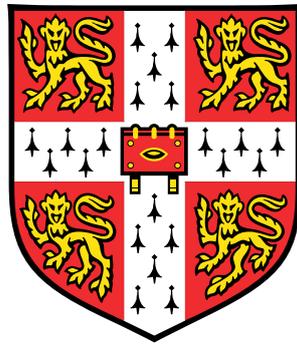


Defining the Transcriptional and Epigenetic Signature of Mouse Embryonic Stem Cells With Compromised Developmental Potency



Maria Anna Schacker

Department of Physiology, Development and Neuroscience
University of Cambridge

This dissertation is submitted for the degree of
Doctor of Philosophy

Lucy Cavendish College

August 2018

Meinen Eltern

*The good thing about science is that it's true
whether or not you believe in it.*
- Neil deGrasse Tyson

Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text. It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. It does not exceed the prescribed word limit for the Degree Committee of Biology.

Maria Anna Schacker
August 2018

Acknowledgements

This PhD would not have been possible without the help and scientific and emotional support of my colleagues, friends and family.

First and foremost my thanks and gratitude go to my supervisors Prof. Bill Colledge and Dr. Mike Snaith for giving me the chance to work with and learn from them. Bill has been an amazing mentor throughout my time in Cambridge. He has guided me through my PhD with his immense support, patience and knowledge. At Medimmune, Mike has given me the chance to experience scientific research in an industry setting and I am very grateful for that. I would also like to thank the BBSRC and Medimmune for funding my research and for their generous studentship.

None of this would have been possible without my great colleagues in the Colledge lab and at Medimmune. Victoria and Shel have been there for me from the start of my PhD until now, first as colleagues and now as friends. I am so thankful for all the support I have gotten from them, especially during the first year of my PhD. Although they were not directly involved with my project, Antonio and Stephen have been important members of our team and significantly contributed to the success of my PhD. Being my mentor at Medimmune, Yi-Han has played an integral part in the success of this project. She has guided me through my PhD with her enthusiasm, patience and of course her extensive scientific knowledge. Thanks also go to Steve for his help with the phage display project and Melanie for helping with the methylation analysis.

Furthermore, I would like to thank everyone in PDN for making my time so much more enjoyable. Special thanks must go to Katerina for being such an amazing friend (and of course for trying to help me overcome my fear of mice) and to Jorge without whom weekends in the lab would have been a lot more lonely and much less fun.

My time in Cambridge definitely would not have been the same without the Lucy Cavendish College Boat Club. It has not only given me the opportunity to take part in the great Cambridge tradition that is rowing, but more importantly I have made so many amazing friends. There are too many to mention but you know who you are! I am so proud of everything that we have accomplished together.

All of this would not have been possible without the support of my family. I want to use this

opportunity to thank my amazing parents for everything that they have done for me and for their never ending love and support throughout not just these past four years but throughout my entire life.

Last, but by all means not least, I would like to thank Lukas who has supported me along the way of this journey (almost) from day one and who has always believed in me. It meant the world to me knowing that I could always count on you - thank you for everything.

Abstract

Mouse embryonic stem (ES) cells have played a crucial role in studying developmental processes and gene function *in vivo*. They are extremely useful in the generation of transgenic animals as they can be genetically manipulated and subsequently microinjected into blastocyst stage embryos, where they combine with the inner cell mass and contribute to the developing embryo. Some of the resulting pups are chimaeric, consisting of a mixture of cells derived from the host blastocyst and the injected ES cells. We have identified several ES cell clones arising from gene targeting experiments with an impaired capacity to generate viable chimaeras. When injected into blastocysts, these clones cause embryonic death during mid to late gestation, suggesting that the cells are able to contribute to the embryo but interfere with normal embryonic development. The aim of this work was to identify the underlying changes in the transcriptome, epigenome or cell surface markers that have occurred in these compromised ES cells and to further define the developmental phenotype of the chimaeric embryos. Different stages during development were analysed and whereas there was little difference in embryonic death at gestational day e13.5, there was a significant decrease in embryos surviving to gestational day e17.5. Additionally, severe haemorrhaging was observed in all the dead embryos and small foci of haemorrhaging could also be seen in a number of embryos that were still alive. This was also observed at e13.5, albeit to a less severe extent. Using RNA sequencing to discover differences in the transcriptome between control ES cells and the compromised ES cells, five genes were identified that were downregulated in the compromised cells. Four of these, *Gtl2*, *Rtl1as*, *Rian* and *Mirg* are all located in the imprinted *Dlk1-Dio3* region on chromosome 12 and are normally expressed from the maternal genome. This pattern was also validated in tissues from e17.5 chimaeric embryos. The expression of this locus is to a large extent regulated by a differentially methylated region located approximately 13kb upstream of the *Gtl2* promoter, the IG-DMR. Whereas this is usually only methylated on the paternal copy, in the compromised ES cells both the paternal and the maternal copy were fully methylated, likely causing the silencing of *Gtl2*, *Rtl1as*, *Rian* and *Mirg*. Using the DNA methyltransferase inhibitor 5-azacytidine, expression of *Gtl2* could be rescued. Injection of those 5-azacytidine treated cells into blastocysts did partially rescue the embryonic lethal phenotype. Additionally, cell surface markers were analysed in a phenotypic screen using phage display. NGS analysis of the phage outputs indicates that there

may be additional differences in cell surface markers between the control and compromised ES cell clones, but their specific details remain to be identified.

Overall, we have identified the maternally expressed genes of the *Dlk1-Dio3* region as markers that can distinguish between ES cells with normal or compromised developmental potency and propose to include these genes in the pre-blastocyst injection screening routine for experiments involving the production of chimaeras or genetically modified mouse strains.

Table of contents

List of figures	xix
List of tables	xxv
List of Abbreviations	xxxii
1 Introduction	1
1.1 A Brief History of Mouse Embryonic Stem Cells	1
1.2 The Relationship Between ES Cells and the Early Embryo	2
1.3 Pluripotency Is Maintained by a Transcription Factor Network	3
1.3.1 OCT4	4
1.3.2 SOX2	4
1.3.3 NANOG	4
1.3.4 Other Factors	5
1.4 Maintaining ES Cells in Culture	6
1.4.1 Serum/LIF Media	6
1.4.2 2i Media	7
1.5 Epigenetic Regulation of Pluripotency and Differentiation	7
1.5.1 DNA Methylation	8
1.5.2 Histone Modifications	8
1.5.3 Imprinted Gene Expression in Embryo Development	9
1.6 Applications of Mouse ES Cells	11
1.6.1 Studying Developmental Processes and Modelling Diseases	11
1.6.2 Therapeutic Applications	12
1.7 Making Chimaeras Using Targeted ES Cells	12
1.7.1 Gene Targeting	12
1.7.2 Blastocyst Injections	13
1.8 Some ES Cell Lines Do Not Contribute to Viable Chimaeras	14

1.9	Hypothesis, Aims and Scope of This Thesis	16
2	Methods	17
2.1	Cell Culture	17
2.1.1	Cell Culture Media	17
2.1.2	Cell Culture Substrates	19
2.1.3	Other Cell Culture Reagents	19
2.1.4	Mouse Embryonic Fibroblast (MEF) Cell Culture and Inactivation	20
2.1.5	ES Cell Culture	20
2.1.6	Karyotyping	20
2.1.7	Differentiation Assays	20
2.1.8	Transfection Methods	22
2.2	Molecular Biology Techniques	23
2.2.1	gDNA Extraction	23
2.2.2	RNA Extraction	23
2.2.3	cDNA Synthesis	24
2.2.4	PCR	24
2.2.5	qPCR	25
2.2.6	Gel Electrophoresis	28
2.2.7	Plasmid Preparation	28
2.2.8	RNA Sequencing	28
2.2.9	DNA Methylation Analysis	29
2.2.10	Phenotypic Screening	31
2.3	Animal Work and Histology	36
2.3.1	Blastocyst Injections	36
2.3.2	Analysis of Chimaeras	37
2.3.3	Paraffin Embedding and Sectioning	37
2.3.4	Haematoxylin and Eosin Staining of Embryo Sections	37
3	Initial Characterisation of the ES Cell Clones	39
3.1	Introduction	39
3.1.1	Effects of Prolonged Culture on Karyotype and Pluripotency	39
3.2	Results	40
3.2.1	Karyotyping	40
3.2.2	Expression of Pluripotency Markers	41
3.2.3	Differentiation Assay	41
3.3	Discussion	44

3.3.1	Karyotype Analysis	44
3.3.2	Pluripotency and Differentiation Potential	45
3.3.3	Chapter Conclusion	46
4	Blastocyst Injections and Histology	47
4.1	Introduction	47
4.2	Results	47
4.2.1	Cell Labelling Using the PiggyBac System	47
4.2.2	Phenotypes of e13.5 and e17.5 Chimaeras	50
4.2.3	Histological Analysis	54
4.2.4	Correlation Between Contribution and Phenotype	56
4.3	Discussion	57
4.3.1	Cell Labelling	57
4.3.2	Phenotypic Differences Between ESC(Ctrl) and ESC(Comp) Cells	57
4.3.3	Correlation Between ES Cell Contribution and Phenotype	59
4.3.4	Chapter Conclusion	59
5	Transcriptomics	61
5.1	Introduction	61
5.2	Results	61
5.2.1	RNAseq Data Processing and Analysis	61
5.2.2	Validation of Expression Levels of Genes in the <i>Dlk1-Dio3</i> Region	67
5.2.3	Additional Compromised Clones Have a Similar Expression Profile at the <i>Dlk1-Dio3</i> Region	67
5.2.4	<i>Gtl2</i> Expression in the Embryos	68
5.3	Discussion	69
5.3.1	Structure and Control of the <i>Dlk1-Dio3</i> Region	70
5.3.2	Literature Review - Previous Publications Confirm Our Results	71
5.3.3	Silencing of Maternally Expressed Genes From The <i>Dlk1-Dio3</i> Region in ESC(Ctrl)4	73
5.3.4	Impact of Other Genes	73
5.3.5	<i>Gtl2</i> Expression in e17.5 Tissues	74
5.3.6	Chapter Conclusion	75
6	Epigenetics	77
6.1	Introduction	77
6.2	Results	77

6.2.1	Bisulfite Sequencing of the <i>Dlk1-Dio3</i> IG-DMR	77
6.2.2	Bisulfite Sequencing and RNAseq Results of Other Imprinted Regions	80
6.2.3	Rescuing Expression of the Maternally Expressed Genes of the <i>Dlk1-Dio3</i> Region	81
6.3	Discussion	85
6.3.1	Bisulfite Sequencing of the IG-DMR	85
6.3.2	Loss of Imprinting at Additional Imprinted Loci	86
6.3.3	<i>Gtl2</i> Expression Rescue	87
6.3.4	Chapter Conclusion	89
7	Phenotypic Screening by Phage Display	91
7.1	Introduction	91
7.2	Introduction to Phage Display	91
7.2.1	Filamentous Bacteriophages	91
7.2.2	Antibody Structure	93
7.2.3	Antibody Fragments for Modern Applications	94
7.2.4	Construction of scFv Phage Libraries	94
7.2.5	Biopanning	95
7.2.6	Applications of Phage Display	95
7.3	Phage Display and Why It is Useful in This Project	96
7.4	Experimental Setup	97
7.5	Results	98
7.5.1	Three Rounds of Cell Surface Selection and Selection Rescue	98
7.5.2	Next Generation Sequencing of Selection Outputs	99
7.5.3	scFv Reformatting to IgG Format	103
7.5.4	High Throughput IgG Expression	105
7.6	Discussion	105
7.6.1	NGS of Selection Outputs	106
7.6.2	Next Steps and Future Work	107
7.6.3	Chapter Conclusion	107
8	Conclusion and Future Work	109
8.1	Overview of Findings	109
8.2	General Discussion and Future Work - Relationship Between Phenotype and Genotype	110
8.2.1	Possible Role of Angiogenesis via VEGF Signalling	110

8.2.2	Possible Effect of <i>Gtl2</i> Silencing on Vasculature via DLK1-Notch signalling	113
8.2.3	Possible Role of a Placental Phenotype	113
8.2.4	Possible Role of a Musculoskeletal Phenotype	114
8.3	Concluding remarks	114
References		117

List of figures

1.1	Development of the early embryo (A) Schematic of the morphological changes that occur in the early embryo from the zygote to the blastocyst stage. ES cells can be isolated from the ICM or Epi compartment of blastocysts. ICM: inner cell mass, TE: trophoctoderm, Epi: epiblast, PrE: primitive endoderm. (B) Development and lineage segregation from the zygote into the three main germ layers.	3
1.2	Signalling pathways in the maintenance of pluripotency during ES cell culture BMP signalling through SMAD4/6 and the RAS/RAF/MEK/ERK pathway affect cell differentiation. LIF signalling both inhibits differentiation and activates pluripotency genes. The canonical WNT signalling pathway is involved in maintenance of pluripotency. Shown in red are the trophic factors that are added to ES media: LIF, the MEK1/2 inhibitor PD0325901 and the GSK3 inhibitor CHIR00921.	5
1.3	Waves of DNA demethylation A first wave of demethylation occurs in the primordial germ cells (PGCs). Parent-specific methylation patterns are imprinted during germ cell development. Shortly after fertilisation, a second wave of demethylation occurs and a new methylation signature is established around implantation. Imprinted genes with parent-specific methylation patterns are spared from the second round of demethylation (dotted lines). Red line is the female genome, blue line is the male genome.	9

