are derived from research conducted in ones own independent research laboratory, established following ones relocation from Prof. Zernicka-Goetz laboratory at the Gurdon Institute in Cambridge (United Kingdom); such work was conducted with the support of competitive grants awards to oneself. It is noteworthy that no such similar work had been conducted in this university prior to ones relocation and that the laboratory, its infrastructure and operating procedures were developed, sometimes rather laboriously, from scratch; albeit it with substantial support (as gratefully acknowledge above). The philosophy of and approach to understanding mechanisms of cell-fate in the early embryo adopted in ones research group

are outlined on our webpage (http://kmb.prf.jcu.cz/en/laboratories/en-bruce-lab) and in a review article written shortly after its establishment[72].

3.8 Evidence supporting the interlinked nature of blastocyst cell-fate derivation during preimplantation mouse embryo development.

The preimplantation stages of mouse embryo development are characterised by the emergence of three blastocyst cell lineages; the extraembryonic and differentiating TE and PrE, plus the pluripotent EPI. As described in the preimplantation mouse embryo introductory sections above (see 3.1 – 3.4), the derivation of these three lineages has traditionally been conceptualised as two distinct and independent cell-fate decisions; the first in which the spatial segregation of blastomeres, after the 8-cell stage, yields outer-TE and inner-ICM cell populations and the second by gene expression refinements and active cell-sorting within the blastocyst ICM that eventually segregates deeply residing EPI from the superficial PrE cells (reviewed[2,4,5]). However, in recent years this 'two-step' model has come under increasing scrutiny; primarily based on observations from non-invasive *in vitro* time-lapse recordings of embryos throughout the entire preimplantation period, that show biases in the ultimate fate of ICM cells (*i.e.* either EPI or PrE) based upon the developmental timing of the internalisation of their ancestral cells (*i.e.* either after the first possible wave of asymmetric division at the 8- to 16-cell transition, largely yielding EPI progenitors, or as a consequence of the second wave of asymmetric division during the transit from the 16- to 32-cell stage, strongly favouring formation of PrE precursors;[42]). Although, not universally accepted[52], these data heavily imply that not all ICM founder cells are generated with equal potential to contribute to EPI and PrE. Moreover, that ancestral cell history influences such potency, so that EPI progenitors require rapid sequestration to the inner embryo compartment (consequent to the first wave of asymmetric cell divisions), whilst PrE progenitors need to be derived from ancestral cells that reside for longer (*i.e.* ~12 extra hours) on the outside of the embryo, receiving TE-promoting cues (*i.e.* actively supressed hippo-signalling[36]), in order to prime them for PrE differentiation (after internalisation following the second asymmetric wave of cell division). This theory has been conceptualised as the 'integrated cell-fate model'[2,3,79,80], by which it is the relative developmental timing of ICM founder cell internalisation that gives rise to the EPI, PrE or TE blastocyst cell lineages (*n.b.* as the natural consequence of an asymmetric division an outer, potential TE founder cell is also generated). Hence, suggesting that all three lineages are derived in a process that is mechanistically interlinked, rather than being functionally independent; *i.e.* integrated.

Accordingly, a project was undertaken (described below) to functionally test the validity of the proposed integrated cell-fate model[44]. The basic experimental design involved creating fluorescently marked clones of cells within developing preimplantation mouse embryos that were unable to initiate TE differentiation and then to assay the ability of such clones to contribute to the late blastocyst ICM lineages (*i.e.* EPI or PrE, identified by the expression of Nanog and Gata4 marker protein expression, respectively). The inability to initiate TE differentiation was intended to model the early sequestration of inner cells from TE instructive cues, as occurs after the 8- to 16-cell transition/ the first wave of possible asymmetric divisions. It was therefore hypothesised that if the extent of ancestral cell TE induction is unimportant for eventual PrE differentiation in the ICM, TE-inhibited cell clones would not be impaired in their potential to contribute to the PrE; however, if exposure to TE

inductive signals acts to primer PrE differentiation, such clones would be less likely to form PrE and more liable to retain pluripotency and populate the EPI. Accordingly, TE-inhibited cell clones were created in the preimplantation mouse embryo by microinjecting a long dsRNA construct specific for the *Tead4* gene. The *Tead4* gene was targeted because, as described in the introductions above (see section 3.3), it encodes the transcription factor that, together with its transcriptional co-activators Yap1 and Taz, constitute the molecular effectors of outer cell suppressed hippo-signalling, that directs the transcriptional activation of TE-specific/required genes (*e.g. Cdx2* & *Gata3*) and thus, TE differentiation[36]; moreover genetic loss of *Tead4* is embryonic lethal at the blastocyst stage and associated with a complete block in TE specification/formation[13]. After first confirming the efficacy of the Tead4 RNAi constructs, TE-inhibited cell clones of varying size were generated by either microinjecting individual 2-cell or 4-cell stage embryo blastomeres (together with fluorescent dextran bead tracers); microinjected embryos were then *in vitro* cultured to the late blastocyst stage, fixed and processed for immuno-fluorescent confocal microscopy to detect the expression of protein marker genes of each cell lineage. The contribution of TE-inhibited clones to each lineage was calculated and compared to control microinjection conditions. Smaller TE-inhibited clones were also generated by similarly microinjecting the Tead4 RNAi construct in both blastomeres of 2-cell stage embryos, i*n vitro* culturing and then disaggregating them at the non-compacted 8-cell stage; followed by combining a single 8-cell stage TE-inhibited blastomere with a individual stage match but non-manipulated embryo (to form a 1+8 cell chimera; *n.b.* hosting embryo had its *zona pellucida* removed in the process) and subsequent *in vitro* culture to the late blastocyst stage (after which the lineage contribution of the clone was assayed, as previously described; Fig. 14). In all experimental cases, the contribution of Tead4 RNAi induced TE-inhibited clones was not only found to preferentially allocate to the ICM over the TE (as may have reasonably been expected given the characterised block in TE specification known to be caused by loss of Tead4[13] but was also importantly found to be biased in favour of populating the EPI rather than the PrE (*i.e.* TE-inhibited cell clones that segregated to the ICM did not exhibit the same potential to populate either the PrE or EPI as was observed for fluorescently marked clones in control microinjected embryos/chimeras). Hence, the prevention of cell clones from responding to TE-induction cues (*i.e.* supressed hippo-signalling in outer cells) did indeed impair the potential of internalised cells from the clone to contribute to PrE; an observation in accord with the integrated cell-fate model. Further in-depth characterisation of cells from TE-inhibited clones confirmed that their propensity to occupy the EPI was not associated with any precocious or elevated expression levels of the pluripotency-related transcription factor, and *de facto* late blastocyst EPI marker gene, *Nanog*; nor was it associated with any defects in the expression or sub-cellular localisation of apical (Ezrin, Prkci/z and Pard6b) and basolateral (Scrib and Cdh1) polarity/cell adhesion proteins in the would be ancestral outer-cells, within the clone (at both the 16- and 32- cell stages), that in turn give rise to ICM founder cells. Outer-residing TE-inhibited clones were however associated with increasing levels of cytoplasmic Yap protein (comparing 16- and 32-cell stage embryos) but this was confirmed not to be phosphorylated and thus did not correlate with aberrantly activated hippo-signalling. However, quantitative RTPCR analyses confirmed the robustly reduced expression of mRNA transcripts of a number of PrE-specific genes (including, *Dab2*, *Lrp2* and *Fgfr2* – the receptor considered responsible for receiving Fgf-based signals in the ICM to promote PrE differentiation[5]) associated with TE-inhibition, at both the 16- and 32-cell stages. Furthermore, high-resolution confocal microscopic

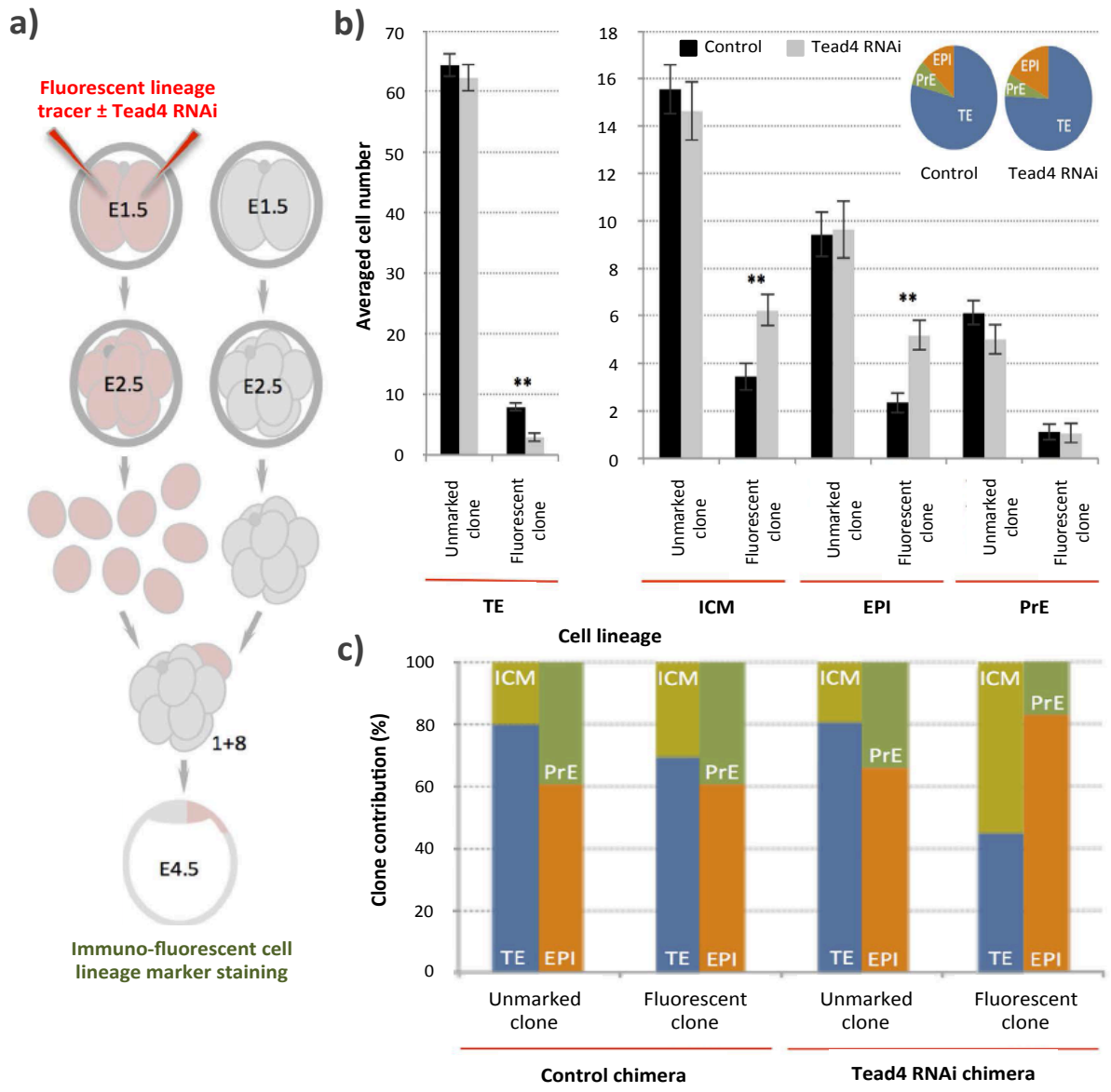

Figure 14: TE-inhibition biases resulting ICM-residing clones against forming PrE in favour of populating EPI: **a)** Experimental strategy to generate chimeras containing fluorescently marked TE-inhibited cell clones (plus fluorescence alone control clones) and to assay their contribution to late blastocyst stage cell lineages. **b)** Averaged number of chimera cells from within ('fluorescent clone') and outwith ('unmarked clone') the fluorescently marked TE-inhibited/Tead4 RNAi (grey bars, n=17) or marked control (black bars, n=30) cell clones in each of the late blastocyst lineages. Errors represent s.e.m. and $p<0.005$ significance markers (**) are shown (2-tailed student t-tests); inset, pie-charts denoting relative average blastocyst lineage contribution of cells within each embryo group. **c)** Averaged percentage contribution of unmarked and fluorescently labelled cell clones in control and TE-inhibited/Tead4 RNAi clone containing chimeras to stated cell lineages (yellow, ICM; blue, TE; green, PrE and orange, EPI); note reduced contribution of marked clones in TE-inhibited/Tead4 RNAi containing embryos to TE and PrE but increased contribution to ICM overall and EPI.

investigations of the ICM cell populations of early blastocyst containing control and TE-inhibited cell clones, confirmed severely reduced presence of Fgfr2 protein at the plasma membrane of TE-inhibited clones (and a heterogeneous pattern of membrane associated localisation in both the non-marked clones and either clone of control embryo groups; potentially indicative of the presence of the emerging EPI and PrE progenitor cell pools known to be present in the ICM at this time: *i.e.* the salt and pepper distribution[41]).

Collectively, these data provide compelling evidence in support of the integrated cell-fate model, and although extra research effort is required to fully understand its mechanistic drivers and its full relevance/significance during the highly regulative landscape of early mouse (and mammal) preimplantation embryo development, current opinion seems to be moving in its favour; as evidence by the publication of other broadly supportive studies[43,45].

3.9 A role for Rho-associated protein kinase in regulating preimplantation mouse embryo intra-cellular apical-basolateral polarity, differential inner- and outer-cell hippo-signalling pathway activation and cell-fate, by regulating Angiomotin sub-cellular localisation.

As referenced in the above introductory section (see 3.3), the establishment of intra-cellular apical-basolateral polarity in compacting late 8-cell (E2.5) stage embryos is an essential prerequisite to assigning functional cell-fate identity to the emerging and spatially distinct inner- and outer- cell populations of the embryo, after the 8- to 16- and 16- to 32-cell transitions[2,5,81]; by ensuring suppression or activation of the hippo-signalling pathway in polarised outer- and apolar inner-cells, respectively[19]. As early as 1999, the importance of small GTPases of the RhoA family in regulating compaction and, largely by inference, apical-basolateral polarity has been known; derived from experiments either inhibiting their activity with the *Clostridium botulinum* C3-transferase toxin or expression of microinjected constitutively activate recombinant proteins[82]; but it was not until 2014 that these initial observations were meaningfully developed by investigating the requirement of the RhoA effector kinases Rock1/2 (Rho-associated protein kinases 1/2) by pharmacologically inhibiting their activity from the 2-cell stage and subsequently assaying preimplantation embryo development[30,83]. Two resulting and contrasting publications have reported failure of Rock1/2 inhibited embryos to either form a blastocyst (associated with apical-basolateral polarity defects[30]) or developmental arrest at stages prior to the 8-cell stage (associated with actin polymerisation defects[83]). Therefore, we decided to independently assay the effects of Rock1/2 inhibition, using the same chemical compound employed in the two original studies (Y-27632), on preimplantation mouse embryo development to provide clarifying insight to the wider field; additionally, we also conducted more comprehensive analyses of the uncovered defective polarity phenotypes in relation to hippo-signalling pathway activity at various developmental time-points, as summarised below.

Thus, we found that Rock1/2 inhibition in embryos *in vitro* cultured from the 2-cell (E1.5) stage results in the failure to form a blastocyst and embryonic cell death during the equivalent developmental period that control treated blastocysts were expanding/maturing; *i.e.* agreeing with the findings of Kono *et al.*[30]. Moreover, we discovered such defective blastocyst formation was associated with a failure of outer cells to specify the TE lineage and as a consequence form a functioning epithelium (that should during unperturbed development comprise tight-junctions plus associated actin between neighbouring outer cells, but was lost after Rock1/2 inhibition) capable of supporting the formation of the fluid filled cavity. Moreover, by performing confocal based microscopy of embryos immuno-stained at the 16- (E3.0) and 32-cell E3.5) stages, we reported a complete breakdown of outer-cell apical basolateral polarity; typified by the homogenous distribution of both normally apical (*e.g.* aPKCλ/ζ, Pard6b and phospho-Ezr) and basolateral (*i.e.* Scrib)

protein markers around the entire cell cortex; although the generally cortical distribution of α-tubulin (*i.e.* microtubules) and basolateral localisations of E-cadherin (Cdh1; a cell-adhesion molecule) and the adherens junction marker β-catenin (Ctnnb1) were unaffected. We also observed that the hippo-signalling pathway activator Angiomotin (Amot), that normally functions to activate the Lats1/2 kinase effectors of the pathway at the adherens junctions of inner cells, was mis-localised to the lateral regions of outer-cells (possibly also at basal regions, but microscopy resolution could not discriminate from inner-cell derived signals); indicating possible aberrant activation of the hippo-signalling pathway in these cells. Consistently, we observed a shift from ordinarily outer-cell nuclear enriched Yap1 localisation (indicative of inactive hippo-signalling that potentiates the required TE-specific gene expression, by the Tead4-Yap1 transcriptional complex) to a predominantly nuclear excluded/cytoplasmic enriched Yap1 immuno-staining pattern, indicating the ectopic activation of the hippo-pathway (reminiscent of the situation normally observed in inner-cells that acts to antagonise TE-required gene expression and permits pluripotency-related gene expression, *e.g. Sox2*[17]) and the mechanistic reason why TE-cell fate specification is blocked in outer-cells in response to Rock1/2 inhibition (Fig. 15). To confirm the mis-localisation of Amot consequent to Rock1/2 inhibition is functionally downstream of the observed breakdown in intra-cellular apical-basolateral polarisation, and that it is responsible for the ectopic outer-cell activation of the hippo-signalling pathway, the following experiment was performed. Two-cell (E1.5) stage embryos were co-microinjected, in both blastomeres, with either Amot-specific dsRNA or negative control GFP-specific dsRNA (*i.e.* has no homology within the murine genome) and rhodamine-conjugated dextran fluorescent tracer beads (RDBs) and *in vitro* cultured until late morula (E3.5 –just prior to initiation of cavity formation) stage in the presence of Rock1/2 inhibitor or DMSO vehicle control; the subcellular localisation of pairwise combinations of Amot, Yap1 (as a read-out of hippo-pathway activation) and Pard6b (as a marker of apical-basolateral polarity) were then assayed by confocal immuno-fluorescent staining microscopy. As can be seen in Fig. 15, the inhibition of Rock1/2 activity in control GFP-dsRNA microinjected embryos resulted in the consistent mis-localisation of Amot to the lateral membranes of outer-cells that was also concomitant with a breakdown in apical-basolateral polarisation (*n.b.* Pard6b signal on the lateral membranes of outer-cells – yellow arrows); such defects were also associated with activated hippo-signalling, as evidenced by Yap1 exclusion from outer-cell nuclei (white arrows). However, in embryos confirmed to be depleted of Amot protein by Amot-dsRNA microinjection, Rock1/2 inhibition was no longer able to activate ectopic hippo-signalling in outer-cells, resulting in nuclear accumulation of Yap1 (white arrow heads; also seen in inner cells, indicating inhibition of the pathway in these cells too – white asterisks); this was despite the presence of defective outer-cell apical-basolateral polarisation persisting (blue arrows; see Pard6b protein localised to outer-cell lateral membranes in Rock1/2 inhibited embryos microinjected with the Amot-specific dsRNA). Hence, it was concluded that the cell-fate related phenotype of activated hippo-signalling observed in outer cells after Rock1/2 inhibition is exclusively mediated by mis-localisation of Amot to adherens junction enriched basolateral surfaces; moreover that is functionally downstream of apical-basolateral polarisation (also confirmed by other groups[36]).

Collectively, these data have provided important mechanistic insight into how a class of RhoA-effector, typified by Rock1/2, is involved in regulating appropriate outer-TE and inner-ICM cell-fate, in the developing preimplantation mouse embryo, by regulating the

appropriate formation and maintenance of apical-basolateral polarity in outer-cell populations. Moreover, this work has opened a rich seam of potential research avenues, to be exploited by our group in the future.

3.10 p38 mitogen-activated-kinases (Mapk14/11) regulate entry into primitive endoderm differentiation in the mouse blastocyst ICM.