1.4	Making chimaeras using targeted ES cells (A) Gene knockout by homologous recombination. The targeting vector contains the disrupted target gene with the selection marker <i>NeoR</i> . The targeted gene is not functional anymore. (B) Gene knockin by homologous recombination. The targeting vector contains the intact target gene as well as the cDNA for an additional gene (GeneX). This results in the expression of both the target gene and GeneX. (C) Targeted ES cells can be injected into blastocysts. They will combine with the ICM and contribute to chimaeras.	14
1.5	ES cell pedigree Pedigree of how the ES cell lines were derived by sequential gene targeting.	15
3.1	Chromosome counting (A) Number of chromosomes per cell for the four ESC(Comp) clones (n=15). Percentages indicate the proportion of cells with exactly 40 chromosomes.(B) Image of a typical chromosome spread of a single cell.	40
3.2	Expression of pluripotency markers (A) <i>Nanog</i> (B) <i>Oct4</i> (C) <i>Sox2</i> (D) <i>Klf4</i> (E) <i>Esrrb</i> (F) <i>Stat3</i> . Gene expression was normalised to <i>Hprt</i> and results were normalised to the values of ESC(Ctrl)1. Error bars represent standard error of the mean. Stars represent significant differences as calculated by ANOVA (p<0.05). n=2 (biological replicates).	42
3.3	Differentiation assay (A) Experimental set up for the differentiation assay (B) Representative images of differentiated cells from each cell line. α SMA in red stains cardiomyocytes, SOX17 in green stains definitive endoderm cells and β III-TUBULIN in green stains neuronal cells.	43
4.1	PiggyBac vectors (A) pPBCAG-H2BtdTomato-IH vector containing the <i>tdTomato</i> transgene, as well as the hygromycin resistance gene. (B) pPBCAG-Venus-IP vector containing the <i>Venus</i> transgene as well as the puromycin resistance gene. (C) pCyL43 PBase (not to scale) (D-E) Standard curves of <i>Kiss1</i> and <i>Venus</i> to calculate transgene copy number	48
4.2	Fluorescently labelled cells and embryos (A) Images of all eight cell lines after labelling with <i>Venus</i> or <i>tdTomato</i> . (B) Images of e13.5 chimaeras for each of the cell lines.	51

- 4.3 **Survival phenotype (A-B)** Phenotypes of chimaeric embryos at e13.5 (A) and e17.5 (B). Numbers above bars are the percentages of live (normal) embryos. (C-F) Percentages of live embryos compared between ESC(Ctrl) and ESC(Comp) groups for both time points: including all live embryos (C), all live embryos but excluding ESC(Ctrl)4 samples (D), only healthy/normal alive embryos (E), only healthy/normal live embryos and excluding ESC(Ctrl)4 samples (F). Stars represent significant difference as calculated by unpaired t-test with Welch's correction ($p < 0.05$). (G) Range of phenotypes: live (normal) e13.5 (i), live (normal) e17.5 (ii), live (haemorrhaging) e17.5 (iii), live (haemorrhaging) e13.5 (iv), dead (not fully resorbed) e17.5 (v, vi), dead (resorbed) e13.5 (vii), dead (resorbed) e17.5 (viii, ix). White arrow heads indicate small foci of haemorrhaging. Other phenotypes include edema (v) and lack of posterior development (viii). (H) Embryo crown-rump length in cm. 53
- 4.5 **Measuring fluorescence intensity by mean grey value (A)** Fluorescent images of e17.5 embryos were analysed in ImageJ. Their outline was drawn (yellow line) and the mean grey value of the defined area was measured. (B) Mean grey values for ESC(Comp) chimaeric embryos. ns means no significant difference and stars represent significant differences as calculated by ANOVA ($p < 0.05$). 56
- 4.4 **Histological analysis at e17.5 (A)** H&E images of sections of the limbs. i) ESC(Ctrl), ii-iii) ESC(Comp). Arrow heads showing blood vessels. iv) hemorrhage in ESC(Comp). v-vi) ESC(Ctrl), vii-viii) ESC(Comp). Arrows showing nucleated blood cells. Scale bar in iv is 250um, all others are 50um. (B-C) Box plots showing diameter of blood vessels in the limbs (B) and liver (C). Whiskers range from 10th to 90th percentile. The line inside the box represents the median, the mean is shown by a cross. Stars represent significant differences as calculated by the Mann-Whitney U test ($p < 0.05$). 56
- 5.1 **Q score distribution** Total Q score distribution for all reads of all samples. 62
- 5.2 **FastQC outputs** Representative example of (A-B) Per base sequence quality and (C-D) Per base sequence content pre (A,C) and post (B,D) trimming. 63
- 5.3 **RNAseq results (A)** PCA plot and (B) MA-plot comparing gene expression in ESC(Comp) samples vs ESC(Ctrl) samples. (C) Pedigree of ES cell clone derivation. 64
- 5.4 **RNAseq results without ESC(Ctrl)4 (A)** PCA plot without ESC(Ctrl)4. (B) MA-plot comparing all ESC(Comp) samples vs ESC(Ctrl)1-3. 65

- 5.5 **Validation of RNAseq results** (A) Maternally and (B) paternally expressed genes. Error bars represent standard error of the mean. Stars represent significant differences as calculated by ANOVA ($p < 0.05$). (C) Results of the PCR for *Rtl1as* (top lane) and *Rtl1* (bottom lane) after primer specific reverse transcription. 66
- 5.6 **Expression of *Gtl2*, *Rian* and *Mirg* in additional clones** Error bars represent standard error of the mean. Stars represent significant differences as calculated by ANOVA ($p < 0.05$). 68
- 5.7 ***Gtl2* tissue expression** (A) *Gtl2* expression in limbs from e17.5 chimeras. Each dot represents a tissue sample. Horizontal bars represent the means for each group of samples. (B) Genotyping results for the three ESC(Comp)4 chimaeras and representative chimaeras generated from other ESC(Ctrl) and ESC(Comp) clones. Top row is the PCR result for a housekeeping control, bottom row is the PCR results for primers specific to each ES cell clone. 69
- 5.8 ***Dlk1-Dio3* locus** *Dlk1-Dio3* imprinted region on the distal part of chromosome 12. The paternal allele encodes the protein coding genes *Dlk1*, *Rtl1* and *Dio3*. The maternal allele encodes the non-coding RNAs *Gtl2*, *Rtl1as*, *Rian* and *Mirg*. *Rian* encodes a number of C/D box snoRNAs. The region also contains two miRNA clusters: one within *Rtl1as* and the other one within *Mirg*. There are two DMRS: the IG-DMR and the *Gtl2*-DMR which are methylated on the paternal but unmethylated on the maternal chromosome. Expression of maternally expressed genes is mainly controlled by the IG-DMR. 70
- 6.1 **Bisulfite sequencing results** (A) IG-DMR of the *Dlk1-Dio3* region (B) Control regions (C) NESPAS-Gnasxl (D) H13 (E) Mest (F) Nap115 (G) H19 (H) Commd1 (I) Peg13 (J) Igf2r. Red outlines indicate aberrant methylation levels. The dotted line indicates 50% methylation. 79
- 6.2 **ES cell culture in different media conditions** *Gtl2* expression in five ES cell clones after culture in standard Serum/LIF conditions as well as 2i and EPSCM. Error bars represent standard error of the mean. Stars represent significant differences as calculated by 2-way-ANOVA ($p < 0.05$). 81
- 6.3 **Treatment with 5-azacytidine** (A) Viability assay for 5-azacytidine concentrations ranging from 0-10 μ M and for 24 h (light grey line) or 48 h (dark grey line). (B-D) *Gtl2*, *Rian* and *Mirg* expression after 48 h of 0.5 μ M 5-azacytidine treatment. Error bars represent standard error of the mean. Stars represent significant differences as calculated by 2-way-ANOVA ($p < 0.05$). 82

- 6.4 **Chimaeric offspring after 5-aza injections** (A) One female (i) and three males (ii-iv) born after injection of 5-aza treated, unselected cells. Photographs were taken at 1 month of age (B) Two very low percentage chimaeras born after injection of a 5-aza treated subclone (ESC(Comp)3-AZA28). Photographs were taken at 2 weeks of age. White arrows show differences in coat colour. 83
- 6.5 **Analysis of 5-azacytidine treated subclones of ESC(Comp)3** (A) *Gtl2* expression in 20 subclones of ESC(Comp)3 after 48h of 0.5uM 5-azacytidine treatment, compared to ESC(Ctrl)3 and the parental untreated ESC(Comp)3 cell line. Error bars represent standard error of the mean. Crosses and stars represent significant differences to ESC(Ctrl)3 and ESC(Comp)3 respectively as calculated by ANOVA ($p < 0.05$). (B) Karyotype analysis of three subclones with rescued *Gtl2* expression ($n \geq 15$). (C-D) *Rian* (C) and *Mirg* (D) expression in the *Gtl2* expressing ESC(Comp)3 subclones. ns means not significantly different from ESC(Comp)3 parental clone. 84
- 7.1 **Bacteriophage life cycle** After a bacteriophage attaches to a bacterial host cell, it injects its DNA which is then replicated and used for phage protein synthesis using the host's DNA replication and protein synthesis machinery. The different phage proteins are then assembled within the host cell and released through a channel in the outer cell membrane that is encoded by one of the phage genes. 92
- 7.2 **Antibody structure** (A) Structure of a full length antibody. The two heavy chains (HC) are shown in blue and the two shorter light chains (LC) are shown in green. The darker shaded ends are the variable regions (Fv) of each the HCs and LCs. Each variable region has three complementarity determining regions (CDRs) and four framework regions (FWs). The two heavy chains are held together by a disulfide bond in the hinge region. (B) Structure of a single chain variable fragment (scFv) containing only the variable regions of the heavy and light chains, held together with a polypeptide linker. 93
- 7.3 **Biopanning** Diverse phage library is added to the cells of interest. Some of the phage will bind to the cells, the rest is washed off. Bound phage is eluted, amplified and used for another round of selection. After several rounds the phage outputs will be enriched for phage binding to the cells of interest. 96

7.4	Experimental setup (A) Hypothesis 1: The ESC(Comp) cells are missing a cell surface marker (shown in green) that is vital for embryonic development. (B) Hypothesis 2: The ESC(Comp) cells overexpress a cell surface marker (shown in red) that is causing the embryonic lethal phenotype. (C) Experimental setup 1: three rounds of selection on the pool of ESC(Ctrl) cells and the pool of ESC(Comp) cells. (D) Experimental setup 2: three rounds of deselection to select for phage binding specifically to the ESC(Comp) cells.	97
7.5	Library diversity and dominant clones (A) The diversity of the original phage library drops each round (B) Graph displaying the ten motifs with the highest read count in the original library and their read counts in the subsequent round one to three outputs. There are a few dominant clones in the original library. These persist in the round one outputs but disappear in rounds two and three.	100
7.6	Evolution over selection rounds Top ten motifs in rounds one and three for Ctrl (A, B), Comp (D, E) and Deselection (G, H) outputs and their counts per round. (C, F, I) Venn diagrams of the top 50 motifs each round for Ctrl (C), Comp (F) and Deselection (I).	102
7.7	Differences between Ctrl, Comp and Deselection outputs Venn diagrams comparing the top 50 motifs of Ctrl, Comp and Deselection outputs after round one (A), round two (B) and round three (C).	102
7.8	scFv to IgG reformatting (A) To create the intermediate pool, the phage display vector is linearised by emulsion PCR amplification and fused with a donor fragment containing the hinge region, the heavy chain constant region and the light chain control elements. (B) The incomplete antibody cassettes from the intermediate pool are again amplified by emulsion PCR and cloned into a pmIgG v.5 vector containing the missing heavy chain control elements and light chain constant regions (λ/κ)	103

List of tables

1.1	Initial blastocyst injection results ES cell contribution (in %) was assessed by coat colour chimaerism.	15
2.1	Tail lysis buffer	23
2.2	RT reaction mix and thermal cycling conditions for Invitrogen and PCR Biosystems reagents	24
2.3	PCR reaction mix and thermal cycling conditions	25
2.4	qPCR reaction mix and thermal cycling conditions	26
2.5	List of primer pairs	27
2.6	PCR amplifying imprint regions	30
2.7	PCR attaching barcoded adapters to samples for sequencing	30
2.8	IgG conversion - Emulsion PCR for intermediate pool	32
2.9	IgG conversion - Preparation of the pmDV.v2 donor fragment	33
2.10	IgG conversion - Emulsion PCR for final pool	34
2.11	IgG conversion - Preparation of the IgG vectors	35
2.12	RT-PCR reaction mix and thermal cycling conditions for NGS samples	36
4.1	Transfection efficiencies	49
4.2	Copy number analysis	49
5.1	Blastocyst injection results of additional clones	68
5.2	Mouse models with defects at the Dlk1-Dio3 region	72
7.1	Output sizes of the three selection rounds	99
7.2	Blaze2 results of round two and round three outputs The numbers show absolute numbers (and %) of sequences.	100
7.3	Output sizes during IgG reformatting	104
7.4	Blaze2 results of intermediate and final pools	105

List of Abbreviations

General Abbreviations

5-aza 5-azacytidine

E. coli Escherichia coli

BMV Bone marrow Vaughan

C region Constant region

CDR Complementarity determining region

cfu Colony forming unit

Comp Compromised

CS Combined spleen

Ctrl Control

DMR Differentially methylated region

DNMT DNA methyltransferase

DS Deselection

EB Embryoid body

EC Embryonal carcinoma

EPSCM Expanded potential stem cell media

EPSC Expanded potential stem cell

EP Electroporation

- ES cell** Embryonic stem cell
- ESC(Comp)** Compromised ES cells
- ESC(Ctrl)** Control ES cells
- Fab** Fragments of antigen binding
- FB** Filamentous bacteriophage
- Fc** Fragment crystallisable
- FISH** Fluorescence *in situ* hybridisation
- FW** Framework region
- H chain** Heavy chain
- H&E** Haematoxylin and Eosin
- ICM** Inner cell mass
- IG-DMR** Intergenic DMR
- IgG** Immunoglobulin G
- iPSC** Induced pluripotent stem cell
- ITR** Inverted terminal repeat sequence
- L chain** Light chain
- LF** Lipofection
- lncRNA** Long noncoding RNA
- MEF** Mouse embryonic fibroblast
- miRNA** Micro RNA
- NF** Nucleofection
- NGS** Next generation sequencing
- PB** PiggyBac
- PCA** Principal component analysis