As introduced above (see 3.4), cells of the early mouse blastocyst (E3.5) stage ICM initially appear uncommitted to either of the segregated cell lineage fates observed by the late blastocyst (E4.5) stage; and thus co-express both Gata6 and Nanog protein, that ultimately represent distinct markers of the PrE and EPI, respectively. However, during blastocyst maturation this initially homogenous and overlapping pattern of lineage marker gene expression becomes resolved, so that by the mid-blastocyst (E3.75 – E4.0) stage, individual cells express Gata6 or Nanog in a mutually exclusive manner revealed as mosaic across the

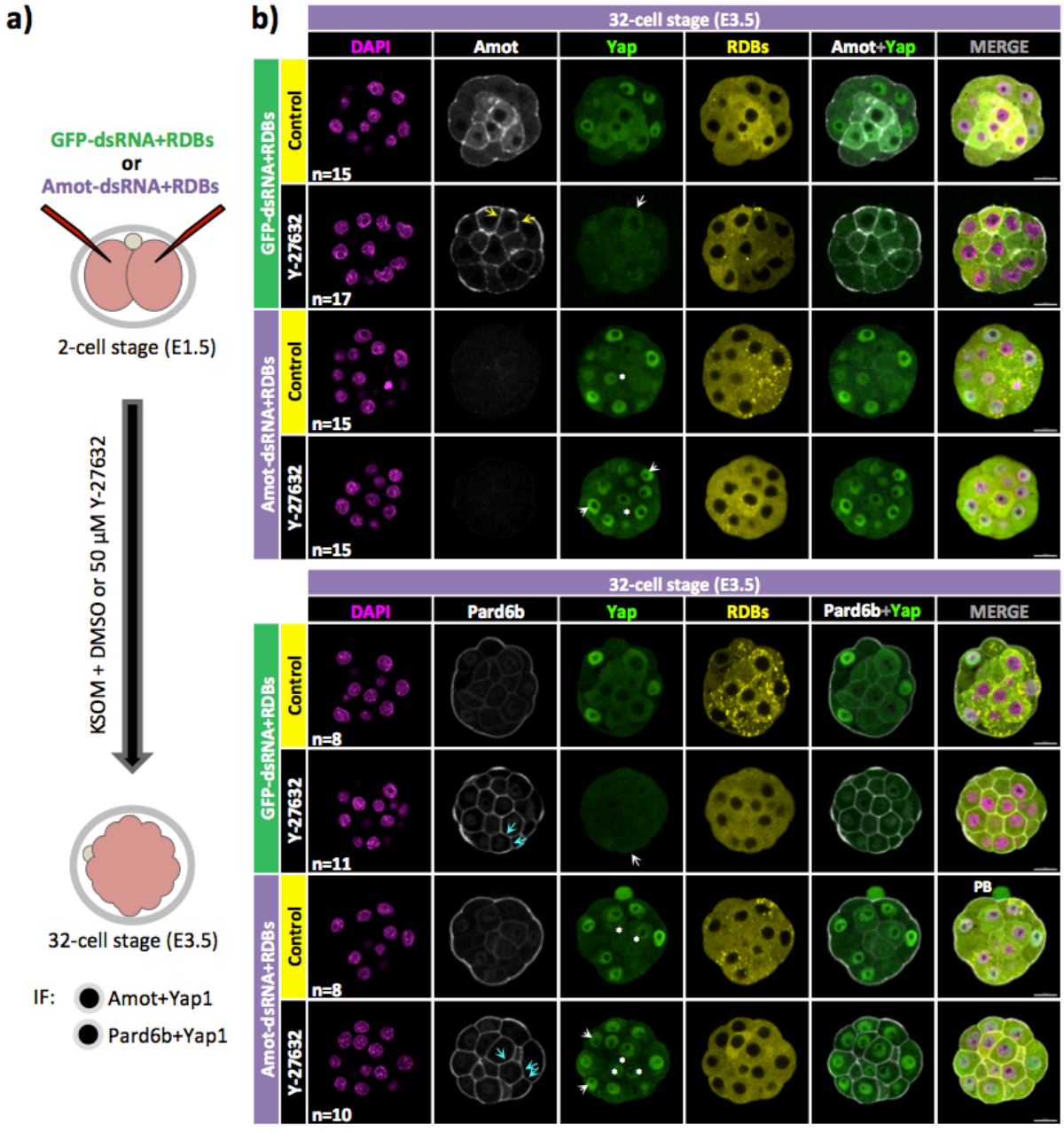

Figure 15: *legend overleaf*

Figure 15: Rock1/2 inhibition results in breakdown of apical-basolateral polarity, mis-localisation of Amot to outer-cell lateral membranes and ectopic outer-cell hippo-pathway activation, mediated by Amot. a) Schema of experimental strategy used; *i.e.* microinjection of Amot-specific or negative control GFP-specific dsRNA + RDBs (fluorescent tracer beads) and *in vitro* culture of embryos ± Rock1/2 inhibition to late morula stage prior to specific immuno-fluorescent (IF) staining (as indicated). **b)** Representative individual IF confocal z-stack micrographs of RNAi negative GFP-control or Amot-knock-down embryos ± Rock1/2 inhibition (experimental 'n' number provided), double IF-stained for the stated protein markers; Amot (assaying outer-cell apical sequestration, grey-scale), Yap1 (hippo-signalling pathway activation, green) and Pard6b (apical-basolateral polarity, grey-scale) – all micrographs are co-stained for DNA (DAPI, pseudo-coloured magenta) and RDBs (confirming microinjected construct delivery, pseudo-coloured yellow) with partial and full channel merges to aid interpretation, scale bar = 20μm and 'PB' denotes second meiotic polar body. Note, yellow arrows in Amot stains denote mis-localisation of Amot to outer-cell lateral membranes after Rock1/2 inhibition; blue arrows highlight disrupted apical-basolateral polarity indicated by outer cell basolateral localisation of Pard6b and ectopic expression on inner-cell membranes, after Rock1/2 inhibition; white arrows in Yap1 stains denote ectopic cytoplasmic Yap1 localisation (*i.e.* not nuclear enriched) in outer-cells, after Rock1/2 inhibition; white arrow-heads and white asterisks denote outer-cell and inner-cell nuclear accumulation of Yap1 irrespective of Rock1/2 inhibition status in Amot-knock-down embryos, respectively – indicating Rock1/2 inhibition is unable to cause ectopic outer cell activation of the hippo-signalling pathway, despite causing a breakdown in apical-basolateral polarity, in the absence of Amot – hence Rock1/2 exerts its effects on TE vs. ICM cell fate via Amot; by regulating its relative subcellular localisation in a polarity dependent mechanism.

ICM, termed the 'salt and pepper' expression pattern[41]. As the salt and pepper pattern emerges, cells within the Gata6 expressing ICM population begin to express later PrE markers (*e.g.* Sox17) and it is at this point the PrE lineage is specified from that of the EPI. As development progresses, later PrE transcription factor markers become activated (*e.g.* Gata4 and Sox7[47]) and the PrE and EPI specified cells actively sort into the cavity-facing and superficial monolayer or deeply encapsulated ICM populations of the late blastocyst, respectively. It is known that ICM cell-lineage specification is under the control active Fgf4-based signalling[43,51,52,84], whereby small inherent heterogeneities between individual ICM cells (the potential origins of which were discussed above – see 3.4) of the early blastocyst (E3.5) dictate if they will be more able respond to extra-cellular Fgf4 ligand and initiate gene expression programmes required for PrE differentiation (*i.e.* support continued *Gata6* expression, induce *Sox17* and repress *Nanog*); or whether they are less able to receive differentiative Fgf4-based inputs and thus retain pluripotency and moreover induce the expression of secreted Fgf4 ligand (*i.e.* ensuring continued expression of *Nanog* and other pluripotency factors, *e.g. Sox2*, to resist differentiation whilst concomitantly producing the extra-cellular differentiation signal required to provide the driving force to differentially specify the PrE and EPI lineages).

The central role played by the extracellular signal-regulated kinases 1/2 (Erk1/2) in mediating PrE differentiation in response to Fgf4-based signals has been known for some time; indeed genetic mutation of the *Grb2* gene (a adaptor protein known to facilitate the transfer of liganded membrane receptor tyrosine kinase signals through to Erk1/2 activation) is associated with a complete failure to specify PrE[41] and moreover pharmacological inhibition of the mitogen-activated kinase-kinases 1/2 (Mek1/2), responsible for phosphorylating and activating Erk1/2, also completely blocks PrE differentiation, resulting in a pan-ICM EPI-specified state[51,52], by the late blastocyst (E4.5) stage. However, we wanted to test for a functional and PrE-related role for the related class of genes in the serine-threonine and tyrosine kinase super-family, represented by the p38-mitogen activated kinases (p38-Mapks); *n.b.* there are four paralogous *p38-Mapk* genes in

the mouse genome (designated; *p38α/Mapk14*, *p38β/Mapk11*, *p38γ/Mapk12* and *p38δ/Mapk13* and all are known to be expressed in the preimplantation mouse embryo with varying dynamics and expression levels[85]). This was primarily because of a then recently published paper describing a role for p38-Mapk (Mapk14/11) in mediating autocrine Fgf2-Fgfr2 based signalling in the specification and derivation of the TE lineage[86], plus other ES cell based studies suggesting p38-Mapk inhibition promotes pluripotency. Therefore, we designed experiments to assay for a role of p38-Mapk in regulating PrE differentiation (given that PrE is a conceptually similar tissue to the TE; *i.e.* it is also an polarised epithelial monolayer of differentiating extra-embryonic cells, that evidence suggests is preferentially derived from differentiating outer- 16-cell stage TE ancestors cells, by being primed to express genes required for the PrE lineage[42-45]) using a similar pharmacological based approach employed by the group reporting a role in the TE[86].

Accordingly, we *in vitro* cultured embryos from the early- (E3.5) to late-blastocyst (E4.5) stage in the presence of either the p38-Mapk specific kinase inhibitor SB220025 (targeting the Mapk14/11 isoforms) or an equivalent volume of DMSO (as the negative vehicle control). Confocal microscopy based immuno-fluorescence analysis revealed a robust reduction in the number of ICM cells expressing the late-PrE marker, Gata4, that was also associated with an increase in the number of ICM cells solely expressing Nanog; indicative of impaired/blocked PrE differentiation. A result that was repeatable using the alternative p38-Mapk inhibitor, SB203580. Complementary analyses detecting Nanog protein expression in combination with the early-PrE marker Gata6, revealed the majority of ICM cells co-expressed both markers and also that the ICM had statistically significant fewer cells exclusively expressing one or other of the EPI or PrE markers; although the effect on the number of solely expressing Gata6/PrE-specified cells was more marked. Thus, these data indicate that most ICM cells in p38-Mapk inhibited blastocysts remain in their initial uncommitted cell-fate state and are unable to induce the successive expression of late PrE markers (*e.g.* Gata4) that is associated with appropriate PrE specification and maturation. Indeed, this was further substantiated by the observation of impaired formation of the salt and pepper pattern of mutually exclusive ICM lineage marker expression in p38-Mapk inhibited embryos, compared to the DMSO vehicle control treated group, when assayed at the mid-blastocyst (E4.0) stage (*i.e.* 80% of ICM cells still co-express Nanog and Sox17; incidentally indicating p38-Mapk activity is not required for initiation of *Sox17* expression, but that this is not maintained to the late blastocyst/E4.5 stage). These data therefore indicated a role for p38-Mapk in regulating the entry of initially uncommitted ICM founder cells into the full PrE differentiative programme, required to appropriately derive the matured blastocyst ICM cell lineages. We next assayed the developmental window of p38-Mapk inhibition sensitivity, directly comparing it against that for Mek1/2 inhibition, to ascertain at which point in blastocyst ICM maturation p38-Mapk activity is required. Such analyses revealed the application of p38-Mapk inhibition after E3.75 had no effect on matured PrE differentiation (assayed at E4.5 by detecting Gata4 protein), or EPI formation (assaying Nanog) and was in contrast to PrE-related Mek1/2 inhibition sensitivity, that persisted to a point after E4.0 (in agreement with previous studies demonstrating ICM cell-fate plasticity in relation to the activity of the Erk1/2 pathway[52]). Thus, demonstrating p38-Mapk activity is required during early blastocyst ICM maturation and consistently, prior to the emergence of the salt and pepper pattern of specific EPI and PrE progenitors; moreover, its activity is distinct from that of Mek1/2 that is required throughout blastocyst maturation.

These data are consistent with p38-Mapk (Mapk14/11) fulfilling an early role in blastocyst maturation that enables initially uncommitted cells to specify and differentiate towards the PrE lineage that is facilitated and driven to completion by on-going signalling through the activated Erk1/2 pathway. Moreover it suggests the requirement for p38-Mapk activation is, at least partially, functionally upstream of Erk1/2 activation as p38-Mapk inhibition efficiently blocks full PrE differentiation (as assayed by Gata4 expression).

As stated above, activation of the Erk1/2 pathway in the context of blastocyst ICM PrE differentiation has been confirmed to be subsequent to activated Fgfr-signalling[51,52]; moreover, exogenous addition of recombinant Fgf4 to maturing blastocysts is sufficient to drive all ICM cells to adopt a PrE cell-fate[51,52]. We therefore wanted to assay if the PrE-related role for p38-Mapk we had uncovered, was also associated with active Fgfr signalling. Accordingly, we devised an experimental strategy to pharmacologically inhibit Fgfr-signalling (using the compound SU5402 given from the 16-cell/E3.0 to late-blastocyst/E4.5 stages) to elicit characteristic inhibition of PrE differentiation (assaying expression of the late PrE marker, Gata4), against which we could test the ability of experimentally activating endogenous levels of p38-Mapk (by expressing a phospho-mimetic and constitutively active form of its specific and functionally upstream kinase, Mkk6; termed Mkk6-EE and confirmed to increase active phosph-p38-Mapk protein levels when micro-injected as a mRNA in the preimplantation mouse embryo) to rescue the PrE differentiation programme. As shown in Fig. 16, treating *in vitro* cultured embryos (that had been microinjected with control GFP mRNA) with SU5402 was associated with the anticipated and robust block in completed PrE differentiation (with ICM cells adopting an alternative Nanog positive, EPI cell-fate). However, the microinjection and expression of recombinant Mkk6-EE mRNA, was able to efficiently rescue the Fgfr inhibition induced PrE differentiation deficient phenotype; moreover this rescue could ablated by subsequent pharmacological inhibition of p38-Mapk activity during blastocyst maturation, demonstrating the dependence of the rescue on active p38-Mapk (Fig. 16). Thus, overall demonstrating a mechanism of p38-Mapk activation that is, like that of Erk1/2 activation, functionally downstream of engaged Fgfr-signalling (also confirmed by assaying and detecting decreased phospho-p38-Mapk levels in Fgfr-inhibited late morula stage embryos); moreover, we also demonstrate a similar requirement for functionally active Tak1 (a mitogen-activated kinase-kinase-kinase) implicated in contributing to PrE differentiation via a non-canonical and Smad-independent Bmp-signalling related mechanism[87] but that may also, or alternatively, be activated during blastocyst ICM maturation by active Fgfr-signalling, given precedent in other systems[88].

In conclusion, this study discovered a completely novel regulatory role for p38-Mapks (Mapk14/11) in regulating cell-fate derivation in the mouse blastocyst ICM; specifically by integrating the same Fgf-based extra-cellular signals already known to promote PrE differentiation via activation of the Erk1/2 pathway, but crucially during a developmental period preceding, or at least partially over-lapping, it during early blastocyst ICM maturation and prior to the emergence of the salt and pepper pattern of specified PrE and EPI cells. As such, the following model of ICM cell-fate derivation was proposed (summarised in Fig. 17); active Fgf-signalling inputs (plus possibly other inputs *e.g.* active Bmp-signalling) received by the essentially cell-fate uncommitted cells of the early blastocyst (E3.5) ICM, in turn activate p38-Mapk. Such activation of p38-Mapk permits/enables ICM cells receiving the extra-cellular inputs to begin the process of PrE specification that is completed by E3.75 and

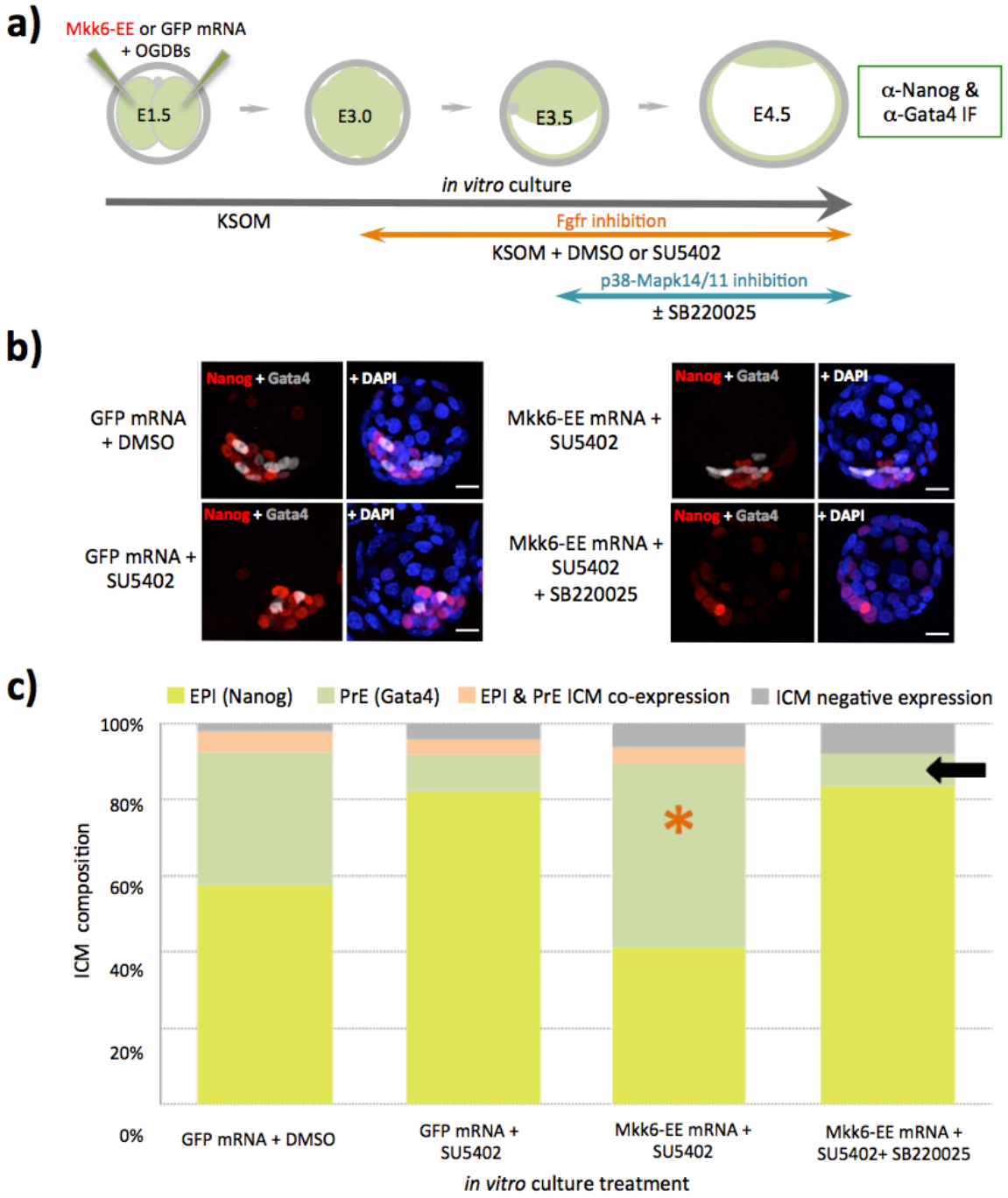

Figure 16: a) Experimental schema by which 2-cell stage embryos, that had been microinjected with either control GFP or constitutively active Mkk6 (Mkk6-EE) mRNA, were cultured to the 16-cell (E3.0) stage and then transferred into media containing the Fgfr-inhibitor SU5402 (or vehicle control DMSO) and then further cultured to the late blastocyst (E4.5) stage and assayed for expression of EPI (Nanog) or late PrE (Gata4) marker protein expression; *n.b.*, one group of Mkk6-EE injected embryos that were subject to Fgfr-inhibition were transferred into media that additionally contained the p38-Mapk inhibitor SB2200225 at the early-blastocyst (E3.5) stage and then cultured to the late-blastocyst (E4.5) stage. **b)** Representative projected confocal micrographs of immuno-fluorescently (IF) stained late-blastocyst (E4.5) embryos; Nanog (red), Gata4 (grey scale) and DNA (blue – DAPI), scale bar = 15μm. Individual experimental conditions: GFP mRNA + DMSO (n = 23), GFP mRNA + SU5402 (n = 23), Mkk6-EE mRNA + SU5402 (n = 23) & Mkk6-EE + SU5402 + SB220025 (n = 22). **c)** Averaged cell lineage composition of the ICM of the experimental groups assayed, assigned according to mutually exclusive, over-lapping or negative expression of EPI (Nanog) and late PrE (Gata4) marker proteins. Orange asterisk denotes rescued PrE differentiation in embryos expressing Mkk-6EE (thus activating endogenous levels of p38-Mapk) under Fgfr-inhibited conditions, compared to control group (GFP mRNA injected) and the black arrow highlights ablation of such rescue by subsequent p38-Mapk inhibition from the early blastocyst stage.

subsequently revealed in the mid-blastocyst (~E3.75 – E4.0) stage as the salt and pepper pattern. Continued, extra-cellular Fgf-based stimulation then drives PrE differentiation (and the associated expression of later PrE marker genes) to its completion, by the late blastocyst (E4.5) stage (revealed after active cell sorting in the separated EPI and PrE tissue layers). Additionally, one further consequence of this study, has been to provide our research group with extensive future research leads, currently under investigation.