PGC Primordial germ cell

PRC2 Polycomb repressive complex 2

PrE Primitive endoderm

Q Phred score

RNAi RNA interference

scFv Single chain variable fragment

SNP Single nucleotide polymorphism

TET-CD TET catalytic domain

UPD Uniparental disomy

V region Variable region

VEC Vascular endothelial cell

Gene and Protein Names

α SMA α smooth muscle actin

 β III-Tubulin

Akt RAC-alpha serine/threonine-protein kinase

bFgf Basic fibroblast growth factor

Bmp Bone morphogenic protein

CD31/Pecam1 Platelet and endothelial cell adhesion molecule 1

Cdx2 Caudal type homeobox 2

Commd1 Copper metabolism domain containing 1

Dio3 Deiodinase iodothyronine type III

Dlk1 Delta like non-canonical Notch ligand 1

Dppa3 Developmental pluripotency associated 3

Egf Epidermal growth factor

Eomes Eomesodermin

Ephrin Eph family receptor-interacting proteins

Eph Erythroprotein-producing human hepatocellular carcinoma

Erk Extracellular-signal-regulated kinase

Esrrb Estrogen related receptor beta

Gata4/6 GATA binding protein 4/6

gp130 Glycoprotein 130

Gsk3 Glycogen synthase kinase 3

Gtl2/Meg3 Maternally expressed 3

H13 Histocompatibility minor 13

H19 H19 imprinted maternally expressed transcript

Hey1/2 Hairy/enhancer-of-split related with YRPW motif protein 1/2

Id Inhibitor of differentiation

Igf2r Insulin-like growth factor 2 receptor

Igf2 Insulin-like growth factor 2

Jagged

Jak Janus-associated kinase

Kiss1 Kisspeptin

Klf4 Kruppel like factor 4

LIF Leukaemia inhibitory factor

Mek Mitogen-activated protein kinase kinase

Mest Mesoderm specific transcript

Mirg miRNA containing gene

Myf5 Myogenic factor 5

Nanog Nanog homeobox

Nap115 Nucleosome assembly protein 1 like 5

neoR Neomycin resistance

Nespas-Gnasxl GNAS complex locus

Notch

Oct4/Pou5f1 Octamer binding protein 4

p53 Tumour protein 53

Pdx1 Pancreatic and duodenal homeobox 1

Peg13 Paternally expressed 13

PI3K Phosphoinositide 3-kinase

Rex1 RNA exonuclease 1 homolog

- Rian** RNA imprinted and accumulated in nucleus
- Rtl1** Retrotransposon gaglike 1
- Smad4/6** SMAD family member 4/6
- Sox17** SRY-Box 17
- Sox2** SRY-Box 2
- Stat3** Signal transducer and activator of transcription 3
- Tcf3** Transcription factor 3
- Tet** Tet methylcytosine dioxygenase
- Tgf- β 1** Transforming growth factor β 1
- Tie1** Tyrosine kinase with immunoglobulin-like and EGF-like domains 1
- Tnf α** Tumour necrosis factor α
- Vegfr1** Vascular endothelial growth factor receptor 1
- Vegf** Vascular endothelial growth factor
- Wnt** Wnt family member
- Zpf57** Zinc finger protein 698

Chapter 1

Introduction

1.1 A Brief History of Mouse Embryonic Stem Cells

The concept of "stem cells" has been around since at least the late nineteenth century when the German biologist Ernst Haeckel used it to describe the unicellular organisms that he believed to be the ancestors to all multicellular organisms [1]. Shortly after that, the term stem cells was used in the context of embryology for the first time [2, 3]. However, it was not until the mid twentieth century that scientists were starting to think about embryonic stem cells as we know them today.

In the 1950s, Leroy Stevens set the scene with his pioneering work on testicular teratocarcinomas. Teratocarcinomas are tumours that occur predominantly in the gonads and can contain a wide mixture of different tissue types derived from all three germ layers, including fully differentiated structures like muscle, hair or bone. Stevens discovered that approximately 1% of males of a particular mouse strain, the inbred 129 strain, develop these tumours sporadically in their testes and that these can be maintained by serial transplantation [4]. He then proceeded to show that grafting the genital ridges of developing embryos into the tissues of adult mice can efficiently induce teratocarcinoma formation. He further noted that some structures in the tumours resembled early embryo development [5]. This work was the first evidence that teratocarcinomas are of embryonic origin. Supporting this theory, another important study demonstrated that single undifferentiated tumour derived cells were able to fully regenerate teratocarcinomas when injected into adult mice [6]. However, the same was not possible using differentiated cells. These findings, as well as the close morphological resemblance of the undifferentiated tumour cells and the pluripotent cells of the early pre-gastrulation mouse embryo, led to the experiments showing that grafting early mouse embryos to extra-uterine sites gives rise to retransplantable teratocarcinomas [7, 8]. It was thus argued that teratocarcinomas originate from the pluripotent cells (or stem cells) of the early mouse embryo and

that undifferentiated teratocarcinoma cells would be a good model to study pluripotency and differentiation.

Within the next couple of years the focus of the field was, therefore, to establish cell lines of the undifferentiated tumour cells, also called embryonal carcinoma (EC) cells. Although somewhat successful, EC cell cultures were very variable and pluripotency was often lost over a few passages. This was improved when EC cells were cultured on a layer of mitotically inactivated mouse embryonic fibroblasts (MEFs or feeder cells) [9]. Subsequently, a number of groups reported that it was possible to make chimaeras from EC cells by injecting them into the blastocoel cavity of a host embryo [10–12].

However, deriving EC cells from teratocarcinomas is a lengthy process and the similarity between EC cells and the pluripotent embryonic cells remained questionable. Despite their capacity to contribute to chimaeras, these are usually relatively low contribution chimaeras and EC cells never contribute to the germline. To undergo this problem, both Martin Evans and Gail Martin independently set out to derive pluripotent embryonic stem (ES) cell lines directly from the embryo [13, 14]. They used the feeder cell system previously established for the culture of EC cells to culture whole blastocysts or just the inner cell mass (ICM) of blastocyst stage embryos, respectively. The derived ES cells have the same characteristics as EC cell lines: they can be maintained as pluripotent cells in culture, they can recover teratocarcinomas upon injection into adult mice and form tissues of all three germ layers. But more importantly, unlike EC cells, they are able to contribute to high contribution chimaeras and also populate the germ line [15]. These discoveries were the beginning of modern ES cell biology.

1.2 The Relationship Between ES Cells and the Early Embryo

We now know that ES cells are derived from the epiblast compartment of the developing embryo, that normally goes on to form the embryonic tissues (Figure 1.1). This epiblast lineage is established during preimplantation development, when the developing blastomeres undergo two lineage segregation events (Figure 1.1A). The first lineage decision occurs at the 8-cell stage at embryonic day e2.75 and begins with compaction and polarisation of the cells. During this process, cells on the outer part of the blastomere form tight junctions, leading to a separation of these cells and thereby providing essential spatial cues which enable cell polarisation and a clear distinction between inside and outside. The outer cells will have a bias towards the trophoblast lineage, giving rise to the placenta, whereas the ICM will later give rise to the embryo proper as well as yolk sac, allantois and amnion [16]. The trophoblast lineage is generally identified by *Eomes* (Eomesodermin) and *Cdx2* (Caudal type homeobox 2) [17] whereas the ICM expresses *Oct4* (Octamer binding protein 4, also known as *Pou5f1*) [18, 19]

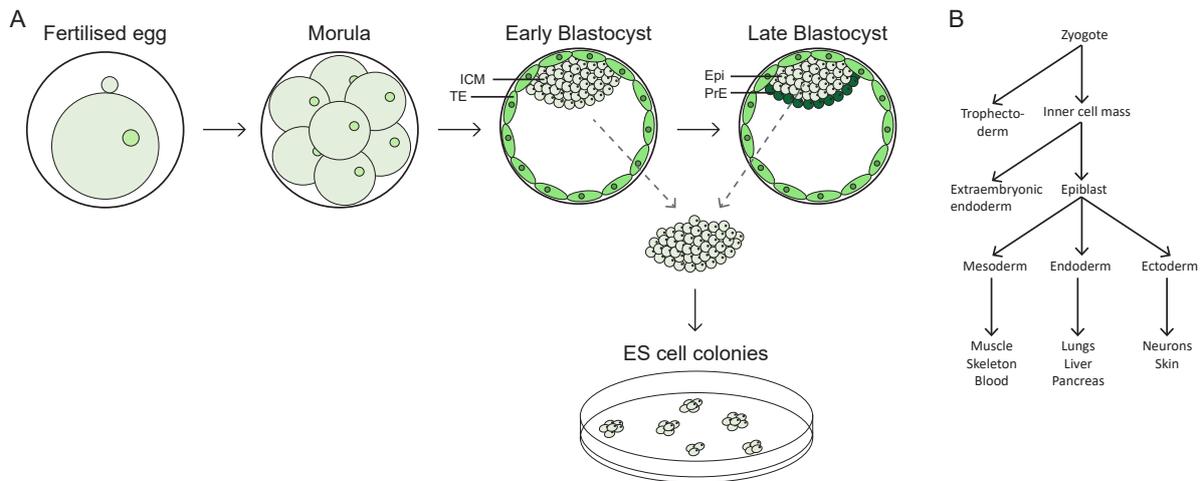


Fig. 1.1 Development of the early embryo (A) Schematic of the morphological changes that occur in the early embryo from the zygote to the blastocyst stage. ES cells can be isolated from the ICM or Epi compartment of blastocysts. ICM: inner cell mass, TE: trophectoderm, Epi: epiblast, PrE: primitive endoderm. (B) Development and lineage segregation from the zygote into the three main germ layers.

and *Nanog* [20]. At e3.25, the trophectoderm starts to secrete a fluid into the intercellular space of the early blastocyst, inducing blastocoel formation and localisation of the ICM to one side of this cavity. Through this process the distinctive blastocyst structure is created. During the second lineage decision, the ICM partitions into the second extraembryonic lineage, called the primitive endoderm (PrE), and the preimplantation epiblast. Only the epiblast will contribute to embryonic tissues, whereas the PrE goes on to form the yolk sac [21, 22]. Separation of these two lineages begins at e3.5 and is complete at e4.5. Initially, the ICM is a heterogeneous population of cells displaying different gene expression patterns. However, at e4.5 there is a clear distinction of epiblast and PrE as the latter forms an epithelial layer at the boundary of ICM and blastocoel. The PrE can be identified by the expression of the markers *Gata-4* and *Gata-6* (GATA binding protein 4/6) [23], whereas the epiblast expresses *Nanog* [24].

1.3 Pluripotency Is Maintained by a Transcription Factor Network

ES cells are generally defined as pluripotent cells that have the capacity to self renew in culture almost indefinitely and to contribute to all three germ layers (Figure 1.1B) both *in vivo* as well as *in vitro*. ES cells as well as epiblast cells in the embryo express a network of transcription factors that act together to sustain this pluripotent state. They work in such a way that they

promote cell proliferation and at the same time prevent the cells from differentiating. Previous studies have identified three transcription factors as the key players in this complex network, OCT4, SOX2 (SRY-Box 2) and NANOG.

1.3.1 OCT4

OCT4 is a POU-domain transcription factor that is expressed throughout the embryo until the first lineage segregation, after which its expression is restricted to the ICM and later to the epiblast cells only. It is never expressed in the trophoctoderm cells. In fact, one of its roles during early development is thought to be the inhibition of differentiation into trophoctoderm. First evidence for this was shown by Nichols et al. in 1998 [19] who showed that the ICM cells of *Oct4*-deficient embryos are not pluripotent but instead differentiate into the trophoblast lineage. Similarly, reduced *Oct4* gene expression in ES cells causes differentiation into trophoctoderm [25]. It was also shown that the precise level of expression is incredibly important. Not just reduced *Oct4* levels affect embryo development, but a 1.5-fold increase can already cause ES cell differentiation into the mesoderm or endoderm lineages.

1.3.2 SOX2

Expression of *Sox2*, which encodes a SRY-box transcription factor, can be detected in all cells of the early embryo from the oocyte to the blastocyst [26]. Loss of *Sox2* results in a phenotype that is almost identical to that of *Oct4*-deficient embryos [27, 28]. It is a binding partner of OCT4. This was first discovered when SOX2 was identified to activate *Fgf4* (Fibroblast growth factor 4) expression together with OCT4 [29, 30]. Since then, a number of target genes of OCT4 and SOX2 have been identified, including themselves, as well as the last of the three core pluripotency genes *Nanog*.

1.3.3 NANOG

NANOG is a homeobox-containing transcription factor. Unlike OCT4 and SOX2 its expression in the early embryo is restricted to the ICM and the epiblast and declines just before implantation. Both *in vivo* and *in vitro*, *Nanog*-deficient cells are not pluripotent. *Nanog*-null embryos develop normal until e3.5 but after that the embryos are lacking the epiblast and extraembryonic tissues develop in a disorganised structure. ES cells lacking *Nanog* expression differentiate into parietal endoderm-like cells [31, 32]. It is believed that the role of NANOG is the inhibition of the *Gata4* and *Gata6* genes which are important for primitive endoderm development [31].