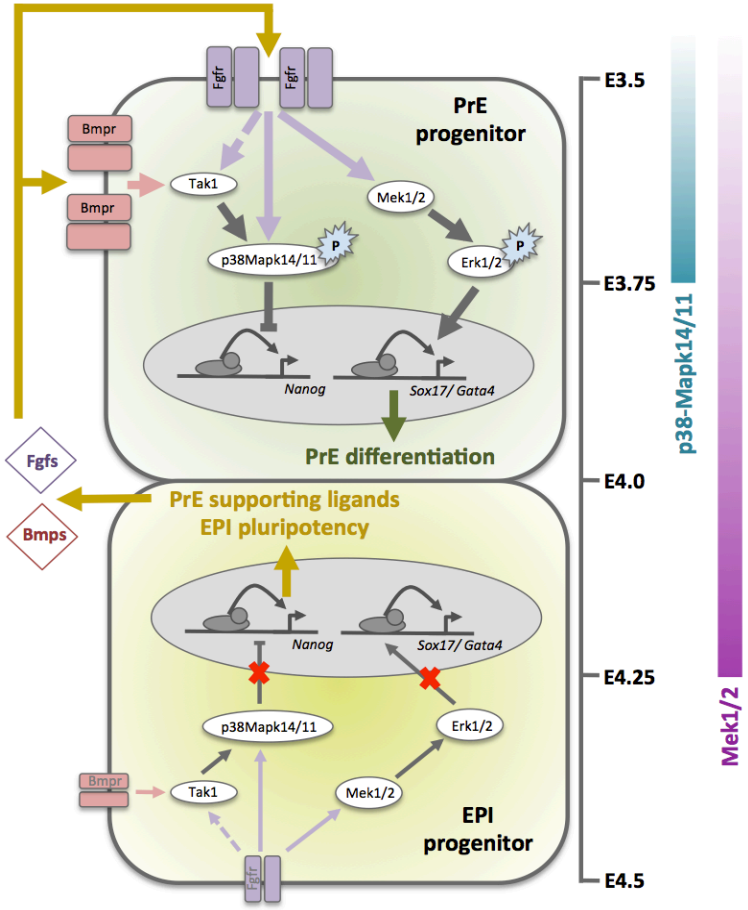

Figure 17: A contemporary model of mouse blastocyst ICM cell-fate derivation, whereby initially uncommitted ICM cells that will ultimately give rise to PrE, receive Fgf- (plus potentially Bmp-) based differentiation signals, during early blastocyst maturation (before E3.75) that cause activation of p38-Mapks. Such activation permits the specification of the PrE cell-fate, by down-regulating Nanog expression (as revealed by the salt and pepper pattern of EPI and early PrE markers by the mid-blastocyst/ ~E3.75 - E4.0) stage, that is then driven to completion, as the blastocyst matures, by Fgf-induced Erk1/2 pathway activation. Conversely, a comparative inability to respond to the early blastocyst signalling events (Fgf- & Bmp-based) in EPI progenitors, leads to the continued expression of pluripotency required genes (antagonising PrE differentiation) and the production of the secreted PrE promoting ligands (Fgfs & Bmps) that drive PrE specification and differentiation, thus resulting in the separation of the two blastocyst ICM lineages (EPI & PrE). Taken from Thamodaran and Bruce (2016)[50].

3.11 A summary of theme II research.

The work described in theme II relates to the point at which one began researching mechanisms of cell-fate derivation in the preimplantation mouse embryo, initially as a senior post-doctoral research associate under the mentorship of Prof. Magdalena Zernicka-Goetz at the Gurdon Institute (University of Cambridge, United Kingdom) but then as an independent principle investigator running ones own research group at the Faculty of Science, University of South Bohemia in České Budějovice (Czech Republic). During ones post-doctoral appointment, work developing observations initially made in the embryo and relating to inter-blastomere heterogeneity in the levels of R17/26H3me2a modified chromatin, catalysed by Carm1, and pluripotent potential[1] were developed in the mouse ES cell system; detailing how Carm1 is required to potentiate the expression of the key pluripotency transcription factors Oct4 and *Sox2*[71]. Additionally, a role for maternally provided Cdx2 mRNA was uncovered, using microinjected Cdx2-specific RNAi and morpholino construct based strategies, in assisting blastomeres of

the 8-cell stage embryo to appropriately compact and polarise along their apical-basolateral axis; our experimental interventions resulting in abrogated separation of TE and ICM cell-fate and arrested development and blastomere/embryo death, starting from the 8- to 16-cell transition, and blocked blastocyst formation[68]. A result that was also supported by a follow-up study in which both maternally and zygotically derived Cdx2 was removed using a genetic deletion strategy[74] (although some disagreement exists in the field[75]). Ones time working in the Zernicka-Goetz group was also defined by the publication of two review articles[2,3], summarising the contemporary state of the field and articulating novel perspectives and hypotheses (*e.g.* relating to the 'integrated cell-fate model, that was further developed in ones own independent research[44]). After moving to the University of South Bohemia, and successfully attracting funding in the form of a 'Marie Curie Career Integration Fellowship' and a Czech Science Foundation supported 'Standard Individual Grant', a paper functionally testing the 'integrated cell-fate model' and presenting evidence for its functional existence in the derivation of the three mouse blastocyst cell lineages was indeed published[44]. This was followed by a study demonstrating the importance of the RhoA effector kinases Rock1/2 on regulating the establishment and maintenance of (outer-) blastomere intra-cellular apical-basolateral polarity and the consequence for outer-TE and inner-ICM cell-fate derivation by regulating the differential activation of the hippo-signalling pathway; solely via the sub-cellular localisation of the hippo-pathway terminal kinase effector, Lats1/2, activator, Amot[39]. The last cited study relates to the identification of a novel role for p38-Mapk in regulating appropriate cell lineage derivation within the ICM of maturing mouse blastocysts; specifically by receiving extra-cellular Fgf-signalling inputs (plus potentially Bmp-related signals) to enable a subset of, essentially cell-fate uncommitted, receptive cells of the early blastocyst ICM to enter a programme of differentiation and thus specify the emerging PrE lineage; that is then driven to completion by the action of the related Erk1/2 pathway in response, at least partially, to the same Fgf-based signals. Hence, leading to the emergence of the pluripotent EPI (that did not respond to Fgf-signalling nor presumably activate p38-Mapks) and differentiating PrE lineages, evident by the late blastocyst (E4.5) cell stage[50]. Data arising from these latter three publications have provided a rich seam of research leads that the ones group is currently developing and that is expected to facilitate our research efforts into the foreseeable future.

3.12 References relevant to theme II.

1 Torres-Padilla, M. E., Parfitt, D. E., Kouzarides, T. & Zernicka-Goetz, M. Histone arginine methylation regulates pluripotency in the early mouse embryo. *Nature* **445**, 214-218 (2007).

2 Zernicka-Goetz, M., Morris, S. A. & Bruce, A. W. Making a firm decision: multifaceted regulation of cell fate in the early mouse embryo. *Nature reviews* **10**, 467-477 (2009).

3 Bruce, A. W. & Zernicka-Goetz, M. Developmental control of the early mammalian embryo: competition among heterogeneous cells that biases cell fate. *Curr Opin Genet Dev* **20**, 485-491 (2010).

4 Lokken, A. A. & Ralston, A. The Genetic Regulation of Cell Fate During Preimplantation Mouse Development. *Curr Top Dev Biol* **120**, 173-202 (2016).

5 Chazaud, C. & Yamanaka, Y. Lineage specification in the mouse preimplantation embryo. *Development* **143**, 1063-1074 (2016).

6 Tarkowski, A. K. Mouse chimaeras developed from fused eggs. *Nature* **190**, 857-860 (1961).

7 Tarkowski, A. K. & Wroblewska, J. Development of blastomeres of mouse eggs isolated at the 4- and 8-cell stage. *Journal of embryology and experimental morphology* **18**, 155-180 (1967).

8 Johnson, M. H. & Ziomek, C. A. The foundation of two distinct cell lineages within the mouse morula. *Cell* **24**, 71-80 (1981).

9 Samarage, C. R. *et al.* Cortical Tension Allocates the First Inner Cells of the Mammalian Embryo. *Developmental cell* **34**, 435-447 (2015).

10 Anani, S., Bhat, S., Honma-Yamanaka, N., Krawchuk, D. & Yamanaka, Y. Initiation of Hippo signaling is linked to polarity rather than to cell position in the pre-implantation mouse embryo. *Development* **141**, 2813-2824 (2014).

11 Maitre, J. L. *et al.* Asymmetric division of contractile domains couples cell positioning and fate specification. *Nature* **536**, 344-348 (2016).

12 Korotkevich, E. *et al.* The Apical Domain Is Required and Sufficient for the First Lineage Segregation in the Mouse Embryo. *Developmental cell* **40**, 235-247 e237 (2017).

13 Nishioka, N. *et al.* Tead4 is required for specification of trophectoderm in pre-implantation mouse embryos. *Mechanisms of development* **125**, 270-283 (2008).

14 Strumpf, D. *et al.* Cdx2 is required for correct cell fate specification and differentiation of trophectoderm in the mouse blastocyst. *Development* **132**, 2093-2102 (2005).

15 Ralston, A. *et al.* Gata3 regulates trophoblast development downstream of Tead4 and in parallel to Cdx2. *Development* **137**, 395-403 (2010).

16 Cao, Z. *et al.* Transcription factor AP-2gamma induces early Cdx2 expression and represses HIPPO signaling to specify the trophectoderm lineage. *Development* **142**, 1606-1615 (2015).