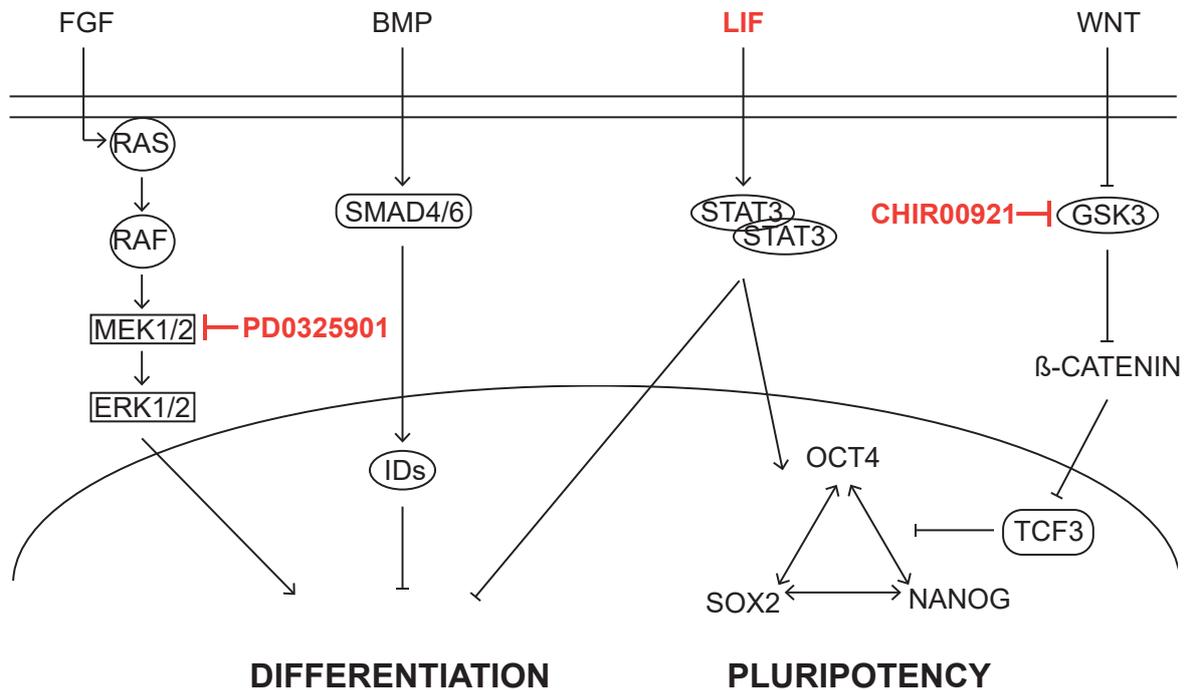


Fig. 1.2 **Signalling pathways in the maintenance of pluripotency during ES cell culture** BMP signalling through SMAD4/6 and the RAS/RAF/MEK/ERK pathway affect cell differentiation. LIF signalling both inhibits differentiation and activates pluripotency genes. The canonical WNT signalling pathway is involved in maintenance of pluripotency. Shown in red are the trophic factors that are added to ES media: LIF, the MEK1/2 inhibitor PD0325901 and the GSK3 inhibitor CHIR00921.

1.3.4 Other Factors

OCT4, SOX2 and NANOG are the master regulators at the heart of the pluripotency network. On the one hand they act through a positive feedback mechanism to regulate each other and on the other hand they work together to regulate downstream gene expression by activating genes involved in maintaining the pluripotent state while at the same time repressing genes that are important for differentiation. A number of other genes have been identified that are very important for establishing and maintaining the pluripotent state, including *Esrrb* (Estrogen related receptor beta), *Stat3* (Signal transducer and activator Of transcription 3), *Klf4* (Kruppel like factor 4), *Rex1* (RNA exonuclease 1 homolog) and many more, but most of these are regulated but at least one of the core pluripotency markers OCT4, SOX2 or NANOG.

1.4 Maintaining ES Cells in Culture

In vivo, the pluripotent state is very tightly regulated and it only really exists for approximately half a day. *In vitro*, external factors are needed to maintain pluripotency and self renewal of ES cells for a prolonged period of time. The first ES cell lines were cultured on MEFs in serum containing media. Since then a few different methods have evolved for culturing ES cells. These are summarised in Figure 1.2.

1.4.1 Serum/LIF Media

The first ES cells [13, 14] were cultured on a monolayer of MEF feeder cells in serum-based media conditions. It was quickly suggested that the feeders produce and provide one or more trophic factors that help maintain the ES cells' pluripotent state. This was later identified to be Leukaemia inhibitory factor (LIF) [33, 34]. It was shown that in the presence of serum containing medium, LIF is sufficient to maintain pluripotency in ES cells in the absence of feeders.

LIF belongs to the IL6 family of cytokines. Upon binding to its receptor, the glycoprotein 130 (gp130) is recruited and forms a heterodimer with the LIF receptor which leads to the recruitment and activation of JAKs (Janus-associated kinases) as well as STAT3. STAT3 is then activated by phosphorylation through JAK and forms homodimers which are translocated to the nucleus where they function as transcription factors. Amongst a number of other pluripotency related genes STAT3 also activates *Nanog* expression and thereby positively regulates the expression of the core pluripotency gene network. Importantly, it also acts as an inhibitor of differentiation into non-neuronal lineages.

Although there is no doubt that LIF is an extremely important regulator to maintain ES cell pluripotency in culture, it is not sufficient to do so on its own. In the absence of serum, ES cells will differentiate even in the presence of LIF [35], suggesting that the serum contains additional factors that are required to maintain pluripotency in culture. Ying et al. [35] have shown that replacing serum with bone morphogenic protein (BMP) is sufficient to block ES cell differentiation. BMP signals through SMAD4/6 (SMAD family member 4/6) to upregulate the expression of *Id* (Inhibitor of differentiation) genes. These suppress the differentiation of ES cells into neuronal lineages. So together the main function of LIF and serum (or BMP) seems to be to inhibit ES cell differentiation, thereby helping to maintain pluripotency.

1.4.2 2i Media

Recent efforts have been looking into more drivers for ES cell differentiation. This has led to the development of defined ES cell media conditions (also called 2i media). It has been shown that ERK (Extracellular-signal-regulated kinase) signalling positively affects the cells' ability to respond to differentiation cues [36] and *Erk2* mutant ES cells only need the addition of LIF to maintain pluripotency. Similarly, inhibition of MEK1/2 (Mitogen-activated protein kinase kinase 1/2) upstream of ERK1/2 combined with LIF supplementation is sufficient to maintain ES cells in culture without differentiation [37].

Another pathway that has been implicated to be important for the balance between ES cell pluripotency and differentiation is the Canonical WNT signalling pathway acting through β -CATENIN [38–40]. In the absence of WNT ligands, cytoplasmic β -CATENIN is phosphorylated through action of GSK3 (Glycogen synthase kinase 3) and thereby marked for degradation. In the presence of WNTs or the absence of GSK3, β -CATENIN is stabilised and moves to the nucleus. In the nucleus, it interferes with transcription factor TCF3-binding (Transcription factor 3) to promoters of important pluripotency genes [41]. TCF3 has a repressive function and by interfering with its DNA binding, β -CATENIN activates the expression of pluripotency genes. Ying et al. show that inhibition of GSK3, together with LIF can maintain ES cell pluripotency [37]. Additionally, LIF is dispensable when combining the two inhibitors of MEK1/2 and GSK3 (2i media). None of these require the addition of serum or BMP.

1.5 Epigenetic Regulation of Pluripotency and Differentiation

The pluripotency gene network in the epiblast and in ES cells is remarkable in that it tightly regulates the expression of genes required for self renewal but represses genes that will become active during differentiation. However, they are not the only key players. Epigenetics also plays a vital role in establishing and maintaining pluripotency as well as controlling differentiation and lineage specification. Two main types of epigenetic modifications exist in ES cells - DNA methylation and histone modifications. Additional epigenetic mechanisms include RNA-mediated regulation of gene expression, however it is beyond the scope of this thesis to discuss these.

1.5.1 DNA Methylation

DNA methylation entails the addition of a methyl group to the cytosine of a CpG dinucleotide. This is done with the help of different DNA methyltransferases (DNMTs). Most CpGs in the genome are methylated. Unmethylated CpGs are mostly clustered in so called CpG islands near promoter regions where they regulate gene expression [42]. Methylation of these CpG islands is usually associated with gene silencing whereas unmethylated CpG islands are associated with active gene expression [43]. This is because for transcription factors to be able to activate gene expression, the promoter needs to be accessible and DNA methylation physically blocks the DNA.

Directly after fertilisation, the DNA is demethylated genome wide (with the exception of imprinted genes, which will be discussed later in this section) and becomes remethylated shortly before implantation, when cell differentiation is initiated (Figure 1.3). DNA methylation reaches its minimum at the blastocyst stage [44]. This is reflected in the fact that ES cells have very low levels of DNA methylation and also have a very active transcriptome. An incredibly high number of genes is expressed, but most of them only at low levels. It is believed that in the embryo and in ES cells, DNA methylation is not important for the establishment and maintenance of pluripotency but more for lineage segregation and differentiation. Upon differentiation the promoters of *Oct4* and *Nanog* become hypermethylated and their transcription is silenced. ES cells that do not express any DNMTs are viable and can proliferate and self renew, but they are unable to differentiate as they cannot upregulate genes required for germline specification, nor silence pluripotency factors such as *Oct4* and *Nanog* [45, 46].

1.5.2 Histone Modifications

The second main type of epigenetic modifications are histone modifications. DNA is wrapped around structural and functional proteins called histones. This tight packaging is called chromatin. Like DNA methylation, histone modifications can alter gene expression by changing chromatin structure. However, epigenetic regulation of gene expression by histone modifications is much more complex. Types of modifications include methylation, phosphorylation, acetylation and ubiquitylation. Depending on the type of modification as well as the histone and amino acid that are modified, gene expression can either be activated or silenced. For example, tri-methylation of the histone 3 lysines 9 and 27 (H3K9me3 and H3K27me3) are generally associated with a closed, repressive chromatin state whereas H3K4me3 and H3K9 acetylation (H3K9ac) are mostly associated with an open and accessible chromatin state which allows transcription factors to bind and activate gene expression.

In ES cells, most of the core pluripotency transcription factors carry the active histone mark

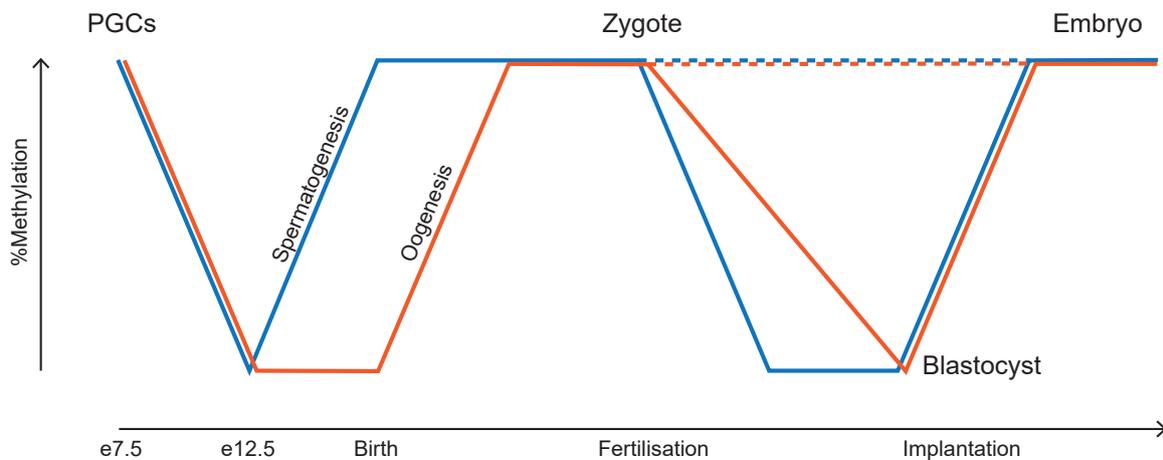


Fig. 1.3 Waves of DNA demethylation A first wave of demethylation occurs in the primordial germ cells (PGCs). Parent-specific methylation patterns are imprinted during germ cell development. Shortly after fertilisation, a second wave of demethylation occurs and a new methylation signature is established around implantation. Imprinted genes with parent-specific methylation patterns are spared from the second round of demethylation (dotted lines). Red line is the female genome, blue line is the male genome.

H3K4me₃, allowing for their active transcription. On the other hand, genes that are not required until later during development are usually stably repressed. Interestingly, most genes that are important for lineage specification in the pre- and peri-implantation embryo usually carry both repressive as well as a smaller number of activating histone marks. This so called bivalent mark keeps them in a state that allows for fast activation upon differentiation from pluripotent cells [47].

1.5.3 Imprinted Gene Expression in Embryo Development

Epigenetic mechanisms, such as DNA methylation and histone modifications, remain crucial during the rest of development as they are vital for further lineage specification and tissue specific gene expression. One particular epigenetic mechanism that relies on DNA methylation and is important during embryonic and fetal development is genomic imprinting. Prior to the wave of DNA demethylation in the early pre-implantation embryo, DNA methylation marks are erased in the primordial germ cells (PGCs) in the male and female germline (of their father and mother, respectively). In the male, remethylation takes place in the prospermatogonia a few days later, whereas in the female this process is only initiated after birth with oocyte development (Figure 1.3). Although largely similar, the methylation signatures of male and female germ cells are distinct and there are a few genes that have a defined parent-specific signature (i.e. they

are methylated in only the male or the female germ cells) - these genes are called imprinted genes. They keep their methylation signature throughout development and adult life. As mentioned before, there is another round of DNA demethylation after fertilization. To maintain the parent-specific signature of the imprinted genes, these are spared from demethylation. The mechanisms for this are not fully understood but *Dppa3* (Developmental pluripotency associated 3) and *Zfp57* (Zinc finger protein 698) have both been suggested to play a role [48, 49].

Imprinting is known to be an incredibly important mechanism that controls and regulates developmental processes. This was first realised in the 1980s when uniparental embryos harbouring either two maternal (gynogenic) or two paternal (androgenic) genomes were created [50, 51]. These are not viable - gynogenic embryos show a lack of extraembryonic tissues, whereas androgenic embryos fail to produce embryonic tissues, suggesting that the correct expression of imprinted genes is important for early lineage segregation and differentiation processes. Since then, it has become clear, that imprinted genes have a variety of different roles but one of the main responsibilities seems to be to regulate embryonic development and growth. Interestingly, paternally expressed genes are generally associated with fetal growth, whereas maternally expressed genes are associated with fetal growth restriction. *Igf2* (Insulin-like growth factor 2) for example is a paternally expressed gene and normally only the one paternal copy is present [52]. Expression from both alleles (for example through hypomethylation of the maternal copy) will lead to an over-dose of IGF2 and a severely growth enhanced embryo. On the contrary, silencing of the paternal allele (for example through hypermethylation) results in a growth restricted embryo [53]. Many imprinted genes are expressed not only in the embryo but also in the placenta and they can indirectly affect embryonic growth and development through regulation of placental development. Similarly to embryonic development, paternally expressed genes enhance placental growth whereas maternally expressed genes negatively regulate placental growth. Since the placenta acts as the interface between mother and embryo, which supplies the embryo with important nutrients and growth factors, regulation of placental growth and development by imprinted genes also has an effect on embryonic growth and development [54]. So far, approximately 150 imprinted murine genes are known and aberrant regulation of almost all of them has consequences for embryonic development.

1.6 Applications of Mouse ES Cells

ES cells have become an incredibly useful tool in a number of scientific areas in recent years. While human stem cells research is often focussed on the therapeutic application of these cells, mouse ES cells are mainly used to study developmental processes or diseases.

1.6.1 Studying Developmental Processes and Modelling Diseases

In Vitro

One of the characteristics of ES cells is that they can differentiate into cell types of the three germ layers and it is well known that ES cells will spontaneously differentiate in culture if they are grown without the addition of a feeder layer or LIF. Often, beating clusters of cardiomyocytes or cells that morphologically resemble neurons can be observed. However, most cell types do not spontaneously arise and the addition of specific growth factors or extracellular matrix proteins is required. Over the years, a number of different protocols have been published for lineage-specific differentiation of ES cells. These include protocols for the generation of cardiomyocytes [55–57], neurons [58–60], pancreatic cells [61, 62], retina [63], adipocytes [64], skeletal muscle [65], renal cells [66] and germ cells [67]. These differentiation models are extremely useful not only because they give access to rare progenitor cell populations, but also because it is now possible to study developmental mechanisms *in vitro* and look at the effect of altered gene expression on developmental processes. This is particularly useful when studying diseases. Specific genes of interest that are known to be affected in diseases can be altered in the ES cells, for example by knocking them out or making them constitutively active. These mutations will then persist throughout differentiation and the effect throughout the developmental stages can be analysed. This can be extremely useful for studying neurodegenerative diseases. In the animal (or of course human patients) a phenotype often only presents after considerable damage has already occurred in the cells. With ES cell culture models, it is possible to study the early stages of diseases and identify changes and defects on a cellular level. Some techniques such as RNA interference (RNAi) even make it possible to look at the effect of aberrant gene expression during specific developmental timepoints by only interfering with its expression for a specified time window.

In Vivo

In addition to studying gene function *in vitro*, ES cells can also be used to study the role of genes *in vivo* in the context of the whole animal. Usually this is done by manipulating genes in the ES cells and then injecting ES cells into blastocyst stage embryos to generate chimaeras. If the ES cells have contributed to the germline, these chimaeras can be bred to homozygous wild-type mice to obtain a heterozygous F1 generation. Subsequent interbreeding of F1 animals will result in 25% of homozygous mutant animals in the F2 generation. This method is widely used to generate a vast number of transgenic mouse lines. The effect of aberrant gene expression can then be studied *in vivo*. The advantage of this over using just a cell culture model as described above is that it is possible to investigate the effect of the mutation on a number of tissues and

on the interaction between cells of different types of tissues. In culture the focus is usually only on cells from one type of tissue. Furthermore, any changes that are observed on the cellular level are more likely to be physiologically relevant.

1.6.2 Therapeutic Applications

Although the focus of research using mouse ES cells lies more on development and disease biology, ES cells also have indirect potential in the therapeutic area. Firstly, the differentiated ES cell disease models can be used for drug discovery and to test drug efficacy. Many diseases can have a number of genetic causes and a lot of drugs are not effective for treating all of them because they may not act on the affected biological pathway. Having a cell culture model with a specific mutation can shed light on which drugs may be effective for which mutation. These can then be further explored *in vivo* and in human samples.

Additionally, research into mechanisms of self renewal and cell proliferation of ES cells can indirectly help cancer research. Most cancers arise from and harbour niches of so called cancer stem cells and it is believed that the mechanisms of self renewal and proliferation in these cells are similar to or the same as in ES cells [68].

1.7 Making Chimaeras Using Targeted ES Cells

1.7.1 Gene Targeting

As mentioned above, to study developmental processes and disease progression *in vivo*, genetically modified ES cells can be injected into blastocyst stage embryos where they will combine with the ICM and contribute to the developing chimaeric embryo. Although this area of research is quickly developing and evolving with the arrival of CRISPR technology for genome editing [69], for a long time gene targeting by homologous recombination has been the method of choice to introduce specific genetic changes into the ES cell genome. All ES cells used in this thesis have been genetically modified that way.

Gene targeting by homologous recombination can be used to introduce almost all kinds of mutations, ranging from knockouts, insertions, large deletions and point mutations to conditional knockouts or knockins that depend on the presence of a marker specific to the tissue or cells of interest. In their simplest version, targeting vectors only contain a selectable marker such as the bacterial neomycin resistance gene (*neoR*), as well as two flanking arms that have sequence homology with the gene or locus of interest. The cell's DNA repair machinery can then replace the targeted locus with the homologous exogenous genomic sequence. These simple vectors are usually replacement targeting vectors that disrupt the normal function of the gene of interest

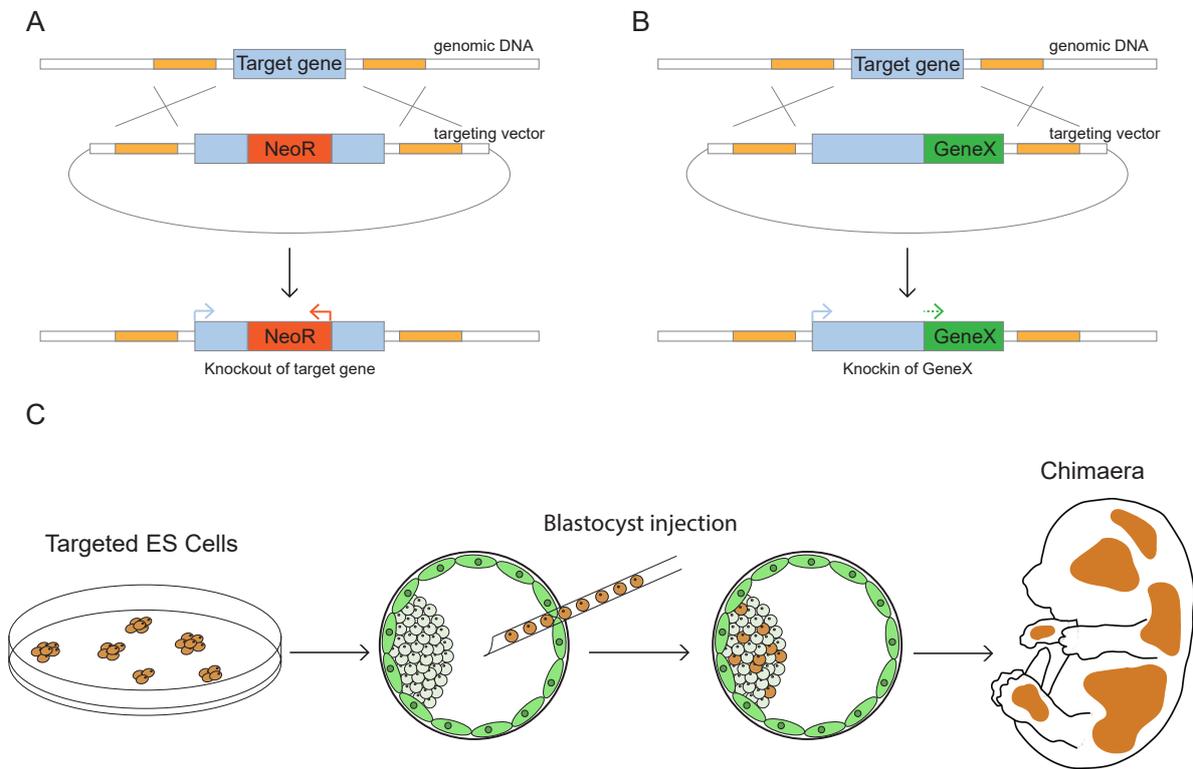


Fig. 1.4 Making chimaeras using targeted ES cells (A) Gene knockout by homologous recombination. The targeting vector contains the disrupted target gene with the selection marker *NeoR*. The targeted gene is not functional anymore. (B) Gene knockin by homologous recombination. The targeting vector contains the intact target gene as well as the cDNA for an additional gene (*GeneX*). This results in the expression of both the target gene and *GeneX*. (C) Targeted ES cells can be injected into blastocysts. They will combine with the ICM and contribute to chimaeras.

through the insertion of the selection marker (Figure 1.4A). Knockin vectors have a similar design but they contain an additional genetic element that will be driven off the promoter of the targeted locus. The exact positioning of the targeting vector into the recombined locus is important both for making sure that the knocked in gene is in frame and to determine whether the targeted gene is still expressed (in addition to the knocked-in gene) or not (Figure 1.4B). These targeting vectors are usually introduced into ES cells by electroporation (other methods are for example lipofection or nucleofection). However, homologous recombination is a very rare event and usually several hundreds of ES cell colonies have to be analysed to identify a small number of clones that contain the correct gene targeting event.

1.7.2 Blastocyst Injections

To generate chimaeras using such targeted ES cells, two main strategies can be used: morula aggregations or, more commonly, blastocyst injections. For morula aggregations, small clumps of ES cells are combined with 8-cell morulae (e2.5) and left over night to aggregate and develop into a blastocyst. All chimaeras described in this thesis were, however, generated by blastocyst injections. For this, blastocysts are collected from plugged females (that were mated with stud males) at 3.5 days post coitum by flushing the uterine horns. Using a micromanipulator, 10-12 individual ES cells are injected into the blastocoel cavity. A number of these are then transferred back into pseudopregnant females that were mated with vasectomised males to develop to term.

The injected ES cells will combine with the ICM of the host embryo to form the epiblast which will later go on to become the embryo proper. The developing chimaeric embryo then contains cells from the host blastocyst as well as cells that stem from the ES cells (Figure 1.4C). The degree of chimaerism can vary but a successful blastocyst injection can very often yield pups with >90% ES cell contribution. ES cells do not contribute to the extraembryonic tissues.

1.8 Some ES Cell Lines Do Not Contribute to Viable Chimaeras

The generation of chimaeric embryos by blastocyst injection is a technique that is well established in our laboratory and used on a regular basis in collaboration with a group at Medimmune (Cambridge, UK). All ES cells used in the experiments described in this thesis have come out of a sequential gene targeting project run by Medimmune, meaning that multiple site-specific mutations were introduced into the cells. Since the genes of interest whose expression was altered in this sequential gene targeting project are not known to be implicated in or necessary for embryonic development, most ES cell clones give rise to viable high contribution chimaeric pups. Occasionally, the cells can be lost from the blastocyst for a variety of reasons, including technical reasons or abnormal karyotype, resulting in pups without ES cell contribution.

During this collaboration, a number of compromised targeted ES cell clones have arisen that repeatedly did not give rise to viable chimaeras. In this case, as shown in Table 1.1, for some ES cell lines no pups were born. It can sometimes be the case that none of the blastocysts implant into the uterus of the pseudopregnant recipient female, which would, of course, also not result in the birth of pups. However, the recipient female mice were routinely weighed to monitor pregnancy and on a number of occasions they gained weight but then lost it again in the second half of pregnancy, meaning that they were pregnant but the embryos were lost mid

Table 1.1 **Initial blastocyst injection results** ES cell contribution (in %) was assessed by coat colour chimaerism.

Cell line	Number of pups	Of which chimera (% ES cell contribution)
ESC(Ctrl) 1	NA	NA
ESC(Ctrl) 2	3	2 (>90%)
ESC(Ctrl) 3	8	5 (>95%)
ESC(Ctrl) 4	4	3 (2x >95%)
ESC(Comp) 1	0	0
ESC(Comp) 2	0	0
ESC(Comp) 3	1	0
ESC(Comp) 4	0	0

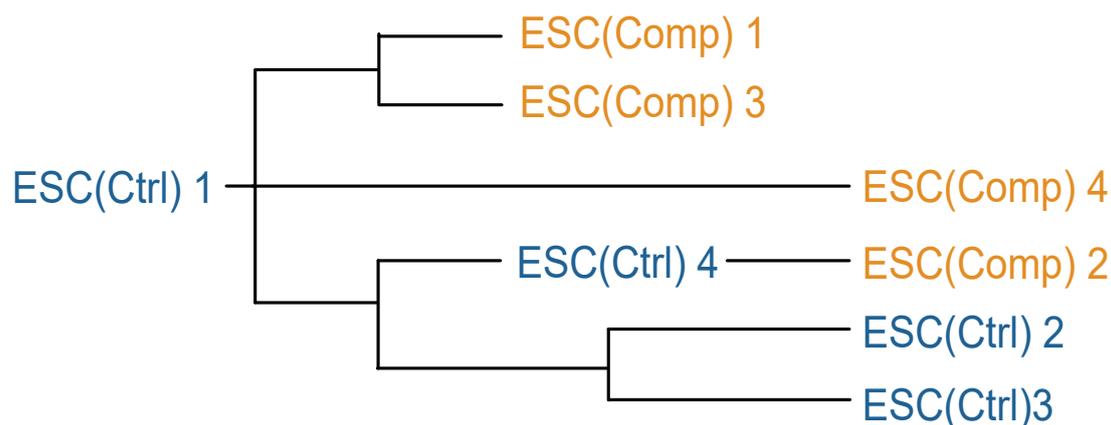


Fig. 1.5 **ES cell pedigree** Pedigree of how the ES cell lines were derived by sequential gene targeting.

to late gestation. When opening the abdominal cavity of these pregnant females at what would have been term, only decidua and dead embryos were observed. This suggests that the ES cells have contributed to the developing embryo but they have died prenatally. In comparison, all of the control cell lines had given rise to a number of excellent chimaeras after just one injection session. One of the control ES cell clones (ESC(Ctrl)1) is the parental ES cell line to all other cell lines used in this project and has not undergone gene targeting (ES cell pedigree shown in Figure 1.5). These were never injected into blastocysts. However, since subclones of this line have previously given rise to high percentage chimaeras, it is assumed to be a good control.