17 Wicklow, E. *et al.* HIPPO Pathway Members Restrict SOX2 to the Inner Cell Mass Where It Promotes ICM Fates in the Mouse Blastocyst. *PLoS Genet* **10**, e1004618 (2014).

18 Loh, Y. H. *et al.* The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nat Genet* **38**, 431-440 (2006).

19 Sasaki, H. Roles and regulations of Hippo signaling during preimplantation mouse development. *Dev Growth Differ* **59**, 12-20 (2017).

20 Bergstralh, D. T. & St Johnston, D. Epithelial cell polarity: what flies can teach us about cancer. *Essays Biochem* **53**, 129-140 (2012).

21 De Vries, W. N. *et al.* Maternal beta-catenin and E-cadherin in mouse development. *Development* **131**, 4435-4445 (2004).

22 Dard, N., Louvet-Vallee, S., Santa-Maria, A. & Maro, B. Phosphorylation of ezrin on threonine T567 plays a crucial role during compaction in the mouse early embryo. *Developmental biology* **271**, 87-97 (2004).

23 Louvet-Vallee, S. ERM proteins: from cellular architecture to cell signaling. *Biol Cell* **92**, 305-316 (2000).

24 Ajduk, A. & Zernicka-Goetz, M. Polarity and cell division orientation in the cleavage embryo: from worm to human. *Mol Hum Reprod*, (2015).

25 Sasaki, H. Position- and polarity-dependent Hippo signaling regulates cell fates in preimplantation mouse embryos. *Semin Cell Dev Biol*, (2015).

26 Fleming, T. P., McConnell, J., Johnson, M. H. & Stevenson, B. R. Development of tight junctions de novo in the mouse early embryo: control of assembly of the tight junction-specific protein, ZO-1. *J Cell Biol* **108**, 1407-1418 (1989).

27 Plusa, B. *et al.* Downregulation of Par3 and aPKC function directs cells towards the ICM in the preimplantation mouse embryo. *Journal of cell science* **118**, 505-515 (2005).

28 Vinot, S., Le, T., Maro, B. & Louvet-Vallee, S. Two PAR6 proteins become asymmetrically localized during establishment of polarity in mouse oocytes. *Curr Biol* **14**, 520-525 (2004).

29 Vinot, S. *et al.* Asymmetric distribution of PAR proteins in the mouse embryo begins at the 8-cell stage during compaction. *Developmental biology* **282**, 307-319 (2005).

30 Kono, K., Tamashiro, D. A. & Alarcon, V. B. Inhibition of RHO-ROCK signaling enhances ICM and suppresses TE characteristics through activation of Hippo signaling in the mouse blastocyst. *Developmental biology* **394**, 142-155 (2014).

31 Alarcon, V. B. Cell polarity regulator PARD6B is essential for trophectoderm formation in the preimplantation mouse embryo. *Biol Reprod* **83**, 347-358 (2010).

32 Dard, N., Le, T., Maro, B. & Louvet-Vallee, S. Inactivation of aPKClambda reveals a context dependent allocation of cell lineages in preimplantation mouse embryos. *PLoS One* **4**, e7117 (2009).

33 Meng, Z., Moroishi, T. & Guan, K. L. Mechanisms of Hippo pathway regulation. *Genes & development* **30**, 1-17 (2016).

34 Nishioka, N. *et al.* The Hippo signaling pathway components Lats and Yap pattern Tead4 activity to distinguish mouse trophectoderm from inner cell mass. *Developmental cell* **16**, 398-410 (2009).

35 Hirate, Y. *et al.* Par-aPKC-dependent and -independent mechanisms cooperatively control cell polarity, Hippo signaling, and cell positioning in 16-cell stage mouse embryos. *Dev Growth Differ* **57**, 544-556 (2015).

36 Hirate, Y. *et al.* Polarity-dependent distribution of angiomotin localizes Hippo signaling in preimplantation embryos. *Curr Biol* **23**, 1181-1194 (2013).

37 Leung, C. Y. & Zernicka-Goetz, M. Angiomotin prevents pluripotent lineage differentiation in mouse embryos via Hippo pathway-dependent and -independent mechanisms. *Nat Commun* **4**, 2251 (2013).

38 Hirate, Y. & Sasaki, H. The role of angiomotin phosphorylation in the Hippo pathway during preimplantation mouse development. *Tissue Barriers* **2**, e28127 (2014).

39 Mihajlovic, A. I. & Bruce, A. W. Rho-associated protein kinase regulates subcellular localisation of Angiomotin and Hippo-signalling during preimplantation mouse embryo development. *Reprod Biomed Online* **33**, 381-390 (2016).

40 Kang, M., Piliszek, A., Artus, J. & Hadjantonakis, A. K. FGF4 is required for lineage restriction and salt-and-pepper distribution of primitive endoderm factors but not their initial expression in the mouse. *Development* **140**, 267-279 (2013).

41 Chazaud, C., Yamanaka, Y., Pawson, T. & Rossant, J. Early lineage segregation between epiblast and primitive endoderm in mouse blastocysts through the Grb2-MAPK pathway. *Developmental cell* **10**, 615-624 (2006).

42 Morris, S. A. *et al.* Origin and formation of the first two distinct cell types of the inner cell mass in the mouse embryo. *Proc Natl Acad Sci U S A* **107**, 6364-6369 (2010).

43 Morris, S. A., Graham, S. J., Jedrusik, A. & Zernicka-Goetz, M. The differential response to Fgf signalling in cells internalized at different times influences lineage segregation in preimplantation mouse embryos. *Open Biol* **3**, 130104 (2013).

44 Mihajlovic, A. I., Thamodaran, V. & Bruce, A. W. The first two cell-fate decisions of preimplantation mouse embryo development are not functionally independent. *Sci Rep* **5**, 15034 (2015).

45 Krupa, M. *et al.* Allocation of inner cells to epiblast vs primitive endoderm in the mouse embryo is biased but not determined by the round of asymmetric divisions (8-->16- and 16-->32-cells). *Developmental biology* **385**, 136-148 (2014).

46 Ohnishi, Y. *et al.* Cell-to-cell expression variability followed by signal reinforcement progressively segregates early mouse lineages. *Nat Cell Biol* **16**, 27-37 (2014).

47 Artus, J., Piliszek, A. & Hadjantonakis, A. K. The primitive endoderm lineage of the mouse blastocyst: sequential transcription factor activation and regulation of differentiation by Sox17. *Developmental biology* **350**, 393-404 (2011).

48 Aksoy, I. *et al.* Oct4 switches partnering from Sox2 to Sox17 to reinterpret the enhancer code and specify endoderm. *EMBO J* **32**, 938-953 (2013).

49 Frankenberg, S. *et al.* Primitive endoderm differentiates via a three-step mechanism involving Nanog and RTK signaling. *Developmental cell* **21**, 1005-1013 (2011).

50 Thamodaran, V. & Bruce, A. W. p38 (Mapk14/11) occupies a regulatory node governing entry into primitive endoderm differentiation during preimplantation mouse embryo development. *Open Biol* **6**, rsob.160190 (2016).

51 Nichols, J., Silva, J., Roode, M. & Smith, A. Suppression of Erk signalling promotes ground state pluripotency in the mouse embryo. *Development* **136**, 3215-3222 (2009).

52 Yamanaka, Y., Lanner, F. & Rossant, J. FGF signal-dependent segregation of primitive endoderm and epiblast in the mouse blastocyst. *Development* **137**, 715-724 (2010).

53 Wu, Q. *et al.* CARM1 is required in embryonic stem cells to maintain pluripotency and resist differentiation. *Stem cells* **27**, 2637-2645 (2009).

54 Mitsui, K. *et al.* The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell* **113**, 631-642 (2003).

55 Chambers, I. *et al.* Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell* **113**, 643-655 (2003).

56 Niwa, H. *et al.* Interaction between Oct3/4 and Cdx2 determines trophectoderm differentiation. *Cell* **123**, 917-929 (2005).

57 Avilion, A. A. *et al.* Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes & development* **17**, 126-140 (2003).

58 Masui, S. *et al.* Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. *Nat Cell Biol* **9**, 625-635 (2007).

59 Goolam, M. *et al.* Heterogeneity in Oct4 and Sox2 Targets Biases Cell Fate in 4-Cell Mouse Embryos. *Cell* **165**, 61-74 (2016).

60 Panamarova, M. *et al.* The BAF chromatin remodelling complex is an epigenetic regulator of lineage specification in the early mouse embryo. *Development* **143**, 1271-1283 (2016).

61 Parfitt, D. E. & Zernicka-Goetz, M. Epigenetic modification affecting expression of cell polarity and cell fate genes to regulate lineage specification in the early mouse embryo. *Mol Biol Cell* **21**, 2649-2660 (2010).