For simplicity, the compromised cell lines will be referred to as ESC(Comp) and the control cell lines to which they are compared will be referred to as ESC(Ctrl) throughout this thesis.

1.9 Hypothesis, Aims and Scope of This Thesis

We are very certain that the targeted mutations are not the cause for the observed embryonic lethal phenotype in the ESC(Comp) chimaeras. All cell lines were screened for off-target mutations or integration sites and for every gene targeting event a number of ES cell clones (sister clones) were picked that have the identical genotype at the mutated locus of interest. For all of the ESC(Comp) cell lines, at least one of those sister clones was also injected and gave rise to a number of viable high percentage chimaeras.

Our hypothesis is that these compromised ES cell clones have undergone genetic or epigenetic changes either during the subcloning or during subsequent culture that prevent them from undergoing normal differentiation processes and therefore cause embryonic lethality. The aims for this PhD project are therefore to further define the developmental phenotype of the chimaeric embryos and to discover the underlying changes in the transcriptome, epigenome or cell surface markers that have occurred in the compromised ES cells (ESC(Comp)), with the overall goal to identify a cell marker that can distinguish between compromised ES cells and ES cells with good developmental potency.

Chapter 2

Methods

2.1 Cell Culture

2.1.1 Cell Culture Media

MEF Medium

Advanced DMEM (Invitrogen)

10% FCS (Millipore)

2 mM GlutaMax (Gibco)

Standard ESC Medium

DMEM (Invitrogen)

15% FCS (PAN Biotech)

4 mM GlutaMax (Gibco)

1X non essential amino acids (Gibco)

1 mM Sodium pyruvate (Gibco)

20 mM HEPES buffer (Gibco)

0.1 mM β -Mercaptoethanol (Gibco)

1.5×10^3 U/ml LIF (Millipore)

2i Medium

NDiff 227 (Takara)

1×10^3 U/ml LIF (Millipore)

3 μ M CHIR99021 (Cayman)

1 μ M PD0325901 (Cayman)

EPSC Medium

DMEM/F12 (Invitrogen)
20% Knockout serum replacement (Gibco)
2 mM GlutaMax (Gibco)
1X non essential amino acids (Gibco)
0.1 mM β -Mercaptoethanol (Gibco)
1x10³ U/ml LIF (Millipore)
1 uM PD0325901 (Cayman)
3 uM CHIR99021 (Cayman)
4 uM JNK Inhibitor VIII (Cayman)
10 uM SB203580 (Generon)
0.3 uM A-419259 (Cayman)
5 uM XAV939 (Sigma)

Differentiation Medium I

IMDM (Invitrogen)
20% FCS (PAN Biotech)
2 mM glutamine (Gibco)
1X non-essential amino acids (Gibco)
450 uM 1-thioglycerol (Sigma)

Differentiation Medium II

DMEM/F12 (Invitrogen)
5 ug/ml insulin (Sigma)
30 M sodium selenite (Sigma)
50 ug/ml transferrin (Generon)
5 ug/ml fibronectin (Invitrogen)

Differentiation Medium III

DMEM/F12 (Invitrogen)
25 ug/ml insulin (Sigma)
30 nM sodium selenite (Sigma)
50 ug/ml transferrin (Generon)
20 nM progesterone (Sigma)

100 uM putrescine (Sigma)
1 ug/ml laminin (Sigma)
10 ng/ml bFGF (R & D Systems)
20 ng/ml EGF (R & D Systems)

Differentiation Medium IV

DMEM/F12 (Invitrogen)
25 ug/ml insulin (Sigma)
30 nM sodium selenite (Sigma)
50 ug/ml transferrin (Generon)
20 nM progesterone (Sigma)
100 uM putrescine (Sigma)
1 ug/ml laminin(Sigma)
2% B27 (Invitrogen)
10 mM Nicotinamide (Sigma)

2.1.2 Cell Culture Substrates

Gelatin

Cell culture plates were coated with 0.1% (w/v) Gelatin (Sigma) in PBS for 15 min at room temperature and then washed with PBS twice. Cells were plated immediately.

Poly-L-ornithine/Laminin

Cell culture plates were coated with 0.1 mg/ml Poly-L-ornithine (Sigma) for 1 h at room temperature and then washed with PBS twice. Poly-L-ornithine plates were then coated with 0.001 mg/ml laminin (Sigma) at 37°C for 1-3 h. After two washes with PBS the plates were either used immediately or stored at 4°C.

2.1.3 Other Cell Culture Reagents

0.25% Trypsin-EDTA (Gibco)
DPBS (Gibco)
5-azacytidine (Sigma)
1.75 ug/ml Puromycin (Fisher)
150 ug/ml Hygromycin (Fisher)
Mitomycin C (Sigma)

2.1.4 Mouse Embryonic Fibroblast (MEF) Cell Culture and Inactivation

Passage 0 MEFs (Taconic) were obtained from Medimmune. The cells were cultured in MEF medium in a 37°C humidified incubator with 5% CO_2 . When 70-80% confluent, the cells were passaged and after two rounds of passaging, MEFs were inactivated by incubating them in normal MEF medium plus 10 ug/ml Mitomycin C for 3 h at 37°C.

2.1.5 ES Cell Culture

ES cells were also obtained from Medimmune and cultured on MEFs in standard ESC medium in a 37°C humidified incubator with 5% CO_2 . When 70-80% confluent cells were passaged. For *Gtl2* rescue experiments, ES cells were cultured feeder-free on gelatin-coated plates in 2i or EPSC media, or on MEFs in standard ES cell media supplemented with 5-azacytidine.

2.1.6 Karyotyping

The cells were passaged one day before karyotyping and fed fresh ESC medium 3 h before chromosome preparation. The cells were incubated for 2.5 h in colcemid (0.2 ug/ml in culture medium; Gibco) at 37°C. After trypsinisation, the cells were spun down at 1100 rpm for 5 min and gently resuspended in 5 ml hypotonic solution (75 mM KCl; Sigma). After 8 min incubation at room temperature, the cells were spun down at 1100 rpm for 3 min and then gently resuspended in 5 ml fixative (3:1 methanol:acetic acid) and left at -20°C for 30 min. The fixative step was repeated and the cells were subsequently resuspended in 500 ul fixative. Microscope slides were heated to 36°C in a water bath. From a height of 15-30 cm, 30 ul of cell suspension was dropped onto each slide and left until the liquid had evaporated. The slides were bathed in Giemsa stain solution for 3 min and washed with water before being analysed.

2.1.7 Differentiation Assays

Differentiation assays were performed as described previously [55, 58, 62].

Embryoid Body Formation

Embryoid bodies (EBs) were formed in hanging drops in differentiation medium I. 20 ul drops containing a defined number of ES cells were placed on the lids of 10 cm bacteriological petri dishes containing 10 ml PBS to avoid evaporation of drops. The ES cells were cultivated in hanging drops for 2 days and then transferred into a 6 cm bacteriological petri dish with differentiation medium I for 2-3 days.

Mesoderm Differentiation

For differentiation into cardiac muscle progenitors 600 ES were used for EBs. After 4 days (2+2) EBs were transferred into gelatine-coated 6 cm tissue culture dishes and cultured in differentiation medium I for 4 days. At day 4+4 the cells were fixed and stained for α -smooth muscle actin.

Ectoderm Differentiation

For neuronal differentiation 200 ES cells were used for EB formation. After 4 days (2+2), the EBs were transferred into gelatin-coated 6 cm tissue culture dishes and cultured in differentiation medium I for one day. At 4+1 days the medium was exchanged with differentiation medium II. At day 4+8, the EBs were dissociated with trypsin and replated onto poly-L-ornithine/laminin coated tissue culture dishes and cultured in differentiation medium III for a further 6 days. At day 4+14 the cells were fixed and immunostained for the neural marker β III-tubulin.

Endoderm Differentiation

For differentiation into endoderm progenitor cells 600 ES cells were used for EB formation. After 5 days (2+3) the EBs were transferred into gelatine-coated 6cm tissue culture dishes and cultured in differentiation medium I for 6 days. At day 5+6, they were dissociated with trypsin and replated onto poly-L-ornithine/laminin coated tissue culture dishes in differentiation medium IV supplemented with 20% FCS. After 24h the medium was replaced with medium IV without FCS. At day 5+10 the cells were fixed and immunostained for the endoderm marker Sox17.

Immunofluorescence on Cells

The cells were grown on coverslips and then fixed in 4% paraformaldehyde in 1X PBS for 10 min, washed twice in PBS/PVP (3 mg/ml PVP, Sigma), permeabilised for 30 min in 1X PBS/PVP with 0.25% Triton X-100 (Sigma) and blocked in blocking buffer for 1 h. Blocking buffer contained 1X PBS with 2% goat serum (Sigma), 1% bovine serum albumin (BSA, Sigma) and 0.1% Tween 20 (Sigma). The primary antibodies were diluted in blocking buffer and added to the cells for overnight incubation at 4°C. The cells were washed three times using blocking buffer for 15 min. The secondary antibody was diluted in blocking buffer and added to the cells for 1 h at room temperature. The cells were washed three times using blocking buffer for 15 min. Coverslips were mounted with Vectashield with DAPI (Vector) on slides,

sealed and left to set overnight at 4°C in the dark.

The primary antibodies used were against α SMA (Abcam, ab5694, 1:100), β -III Tubulin (Abcam, ab18207, 3.25 ug/ml) and Sox17 (R&D Systems, MAB1924, 10 ug/ml). The secondary antibodies were goat anti-rabbit IgG H&L AF568 (Abcam, ab175471, 1:500), goat anti-rabbit IgG H&L AF488 (ab150081, 1:500) and goat anti-mouse IgG H&L FITC (ab6785, 1:500).

2.1.8 Transfection Methods

For all methods, the cells were split and plated at a density of 5×10^6 cells per 10cm MEF plate 24 h prior to the transfection. The vectors were used at the concentration of 2 ng of pPBCAG-Venus-IP and pPBCAG-H2BtdTomato-IH and 1 ng of CAGG-PBase (pCyL43) [70] per 5×10^6 cells. All vectors were kindly supplied by Professor Azim Surani (The Gurdon Institute, University of Cambridge). After 24 h without selection, the medium was replaced with fresh ESC medium containing the appropriate antibiotics (1.75 ug/ml Puromycin or 150 ug/ml Hygromycin). Cells were selected for 9 days with medium change every 1-2 days.

Electroporation

2×10^7 cells were resuspended in 700 ul PBS (Gibco) and mixed with the targeting vectors which were diluted in 100 ul PBS. After incubation on ice for 5 min the cell/DNA suspension was transferred into an electroporation cuvette and electroporated in a BioRad Gene Pulser at 240 V, 500 uF with a time constant in the range of 11-13 ms. The cells were seeded onto pre-prepared MEF plates at 5×10^6 cells per plate.

Nucleofection

Nucleofection was carried out using the Mouse ES Cell Nucleofector Kit (Lonza). 5×10^6 cells were resuspended in 90 ul Mouse ES Nucleofector Solution and mixed with the targeting vectors which were diluted in 10 ul Mouse ES Nucleofector Solution. The cells were nucleofected using the A-013 program and then seeded onto pre-prepared MEF plates at 5×10^6 cells per plate.

Lipofection

Lipofectamine 3000 (Invitrogen) was used according to the supplier's protocol. Diluted DNA and Lipofectamine were incubated for 5 min at room temperature and then directly added to ES cells growing on MEFs. After 24 h the cells were split and 5×10^6 cells were seeded onto each pre-prepared MEF plate.

2.2 Molecular Biology Techniques

2.2.1 gDNA Extraction

gDNA was extracted from cells using the PureLink Genomic DNA Kit (Invitrogen). Using a NanoDropND-1000 spectrophotometer, quality and quantity of the gDNA was tested. A 260/280 ratio above 1.8 was accepted for gDNA.

For genotyping of embryos, gDNA from tail samples was extracted by incubating the tissue with 100 μ l of lysis buffer over night at 55°C and subsequent Proteinase K inactivation at 94°C for 15 min. Components of the lysis buffer are outlined in Table 2.1.

Table 2.1 Tail lysis buffer

Component	Concentration (in H_2O)
KCl (Sigma)	50 mM
Tris-HCl pH 8.3 (Sigma)	10 mM
$MgCl_2$ (Sigma)	2 mM
Gelatin (Sigma)	0.1 mg/ml
Tween-20 (Sigma)	0.45%
Nonidet P-40 (Roche)	0.45%
Proteinase K (Sigma)	20 mg/ml%

2.2.2 RNA Extraction

RNA was isolated from cells or tissues using the RNeasy Plus Mini Kit (QIAGEN) or the PureLink RNA Mini Kit (Invitrogen) and RNase Free DNase I(QIAGEN) following the manufacturers' protocols. Using a NanoDropND-1000 spectrophotometer, quality and quantity of the RNA was tested. A 260/280 ratio above 2.0 was accepted for RNA.

2.2.3 cDNA Synthesis

First strand cDNA synthesis was performed using SuperScript III First-Strand cDNA Synthesis for RT-PCR (Invitrogen) or the qPCRBIO cDNA Synthesis Kit (PCR Biosystems). RT-PCR was performed in a Veriti 96Well Thermal Cycler (Thermo Fisher). Components of the reaction mix and thermal cycling conditions are outlined in Table 2.2. cDNA quality and quantity was measured using a NanoDropND-1000 spectrophotometer. A 260/280 ratio above 1.8 was accepted for cDNA.

Table 2.2 **RT reaction mix and thermal cycling conditions for Invitrogen and PCR Biosystems reagents**

Reaction mix (per reaction)	Thermal cycling conditions
1 ug total RNA	65°C for 5 min
1 ul 50 mM oligo(dT) (Thermo Scientific)	On ice for 1 min
1 ul 500 ug/ml random primers (Promega)	
1 ul 10 nM dNTP mix (Promega)	
H ₂ O (to obtain volume of 13 ul)	
Add:	25°C for 5 min
4 ul 5X First Strand Buffer (Invitrogen)	50°C for 60 min
1 ul 0.1 mM DTT (Invitrogen)	70°C for 15 min
1 ul 40 U/ul RNaseOUT (Invitrogen)	
1 ul 200 U/ul Superscript III RT (Invitrogen)	
Add:	37°C for 20 min
1 ul RNase H (New England Biolabs)	
Reaction mix (per reaction)	Thermal cycling conditions
1 ug total RNA	42°C for 30 min
4 ul 5x cDNA Synthesis Mix (PCR Biosystems)	85°C for 10 min
1 ul 20x RTase (PCR Biosystems)	
H ₂ O (to obtain volume of 20 ul)	
Add:	37°C for 20 min
1 ul RNase H (New England Biolabs)	

2.2.4 PCR

Standard PCRs using cDNA (for gene expression analysis) or gDNA (for genotyping analysis) were performed as outlined in Table 2.3. Primer sequences are listed in Table 2.5. The reactions were run in a Veriti 96Well Thermal Cycler (Thermo Fisher).

2.2.5 qPCR

qPCR was performed using qPCRBIO SyGreen Mix Lo-ROX (PCR Biosystems) on the 7500 Fast Real-Time PCR System (Applied Biosystems). Components of the reaction mix and thermal cycling conditions are outlined in Table 2.4.

All primers were designed using the NCBI Primer-Blast tool and are listed in Table 2.5. The efficiency of all primers was assessed using serial dilutions to generate a standard curve.

Relative fold changes were calculated using the delta-delta-Ct method. Appropriate statistical

analyses were performed using GraphPad Prism and results were considered significant where the P-value was <0.05.

Table 2.3 PCR reaction mix and thermal cycling conditions

Reaction mix (per reaction)	Step	Thermal cycling conditions
2 ul 10X Buffer B (NEB)	DNA polymerase activation	95°C for 3 min
0.4 ul dNTPs (Promega)	40 cycles of:	40 cycles of:
0.8 ul 10 uM Forward Primer	DNA denaturation	95°C for 15 sec
0.8 ul 10 uM Reverse Primer	Primer annealing	60°C for 30 sec
0.08 ul Taq DNA polymerase (NEB)	Primer extension	68°C for 1 min
14.92 ul Nuclease free H_2O	Final extension step	68°C for 5 min
1 ul cDNA/gDNA template		

Table 2.4 qPCR reaction mix and thermal cycling conditions

Reaction mix (per reaction)	Step	Thermal cycling conditions
7.5 ul SYBR Green master mix (PCR Biosystems)	DNA polymerase activation	95°C for 2 min
0.15 ul 10 uM Forward Primer	40 cycles of:	40 cycles of:
0.15 ul 10 uM Reverse Primer	DNA denaturation	95°C for 5 sec
6.2 ul DNA/RNA-free H_2O	Primer annealing/extension	60°C for 30 sec
1 ul cDNA	Melt curve	95°C for 15 sec 60°C for 1 min 95°C for 30 sec 60°C for 15 sec

Table 2.5 List of primer pairs

Gene	Forward Primer (5'-3')	Reverse Primer (5'-3')
qPCR		
HPRT	GATTAGCGATGATGAACCA	CCTCCCATCTCCTTCATGAC
Oct4	GAGCACGAGTGGAAAGCAAC	CTCTTCTGCTTCAGCAGCTTG
Sox2	AACGGCAGCTACAGCATGATGC	CGAGCTGGTCATGGAGTTGTAC
Klf4	GTGACTATGCAGGCTGTGGC	TGTGTGTTTGCGGTAGTGCC
Nanog	GAACGCCTCATCAATGCCTGCA	GAATCAGGGCTGCCTTGAAGAG
Esrrb	GACATTGCCTCTGGCTACCACT	ACTTGCGCCTCCGTTTGGTGAT
Stat3	AGGAGTCTAACAACGGCAGCCT	AGGAGTCTAACAACGGCAGCCT
Kiss1	TAAGACCCGGGTAGCCATCA	ACAGACAGTTGATGGCCCTG
Venus	CAAGATCCGCCACAACATCG	CTCAGGTAGTGGTTGTCTGGG
Gtl2	AGTGCCTTGTAATCGCCCG	CACCTACTGGGTGCTCACTG
Rian	TGTCACGGTCAGCTCTGTTC	ACCAAGGTGTACGCAACGAT
Mirg	TCGCTTACGACAACCGACAA	GGGTGAGAAGTTGGGGACTC
Dlk1	AGCACCTATGGGGCTGAATG	CACTTGTCACAGAGGGGACC
Dio3	TGCGTATCAGACGACAACCGTC	TGGAAGCCATCAGGTCGGACAA
Rtl1/Rtl1as	AGGCTATCAACGAAGGTCGC	TTCACCCGCAGCTCATCATT
Myf5	AAGGCTCCTGTATCCCCTCAC	TGACCTTCTTCAGGCGTCTAC
PCR		
Rtl1/Rtl1as [71] (primer-specific)	GAGAGTGGACCCCTACCACA	GGCAAACCTCTCATCCATGT
Genotyping A	TCTTTCTCAGTCAAAGGAAGTT GCCAATCA	TCACCATACCCCAAAGGTCACT CTGGTATC
Genotyping B	CCCTGACCAAGTCAAATCCA	GTTGCTACGCCTGAATAAGTG
Genotyping C	TTGACTCATGGTAGCCAGTTGGG	AGCGGCCGCAAATTTATTAGAGC

Transgene Copy Number Analysis by qRT-PCR

To assess the relative concentration of *Venus* compared to *Kiss1* in individual transfected colonies, standard curves for each gene were plotted. By rearranging the formula for each of the standard curves, it was possible to calculate the DNA starting quantity as $x=10^{(C_t-b)/m}$. Assuming that 2 copies of *Kiss1* are present it is then possible to predict the copy number of *Venus* by calculating the ratio of $x_{Venus} : x_{Kiss1}$.

2.2.6 Gel Electrophoresis

Agarose (Sigma) was melted at the desired concentration (usually 1-2%, for fragments smaller than 100bp a 4% gel was used) in 0.5X TBE buffer (1X TBE: 89 mM Tris, 89 mM boric acid, 2 mM EDTA in H_2O). 1X Sybr Safe nucleic acid stain (Life Technologies) was added and the gel was left to solidify in a casting tray. The DNA was mixed with loading dye and loaded into the gel. Electrophoresis was performed at 130-140 V in 0.5X TBE buffer. Bands were visualised by UV transillumination.

2.2.7 Plasmid Preparation

Transformation of Bacteria

Plasmids were transformed into One Shot TOP10 Chemically Competent *E. coli* bacteria (Invitrogen). 50 ul of bacteria cells were incubated on ice with 0.5 ul plasmid DNA for 30 min, heat shocked at 42°C for 30 sec and incubated on ice again for 2 min. 250 ul pre-warmed SOC medium was added to the cells before shaking them at 37°C for 1 h. They were then spread on ampicillin containing terrific-broth agar plates and incubated at 37°C over night. Terrific-broth agar plates were prepared as described by Tartof and Hobbs [72]. Ampicillin was added at a concentration of 50 ug/ul.

Plasmid Purification

Single bacterial colonies were picked and grown in 50 ug/ul ampicillin containing terrific broth over night in a shaking incubator at 37°C. Plasmids were isolated using the Purelink HiPure Plasmid Miniprep Kit (Invitrogen) according to the manufacturer's instructions.

2.2.8 RNA Sequencing

Sample Preparation

RNA was extracted from three replicates of all eight cell lines (ESC(Ctrl)1-4 and ESC(Comp)1-4). Cells were passaged twice after thawing before RNA extraction.

Library Preparation and Sequencing

Library preparation was performed using the Illumina TruSeqHT kit according to the manufacturer's protocol. Single-end sequencing was performed and each sample was sequenced in four lanes on a NextSeq500 (Illumina) instrument in the same High-Output 75 cycle sequencing run. Sequencing depth was approximately 10M reads per sample. Library preparation and

sequencing was performed by the Cambridge Genomics Service (Department of Pathology, University of Cambridge, UK).

Data Analysis

Mapping of sequence reads and statistical analysis was performed using the free bioinformatics platform Galaxy (www.usegalaxy.org/).

After concatenating the sequencing output files from the four lanes for each sample, the quality of sequenced reads was assessed with FastQC (www.bioinformatics.babraham.ac.uk/projects/fastqc/) to analyse basic statistics such as per base sequence quality, per sequence quality scores, per base sequence content, sequence length distribution, Kmer content and adapter content. Reads were trimmed by 8 and 4 bases at the start and end of each read respectively to improve quality. TopHat2 [73] Version 0.9 with default settings was used to align reads to the mouse genome(GRCm38/mm10). Reads that mapped uniquely were used for downstream analysis. htseq-count [74] Version 0.6.1galaxy1 was applied to count the number of reads mapping to each gene. Intersection-strict mode was chosen to exclude reads that overlap more than one gene or reads that do not intersect a given gene with each read position. The Gencode release M7 was used to annotate the genome. Differential expression analysis was subsequently performed with DESeq2 [75] Version 2.1.8.3.

2.2.9 DNA Methylation Analysis

Sample and Library Preparation

gDNA was extracted from three replicates of all 15 cell lines (ESC(Ctrl)1-8 and ESC(Comp)1-7). Cells were passaged twice after thawing before gDNA extraction. 1 ug of DNA was bisulfite converted using the EZ DNA Methylation Kit (Zymo Research, D5020), according to the manufacturer's instructions. Samples were eluted in 66 ul elution buffer to get a concentration of 15 ng/ul. For each sample, each imprinted region was amplified in a separate PCR reaction (Table 2.6) and then pooled. 1.5X AMPure beads were used for PCR clean up. In a second PCR reaction, barcoded adapters were attached to the samples for sequencing (Table 2.7). 1X AMPure beads were used for PCR clean up. Libraries were sequenced as paired-end 150bp reads on a MiSeq Instrument. 10% PhIX spike-in was included due to the low-complexity libraries.

Data Analysis

Data analysis was performed by Dr Eckersley-Maslin at the Babraham Institute Cambridge. In brief, poor quality base calls and adapters were trimmed. Using Bismark (www.bioinformatics.babraham.ac.uk/projects/bismark/) the sequences were subsequently aligned to the mouse genome (GRCm38/mm10) and deduplication was performed in order to remove PCR bias. Methylation calls were imported into Seqmonk (www.bioinformatics.babraham.ac.uk/projects/seqmonk/) and probes were defined for every CpG that was included in the amplicons. Percentage methylation was calculated for each of the probes that had at least 30 reads. The overall percentage methylation for each imprinted region was calculated as the mean of individual CpG sites.

Table 2.6 PCR amplifying imprint regions

Reaction mix (per reaction)	Step	Thermal cycling conditions
4 ul 2x KAPA HiFi Uracil+ 1.2 ul Primer Pool (sequences are confidential) 0.8 ul DNA/RNA-free H_2O 2 ul Bisulphite converted DNA	Initial Denaturation	95°C for 5 min
	30 cycles of:	30 cycles of:
	DNA denaturation	98°C for 20 sec
	Primer annealing	55-62°C for 15 sec
	Primer extension	72°C for 1 min
Final extension step	72°C for 10 min	

Table 2.7 PCR attaching barcoded adapters to samples for sequencing

Reaction mix (per reaction)	Step	Thermal cycling conditions
25 ul 2x KAPA HiFi Uracil+ 1 ul PE1.0 1 ul iTAG 3 ul DNA/RNA-free H_2O 20 ul DNA	Initial Denaturation	98°C for 45 sec
	5 cycles of:	5 cycles of:
	DNA denaturation	98°C for 15 sec
	Primer annealing	65°C for 30 sec
	Primer extension	72°C for 30 sec
Final extension step	72°C for 5 min	

2.2.10 Phenotypic Screening

Cell Surface Selection

1×10^7 cells were used per selection and blocked in ES media for 1 h at room temperature on a rotary mixer at 25 rpm. Phage library aliquots containing 10^{12} phage (libraries used were scFv Bone Marrow Vaughan (BMV) and combined spleen (CS) libraries, available from Medimmune, Cambridge, UK) were also blocked in ES media for 1 h at room temperature. The cells and phage were then incubated together at room temperature on a rotary mixer at 20 rpm for 1 h, allowing the phage to bind to the cells.

For deselection, the supernatant of the ESC(Ctrl) cells was harvested and incubated with the ESC(Comp) cells for 1 h at room temperature on a rotary mixer at 20 rpm.

Unbound phage was washed off by repeated rounds of PBS washes of the cells. To elute the bound phage, the cells were incubated with trypsin for 20 minutes. After inactivation of trypsin, the cells were again spun down and the supernatant containing the eluted phage was kept.

Selection Rescue

E. coli TG1 cells were infected with the eluted phage and grown up over night on 2TYAG bioassay plates (2TY agar plates with 100 ug/ml ampicillin and 2% glucose) at 30°C. The bacteria were scraped off the plates using 2:1 2TY:Glycerol media and grown up in 2TYAG media to mid-log phase at 37°C, 280rpm. After addition of M13KO7trp helper phage the cells were grown for a further 1 h at 37°C, 150 rpm. For over night cultures at 25°C, 280 rpm, growth media was changed to 2TYAK medium (2TY with 100 ug/ml ampicillin and 50 ug/ml kanamycin).