62 Beck, F., Erler, T., Russell, A. & James, R. Expression of Cdx-2 in the mouse embryo and placenta: possible role in patterning of the extra-embryonic membranes. *Dev Dyn* **204**, 219-227(1995).
63 Chawengsaksophak, K., James, R., Hammond, V. E., Kontgen, F. & Beck, F. Homeosis and intestinal tumours in Cdx2 mutant mice. *Nature* **386**, 84-87 (1997).
64 Tamai, Y. *et al.* Colonic hamartoma development by anomalous duplication in Cdx2 knockout mice. *Cancer Res* **59**, 2965-2970 (1999).
65 Dietrich, J. E. & Hiiragi, T. Stochastic patterning in the mouse pre-implantation embryo. *Development* **134**, 4219-4231 (2007).
66 Ralston, A. & Rossant, J. Cdx2 acts downstream of cell polarization to cell-autonomously promote trophectoderm fate in the early mouse embryo. *Developmental biology* **313**, 614-629 (2008).
67 Jedrusik, A. *et al.* Role of Cdx2 and cell polarity in cell allocation and specification of trophectoderm and inner cell mass in the mouse embryo. *Genes & development* **22**, 2692-2706 (2008).
68 Jedrusik, A. *et al.* Maternally and zygotically provided Cdx2 have novel and critical roles for early development of the mouse embryo. *Developmental biology* **344**, 66-78 (2010).
69 Wang, Q. T. *et al.* A genome-wide study of gene activity reveals developmental signaling pathways in the preimplantation mouse embryo. *Developmental cell* **6**, 133-144 (2004).
70 Svoboda, P., Franke, V. & Schultz, R. M. Sculpting the Transcriptome During the Oocyte-to-Embryo Transition in Mouse. *Curr Top Dev Biol* **113**, 305-349 (2015).
71 Wu, G. *et al.* Initiation of trophectoderm lineage specification in mouse embryos is independent of Cdx2. *Development* **137**, 4159-4169 (2010).
72 Bruce, A. W. What is the role of maternally provided Cdx2 mRNA in early mouse embryogenesis? *Reprod Biomed Online* **22**, 512-515 (2011).
73 Wu, G. & Scholer, H. R. Role of mouse maternal Cdx2: what's the debate all about? *Reprod Biomed Online* **22**, 516-518; discussion 519-520 (2011).
74 Jedrusik, A., Cox, A., Wicher, K. B., Glover, D. M. & Zernicka-Goetz, M. Maternal-zygotic knockout reveals a critical role of Cdx2 in the morula to blastocyst transition. *Developmental biology* **398**, 147-152 (2015).
75 Blij, S., Frum, T., Akyol, A., Fearon, E. & Ralston, A. Maternal Cdx2 is dispensable for mouse development. *Development* **139**, 3969-3972 (2012).
76 Gao, N., White, P. & Kaestner, K. H. Establishment of intestinal identity and epithelial-mesenchymal signaling by Cdx2. *Developmental cell* **16**, 588-599 (2009).
77 Consortium, E. P. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74 (2012).
78 Birney, E. *et al.* Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**, 799-816 (2007).
79 Morris, S. A. Cell fate in the early mouse embryo: sorting out the influence of developmental history on lineage choice. *Reprod Biomed Online* **22**, 521-524 (2011).
80 Bruce, A. W. Generating different genetic expression patterns in the early embryo: insights from the mouse model. *Reprod Biomed Online* **27**, 586-592 (2013).
81 Johnson, M. H. From mouse egg to mouse embryo: polarities, axes, and tissues. *Annual review of cell and developmental biology* **25**, 483-512 (2009).

82 Clayton, L., Hall, A. & Johnson, M. H. A role for Rho-like GTPases in the polarisation of mouse eight-cell blastomeres. *Developmental biology* **205**, 322-331 (1999).

83 Duan, X. *et al.* ROCK inhibition prevents early mouse embryo development. *Histochem Cell Biol* **142**, 227-233 (2014).

84 Morris, S. A., Guo, Y. & Zernicka-Goetz, M. Developmental plasticity is bound by pluripotency and the Fgf and Wnt signaling pathways. *Cell Rep* **2**, 756-765 (2012).

85 Natale, D. R., Paliga, A. J., Beier, F., D'Souza, S. J. & Watson, A. J. p38 MAPK signaling during murine preimplantation development. *Developmental biology* **268**, 76-88 (2004).

86 Yang, J. *et al.* Binding of FGF2 to FGFR2 in an autocrine mode in trophectoderm cells is indispensable for mouse blastocyst formation through PKC-p38 pathway. *Cell Cycle* **14**, 3318-3330 (2015).

87 Graham, S. J. *et al.* BMP signalling regulates the pre-implantation development of extra-embryonic cell lineages in the mouse embryo. *Nat Commun* **5**, 5667 (2014).

88 Salazar, L. *et al.* Fibroblast growth factor receptor 3 interacts with and activates TGFbeta-activated kinase 1 tyrosine phosphorylation and NFkappaB signaling in multiple myeloma and bladder cancer. *PLoS One* **9**, e86470 (2014).

CONCLUDING REMARKS

Contained within the pages of this habilitation thesis is a summary of ones scientific research; conducted in the period following the award of ones Ph.D. degree and encompassing ones time as a post-doctoral researcher in the United Kingdom, before culminating in ones transition to an independent research group leader here in the Faculty of Science at the University of South Bohemia in České Budějovice (Czech Republic). As stated, this research has been performed in two phases, each represented by one of the two detailed scientific themes (with particular emphasis placed on ones contemporary research efforts, described in theme II):

- ***Theme I*: Regulation of specific gene transcription & chromatin structure.**
- ***Theme II*: Pluripotency & the acquisition of cell-fate in the preimplantation mouse embryo.**

It is hopefully anticipated the descriptive summaries of ones research output contained within this thesis, together with the attendant provision of the additional details of ones professional career and activities (including ones *curriculum vitae* and summaries relating to ones, publication history, teaching schedule and performance, research student supervision, conference attendance and data presentation history; plus the abstracts pertaining to ones proposed pedagogical and research thesis defence lectures) will be deemed satisfactory for the award of ones habilitation at the Faculty of Science at the University of South Bohemia in České Budějovice.

SUMMARIES OF ARTICLES REFERED TO IN THE THESIS

4.1 THEME I: REGULATION OF SPECIFIC GENE TRANSCRIPTION & CHROMATIN STRUCTURE.

4.1.1 Primary research articles:

1. *** Follows GA, Dhami P, Göttgens B, Bruce AW, Campbell PJ, Dillon SC, Smith AM, Koch C, Donaldson IJ, Scott MA, Dunham I, Janes ME, Vetrie D, Green AR. Identifying gene regulatory elements by genomic microarray mapping of DNaseI hypersensitive sites. *Genome Res.* 16, 1310-1319 (2006).** This study describes the development of a novel technique to characterise protein-genome interactions, typical of those found at DNA regulatory elements within *in vivo* intact nuclear genomes, using a modified version of classic DNaseI foot-printing. Briefly, genomic DNA is isolated from isolated and DNaseI treated cell nuclei, blunt-ended and ligated to an asymmetric double-stranded oligonucleotide linker. A second biotinylated oligonucleotide primer is then used to derive primer extension products between 200-450 base-pairs in length, that are then extracted, fluorescently labelled and hybridised to genomic DNA microarrays; thus hybridising probe signals on the microarray reflect the location of DNAseI hypersensitive sites in the intact nuclei. This approach was validated using a genomic microarray comprising a tile path of the human and mouse *SCL (Scl)/TAL1 (Tal1)* gene loci (and human beta-globin locus) and identified all known regulatory elements plus previously uncharacterised regulator elements.

** Highlighted articles are referred to in the main body of the thesis but do not have a dedicated section related to them.*

2. **Bruce AW, Krejčí A, Ooi L, Deuchars J, Wood IC, Doležal V and Buckley NJ. The transcriptional repressor REST is a critical regulator of the neurosecretory phenotype. *J. Neurochem.* 98, 1828-1840 (2006).** This paper represents a study emerging from ones Ph.D. studies in which DNA sequence elements bound by the transcriptional repressor protein REST/NRSF were bioinformatically identified in the newly available human and mouse genome sequence. It was noted that such elements (RE1/NRSEs) were statistically enriched at gene loci with functions in the neurosecretory pathway. This paper details functional evidence that REST does indeed regulate the transcription of these genes and as such, the functioning of the neurosecretory pathway; as confirmed by functional assays of regulated secretion under conditions of dysregulated *REST* expression.

3. *** Johnson RJ, Gamblin RJ, Ooi L, Bruce AW, Donaldson IJ, Westhead DR, Wood IC, Jackson RM, Buckley NJ. Identification of the REST regulon reveals extensive transposable element-mediated binding site duplication. *Nucleic Acids Res*. 34, 3862-3877 (2006).** This paper describes follow-on and refining work from another paper (published as part of my Ph.D. thesis studies – Bruce et al., 2004 – *Proc. Natl. Acad. Sci. U. S. A.*) that had been aimed at bioinformatically identifying all DNA sequence binding elements for the master neuronal transcriptional repressor factor, REST/NRSF (that are atypically long; known as RE1/NRSE), in newly available genome sequence. The paper represented an improved approach to identifying such sites and their associated, and thus likely target genes, and the results were collated into an online relational database resource. The main novel finding was the existence of RE1/NRSEs associated with transposable elements, providing an evolutionary explanation as how RE1/NRSEs duplicated and brought additional genes into the REST/NRSF regulon.

4. **Koch CM, Andrews RM, Flicek P, Dillon SC, Karaoz U, Clelland GK, Wilcox S, Beare DM, Fowler JC, Couttet P, James KD, Lefebvre GC, Bruce AW, Dovey OM, Ellis PD, Dhami P, Langford CF, Weng Z, Birney E, Carter NP, Vetrie D and Dunham I. The landscape of histone modifications across 1% of the human genome in five human cell lines. *Genome Res.* 17, 691-707 (2007).** This paper describes in comprehensive detail the findings of the Wellcome Trust Sanger Institute's contribution to the international ENCODE pilot project (see above). Specifically, the then as yet unknown relationships between specific post-translational epigenetic histone modifications and underlying genetic features, such as transcriptional start sites and exons/introns in both transcriptionally active and inactive gene loci.

5. **The ENCODE Project Consortium (*inc. Bruce AW*). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature.* 447, 799-816 (2007).** This paper outlines the preliminary findings of the pilot phase of the human genome ENCODE (Encyclopaedia of DNA elements) international consortium project, ultimately aimed at identifying all DNA regulatory elements within the human genome. As a pilot phase project it constituted a feasibility study in which 1% of the human genome was interrogated in multiple human derived cell samples. The Wellcome Trust Sanger Institute was a member of this consortium and designed and manufactured a tile-path genomic DNA microarray, constituting 44 individual and

** Highlighted articles are referred to in the main body of the thesis but do not have a dedicated section related to them.*

contiguous genomic regions, and used it to interrogate the distribution of distinct post-translational epigenetic chromatin modifications and selected transcription factor binding, using the chromatin immuno-precipitation coupled to microarray hybridisation (ChIP:Chip) technique. This contribution is reflected in this paper and more comprehensively described in Koch *et al.,* 2007 (see above). The project was the primer for the report of the full ENCODE project's findings (ENCODE Project Consortium 2012 – *Nature*) and defined a path for countless further more in-depth studies of the human genome.

6. **Bruce AW, Lopez-Contreras AJ, Flicek P, Down T, Dhami P, Dillon SC, Koch CM, Langford, CF, Dunham I, Andrews RM and Vetrie D. Functional diversity for REST (NRSF) is defined by *in vivo* binding affinity hierarchies at the DNA sequence level. *Genome Res.* 19, 994-1005 (2009).** This paper was amongst the first to demonstrate the existence of functionally relevant *in vivo* binding affinity hierarchies for any given transcription factor and distinct subsets of its chromatin/DNA binding sites, based on sequence variations in its DNA sequence binding motif, in the human genome. The study used the master neuronal repressor transcription factor REST as its model (due to its uncommonly long DNA sequence binding motif). The study also showed how the existence of such hierarchies functionally contribute to its lineage specific roles, by demonstrating association with the highest affinity and most evolutionarily conserved sites of genes commonly regulated by REST amongst multiple human cell types but tissue specific association with weaker and less well conserved sites.

7. **Dhami P, Bruce AW, Jim JH, Dillon SC, Hall A, Cooper JL, Bonhoure N, Chiang K, Ellis PD, Langford CF, Andrews RM and Vetrie D. Genomic approaches uncover increasing complexities in the regulatory landscape at the human *SCL* (*TAL1*) locus. *PLoS One.* 5(2), e9059 (2010).** The described data relate to chromatin immuno-precipitation coupled with genomic microarray hybridisation (so-called ChIP:Chip) experiments, interrogating the chromatin landscape surrounding the key haematopoiesis transcription factor regulator gene *SCL*/*TAL1* (an oncogene frequently dysregulated in multiple lymphoid leukaemias) in various haematopoietic cell lines (both positive and negative for *SCL* gene expression). The analyses led to the identification of six novel *SCL* regulatory elements and further in-depth characterisation of already classified enhancers.

8. **Dhami P[§], Saffrey P[§], Bruce AW[§], Dillon SC[§], Chiang K, Bonhoure N, Koch CM, Bye J, James K, Foad NS, Ellis PD, Watkins NA, Ouwehand WH, Langford CF, Andrews RM, Dunham I and Vetrie D. Complex exon-intron marking by histone modifications is not determined solely by nucleosome distribution. *PLoS One.* 5(8), e12339 (2010). [§]*equal contribution.*** This paper describes chromatin immuno-precipitation coupled with genomic microarray hybridisation (so-called ChIP:Chip) experiments in three human cell types. It confirms the presence of gene exon-intron marking associated with a comprehensive panel of epigenetic post-translational histone modifications, but to hitherto unrecognised levels of complexity (*e.g.* demonstrating novel marks enriched in introns, rather than previously characterise exon marks, and the fact the marking system is laid down in the absence of underlying gene transcription, although later transcription can lead to dynamic changes). Most importantly, the data demonstrate

that the observed marking system is not solely accounted for by nucleosome density and likely provides a link between overall chromatin accessibility, RNA polymerase II processivity and co-transcriptional splicing.

4.2 THEME II: PLURIPOTENCY & THE ACQUISITION OF CELL FATE IN THE PREIMPLANTATION MOUSE EMBRYO.

4.2.1 Primary research articles:

1. **Wu Q[§], Bruce AW[§], Jedrusik A, Ellis PD, Andrews RM, Langford CF, Glover DM and Zernicka-Goetz M. CARM1 is required in ES cells to maintain pluripotency and resist differentiation. *Stem Cells.* 27, 2637-2645 (2009) [§]*equal contribution.*** The paper demonstrates that the arginine-methyl-transferase Carm1 is required in mouse embryonic stem (ES) cells to promote self-renewal and pluripotency (consistent with findings in the early preimplantation mouse embryo demonstrating inter-blastomere asymmetry of Carm1 deposited histone H3 methyl argine-17 and -26 substrate levels that positively correlate with enhanced pluripotency). Moreover, via both gain and loss of function analyses in mouse ES cells, the paper also shows Carm1 regulates the expression of the pluripotency related triad of transcription factor genes (*Oct4/Pou5f1*, *Sox2* and *Nanog*) and that over-expression of Carm1 is sufficient to delay ES cells in their response to differentiation cues.

2. **Jedrusik A, Bruce AW[§], Tan MH[§], Leong DE, Skamagki M, Yao M and Zernicka-Goetz M. Maternally and zygotically provided Cdx2 have novel and critical roles for early development of the mouse embryo. *Dev. Biol.* 344, 66-78 (2010). [§]*equal contribution.*** The data described in this paper show that the depletion of both maternal and zygotic mRNA, using RNAi approaches (and the blocking of transcript translation with specific morpholino constructs), for the trophectoderm specific transcription factor gene *Cdx2,* leads to earlier and more severe developmental phenotypes (around the 16-cell morula stage) associated with apical-basolateral polarity defects (that could be rescued by overexpression of Cdx2 mRNA), than observed in zygotic alone genetic null mutants. These data inform a novel and distinct role for detected maternal Cdx2 mRNA constructs in preimplantation mouse development.

3. **Mihajlović AI, Thamodaran V and Bruce AW. The first two cell-fate decisions of preimplantation mouse embryo development are not functionally independent. *Sci. Rep.* 13(5): 15034 (2015).** The data in this paper describe experiments in which varyingly sized cell clones, in which trophectoderm differentiation had been inhibited by RNAi-mediated knockdown of the Tead4 transcription factor, were generated within the developing preimplantation mouse embryo and the ability of internalised cells, from this clone, to contribute to the second extraembryonic primitive endoderm lineage was assayed. It was found that such cells were significantly compromised in this regard, demonstrating a functional relationship between the initiation of trophectoderm differentiation and the subsequent differentiation of internalised progeny cells towards the primitive endoderm, via a potential priming mechanism, likely involving receptiveness to fibroblast growth factor signalling; supportive of the so-called 'integrated' cell fate model that postulates the earliest possible

segregation/generation of inner-cell-mass cells in the developing embryo is associated with retention of pluripotency.

4. **Mihajlović AI and Bruce AW. Rho-associated protein kinase regulates subcellular localisation of Angiomotin and Hippo-signalling during preimplantation mouse embryo development. *Reprod. Biomed. Online.* 33(3), 381-390 (2016).** This paper was amongst those to describe a functional cell-fate role for the Rho-associated kinases Rock1/2 in the acquisition of trophectoderm versus inner cell mass identity; bring resolution to the field following conflicting reports in the literature. Moreover, it functionally confirmed, for the first time, the hippo-signalling activator Amot as the mediator the reported Rock1/2 role, in addition to reporting an extended analysis of identified Rock1/2 inhibition phenotypes at developmental stages prior to the morula to blastocyst transition.

5. **Thamodaran V and Bruce AW. p38 (Mapk14/11) occupies a regulatory node governing entry into primitive endoderm differentiation during preimplantation mouse embryo development. *Open Biology.* 6(9) /rsob.160190 (2016).** This paper was the first to describe the pivotal role of p38-mitogen-activated-kinases in permitting/enabling primitive endoderm differentiation within the mouse blastocyst inner-cell-mass. A role distinct from the previously reported primitive endoderm promoting properties of the related Erk1/2 kinases, shown here to drive the differentiation, but nevertheless also under the control of fibroblast growth factor-based signalling.

4.2.2 Review articles:

1. *** Zernicka-Goetz M, Morris SA and Bruce AW. Making a firm decision: Transcriptional regulation of cell fate in preimplantation development. *Nat. Rev. Genet.* 10, 467-477 (2009).** This is a comprehensive and highly cited (139 as of 05.04.2017) review dedicated to summarising the then current understanding of how the three mouse blastocyst embryo cell lineages (epiblast, trophectoderm and primitive endoderm) arise; divided into two cell-fate decisions (outer trophectoderm vs. inner cell mass and the segregation of epiblast vs. primitive endoderm within the inner cell mass, respectively). The review is set against reconciling the data against the three classically proposed models of cell lineage formation; *i.e.* the early asymmetry, inside-outside and polarity models.

2. *** Bruce AW and Zernicka-Goetz M. Developmental control of the early mammalian embryo: competition among heterogeneous cells that biases cell fate. *Curr. Opin. Genet. & Dev*. 20, 485-491 (2010).** A review article that discusses the evidence for molecular inter-cell heterogeneities between the blastomeres of the early preimplantation stage embryo and how they may contribute to the emergence of the three blastocyst lineages via mechanisms involving competition between such cells for particular niches and hence lineages. The article discusses these themes against published findings suggesting the emergence of the two ICM lineages, epiblast and primitive endoderm, may be informed by the relative developmental timing of ICM founder cell internalisation within the preimplantation stage embryo.

** Highlighted articles are referred to in the main body of the thesis but do not have a dedicated section related to them.*

3. * **Bruce AW. Generating different genetic expression patterns in the early embryo; insights from the mouse model. *Reprod. Biomed. Online.* 27(6), 586-592 (2013).** A review article addressing what was known about how the necessary transcription factor gene expression patterns required to confer identity to the three emerging blastocyst lineages, whilst also proposing a model similar to that expounded in Mihajlović et al., (2015). The article was also given as a seminar at the *Futures in Reproduction Meeting; Celebrating the award of the 2010 Nobel Prize for Physiology or Medicine to Professor Sir Robert Edwards* at Churchill College, University of Cambridge, Cambridge, United Kingdom.

4.2.3 Commentaries:

1. * **Bruce AW. What is the role of maternally provided *Cdx2* mRNA in early mouse embryogenesis? *Reprod. Biomed. Online.* 22(6), 512-515 (2011)a.** A commentary article addressing the functional role of maternally provided Cdx2 mRNA in the preimplantation mouse embryo, following on from the paper one published with co-workers in *Developmental Biology* (Jedrusik *et al.,* 2010 – see above) arguing an important role in the initiation of the first cell-fate decision (trophectoderm versus inner cell mass) and written in the light of somewhat contrasting results of others (Wu *et al.,* 2010 – *Development*) demonstrating more mild phenotypes associated with knockdown of maternal and zygotic Cdx2 mRNA).

2. * **Bruce AW. Response: Role of mouse maternal Cdx2: what's the debate all about? *Reprod. Biomed. Online.* 22(6), 519-5120 (2011)b.** A response to the letter published by Wu and Schöler in *Reproductive BioMedicine Online* (2011), that itself was a response to an earlier commentary article of my own (Bruce 2011, *Reproductive BioMedicine Online* – see above), trying to resolve differences in data and interpretation, undertaken by ourselves and themselves regarding the role of maternally provided *Cdx2* mRNA.

** Highlighted articles are referred to in the main body of the thesis but do not have a dedicated section related to them.*

ATTACHED PUBLICATIONS

The following is an attachment/appendix of print publications in relation to the presented thesis themes:

- Theme I: Regulation of specific gene transcription & chromatin structure.
- Theme II: Pluripotency & the acquisition of cell-fate in the preimplantation mouse embryo.