The overnight cultures were spun down. Supernatant containing the rescued phage was used in the next round of cell surface selection. Three rounds of selection and rescue were performed. The outputs of rounds 2 and 3 were sequenced and their quality was analysed using BLAZE.

Phage to IgG Reformatting

All primer sequences used here are confidential. Phage to IgG reformatting was performed in two cloning steps:

Glycerol stocks from the round 2 outputs were used to prepare plasmid DNA of the pCANTAB6 population, using the QIAprep Spin Miniprep Kit (QIAGEN, 27104). The pCANTAB6-scFv construct was linearised by emulsion PCR, using the Micellula DNA Emulsion & Purification Kit (Roboklon, 3600), following the manufacturer's instructions. Lambda and Kappa populations were prepared in two separate PCR reactions. Table 2.8 outlines the reaction mix and

Table 2.8 **IgG conversion - Emulsion PCR for intermediate pool**

Lambda PCR	ul for 1 population
2X HiFi PCR Premix	111
L1 FW1 (10 uM)	9
L2 FW1 (10 uM)	9
L3 FW1 (10 uM)	9
Lpool FW1 (10 uM)	9
Hpool FW4 (10 uM)	9
Nuclease free H_2O	33
BSA (0.2 mg/ml)	11
DNA template (5 ng/ul)	20 (add after removing 20 ul for no template ctrl)
Kappa PCR	ul for 1 population
2X HiFi PCR Premix	111
K1 FW1 (10 uM)	9
Kpool FW1 (10 uM)	9
Hpool FW4 (10 uM)	9
Nuclease free H_2O	33
BSA (0.2 mg/ml)	11
DNA template (5 ng/ul)	20 (add after removing 20 ul for no template ctrl)

Step	Thermal cycling conditions
Initial Denaturation	94°C for 2 min
28 cycles of:	28 cycles of:
DNA denaturation	94°C for 10 sec
Primer annealing	60°C for 30 sec
Primer extension	72°C for 5 min
Final extension step	68°C for 3 min

thermal cycling conditions. After breaking the emulsions and PCR purification, DNA was eluted in 40 μ l of DNA/RNA-free H_2O .

4.5 μ l of 10X CutSmart Buffer (NEB) and 2 μ l of Dpn1 (10 U/ μ l, NEB) were added to each sample and incubated at 37°C for 30 min to remove the pCANTAB6 template. The amplified and digested populations were run on a gel and the 5kb DNA band was purified using the QIAquick Gel Extraction Kit (QIAGEN). Purified DNA was eluted in 30 μ l of DNA/RNA-free H_2O .

The pmDV.v2 donor fragment was prepared as outlined in Table 2.9. For the In-Fusion (Clontech) cloning reactions were set up as described in the manufacturer's protocol. 100 ng were used each of the pmDV.v2 donor and pCANTAB6-scFv vector. These were incubated at 50°C for 15 min. 2.5 μ l of the In-Fusion reaction was transformed into 100 μ l aliquots of Stellar competent cells (Novagen 636763) and incubated on ice for 30 min. The cells were then heat shocked in a water bath at 42°C for exactly 60 sec and again incubated on ice for 2 min. After addition of 400 μ l SOC medium, the cells were incubated stationary at 30°C for 10 min. Transformed cells were plated on 2TYAG bioassay plates at 30°C for 24 h. The bacteria were then scraped off the plates using 2:1 2TY:Glycerol media and stored in aliquots at -80°C.

Table 2.9 IgG conversion - Preparation of the pmDV.v2 donor fragment

Components	μl per reaction
2X HiFi PCR Premix	50
MM018 Forward primer (10 μ M)	2
pmDv.v2 REV reverse primer (10 μ M)	2
Nuclease free H_2O	44
pmDV.v2 plasmid (10 ng/ μ l)	2

Step	Thermal cycling conditions
Initial Denaturation	94°C for 2 min
25 cycles of:	25 cycles of:
DNA denaturation	98°C for 10 sec
Primer annealing	60°C for 30 sec
Primer extension	72°C for 3 min
Final extension step	68°C for 5 min

Table 2.10 IgG conversion - Emulsion PCR for final pool

PCR Components	ul for 1 population
2X HiFi PCR Premix	111
H Fwd 1 (10 uM)	9
H Fwd 2 (10 uM)	9
VL-Rev or VK-Ref (10 uM)	9
Nuclease free H_2O	51
BSA (0.2 mg/ml)	11
DNA template (5 ng/ul)	20 (add after removing 20 ul for no template ctrl)

Step	Thermal cycling conditions
Initial Denaturation	94°C for 2 min
25 cycles of:	25 cycles of:
DNA denaturation	98°C for 10 sec
Primer annealing	62°C for 30 sec
Primer extension	72°C for 3 min
Final extension step	68°C for 5 min

Glycerol stocks of these intermediate pools were used to prepare plasmid DNA, using a QIAprep Spin Miniprep Kit as before. The antibody cassette of the plasmids was amplified by emulsion PCR. Table 2.10 outlines the reaction mix and thermal cycling conditions. The lambda and kappa PCR products were run on a gel and the DNA from the 3kb bands was purified using the QIAquick Gel Extraction Kit. Purified DNA was eluted in 30 ul of DNA/RNA-free H_2O .

The pmIgG κ _v5 or pmIgG λ _v5 vectors were prepared as outlined in Table 2.11. In-Fusion reactions with linearised pmIgG κ _v5 or pmIgG λ _v5 were set up for kappa and lambda samples respectively as described in the manufacturer's protocol. 100 ng were used each of the pmIgG linearised vectors and the respective antibody cassette insert. These were incubated at 50°C for 15 min. 2.5 ul of the In-Fusion reaction was transformed into 100 ul aliquots of Stellar competent cells (Novagen 636763) in duplicates and incubated on ice for 30 min. The cells were then heat shocked in a water bath at 42°C for exactly 60 sec and again incubated on ice for 2 min. After addition of 400 ul 2TY medium, the cells were incubated stationary at 37°C for 10 min. Transformed cells were plated on 2TYAG bioassay plates at 37°C over night. Bacteria were then scraped off the plates using 2:1 2TY:Glycerol media and stored in aliquots at -80°C. Individual colonies for HT IgG screening were picked into 2TYAG in 96 well plates and grown over night at 37°C, 120 rpm. The intermediate and final expression pools were sequenced and

their quality was analysed using BLAZE. HT IgG expression was performed by the scFV HTE team at Medimmune as described by Xiao et al [76].

Table 2.11 IgG conversion - Preparation of the IgG vectors

Components	ul per reaction	Incubation conditions
Nuclease free H_2O	to 200 ul	37°C for 2 h
10X NEB Buffer 2	20	
100X BSA	2	
pmIgG κ _v5 maxiprep	5 ug	
BsrGI (NEB)	2	
Add BsiW1 (NEB)	2	55°C for 2 h

Components	ul per reaction	Incubation conditions
Nuclease free H_2O	to 200 ul	37°C for 4 h
10X NEB Buffer 2	20	
100X BSA	2	
pmIgG λ _v5 maxiprep	5 ug	
BsrGI (NEB)	2	
KasI (NEB)	4	

Next Generation Sequencing - Sample Preparation

To prepare samples for next generation sequencing (NGS) of the phage outputs, a miniprep of the glycerol stocks was performed using the QIAprep Spin Miniprep Kit (QIAGEN 27104). The DNA was amplified by PCR (components of the reaction mix and thermal cycling conditions are outlined in Table 2.12) and run on a 1% agarose gel. After gel extraction with the QIAquick gel extraction kit (QIAGEN 28704) the DNA was further purified using AMPure beads (A63882). The samples were submitted for sequencing to the DNA Sequencing Facility in the Department of Biochemistry, University of Cambridge. Paired-end sequencing (2x300bp) was performed using the MiSeq system and the MiSeq reagent kit v3 (Illumina).

Next Generation Sequencing - Data analysis

NGS data analysis was performed by Dr Samborsky at the Department of Biochemistry, University of Cambridge. In brief, read quality control and adapter trimming was performed and paired-end sequences were merged to obtain complete scFv library sequences. These

were mapped against a scFv processing pipeline that can separate reads into CDR and non-CDR regions and automatically annotates FW1-4 and CDR1-3. Out of frame sequences were excluded. In order to perform CDR frequency analysis, the CDR3 heavy and light chains were concatenated, the number of occurrences of each was counted and the frequency was calculated for each sample.

Table 2.12 RT-PCR reaction mix and thermal cycling conditions for NGS samples

Reaction mix (per reaction)	Step	Thermal cycling conditions
25 ul Q5 Hotstart 2X Master Mix (NEB) 1 ul 10 uM LSeq Primer 1 ul 10 uM HLink Primer 18 ul DNA/RNA-free H_2O 5 ul DNA template (5 ng/ul)	Initial Denaturation	98°C for 30 sec
	20 cycles of:	20 cycles of:
	DNA denaturation	98°C for 10 sec
	Primer annealing	55°C for 30 sec
	Primer extension	72°C for 1 min
Final extension step	72°C for 2 min	

2.3 Animal Work and Histology

2.3.1 Blastocyst Injections

Animals

Timed matings of the inbred C57/BL6 mouse strain were set up to give blastocysts at 3.5 dpc (days post coitum). One day later, 2.5 dpc pseudopregnant F1 females (C57/Bl6 x CBA) were derived by mating with vasectomised males.

Blastocyst Flushing

Pregnant 3.5 dpc C57/BL6 females were killed by cervical dislocation and their uterine horns were removed in order to flush blastocysts. This was done using a 27 gauge needle that was inserted into one end of each uterine horn and flushing it with M2 media (Millipore). Blastocysts were then transferred into ESC media and either injected immediately or kept for a few hours in a 37°C humidified incubator.

ES Cell Injections and Embryo Transfers

Blastocyst injections and embryo transfers were carried out by Professor W.H. Colledge. ES cell medium was replaced on the cells 1-2 h prior to the injections. The cells were then trypsinised and resuspended in a small volume of ESC media. Using a Leitz Micromanipulator, 10-12 ES cells were injected into each blastocyst. After 2-3 h of recovery to give them time to re-expand, the injected blastocysts were transferred into 2.5 dpc pseudopregnant F1 female recipients using a NSET (non surgical embryo transfer) device.

2.3.2 Analysis of Chimaeras

In the first instance, pregnant females were weighed at gestational days e10.5, e13.5, e15.5. and e17.5 to monitor pregnancy status. Chimaerism of born pups was then assessed by coat colour. Later, pregnant females were sacrificed by dislocation of the neck at two time points during gestation (e13.5 and e17.5) in order to define the embryonic phenotype. The embryos were imaged using both the Zeiss SteREO Discover.V8 Stereomicroscope and the Leica M205 FA fluorescence microscope with a DFC7000T camera .

2.3.3 Paraffin Embedding and Sectioning

Embryonic tissues were fixed in 4% PFA for 5 h and were then transferred into 70% EtOH over night. The tissues were then transferred into fresh 100% EtOH for 4 h, before exchanging EtOH for 50% HistoClear in EtOH for 1 h and subsequently 100% HistoClear overnight. The tissues were transferred to 54-58°C warm paraffin wax twice for 3.5-4 h, before being embedded. Sections were cut to a thickness of 7-8 μm using a microtome and dried at 37°C for 24 h.

2.3.4 Haematoxylin and Eosin Staining of Embryo Sections

The tissue sections were placed in Xylene for two times 5 min each to remove paraffin wax. They were then rehydrated by progressively moving the slides through series of ethanol (100% EtOH, 100% EtOH, 90% EtOH, 70% EtOH, 50% EtOH, 30% EtOH) and PBS for 3 min each. The slides were rinsed with water before staining with haematoxylin for 8 min. After a quick wash in cold ddH₂O they were placed into acid alcohol (70% EtOH with 1% HCl) for a few seconds to force the stain to enter the nucleus. The tissue slides were extensively rinsed with cold tap water for 6-10 min for blueing to occur. They were then stained for 45 sec with Eosin and quickly washed with tap water twice. To dehydrate tissue sections, they were again progressively moved through series of ethanol (50% EtOH, 70% EtOH, 90% EtOH, 100%

EtOH, 100% EtOH) for 1 min each. The slides were then twice placed in Xylene for 2 min each, mounted with coverslips using DPX mounting medium and let to dry and set over night.

Chapter 3

Initial Characterisation of the ES Cell Clones

3.1 Introduction

This chapter provides an overview of the initial characterisation and quality control experiments that were performed on the ES cells used for the experiments described in this thesis. We show that all ES cell lines have a normal karyotype and retain the main characteristics of pluripotent stem cells, that is self renewal and the ability to differentiate into cell types of all three germ layers.

3.1.1 Effects of Prolonged Culture on Karyotype and Pluripotency

During gene targeting, ES cells are exposed to subcloning and repeated passaging. It is a well established fact that long term culture can lead to chromosome instability and abnormal chromosome numbers [77–79]. The most common chromosome number mutations appear to be trisomy 8 or trisomy 11 [77]. A normal karyotype is not only important for the maintenance and differentiation of cells *in vitro* [80] but also for *in vivo* development. The best known example for this is probably trisomy 21 which leads to Down syndrome in humans. Similarly, other karyotypical abnormalities can also cause more or less severe developmental defects, often leading to embryonic lethality as numerous studies in both mouse and human have shown [81–83].

Prolonged ES cell culture can not only lead to changes in the cells' karyotype, but it can also affect their pluripotency [84, 85]. This is thought to be due to the accumulation of epigenetic changes [86].

3.2 Results

3.2.1 Karyotyping

Karyotyping was performed on the ESC(Comp) clones by arresting the cells in metaphase using colcemid. Metaphase is the stage of the cell cycle when the chromosomes are in their most condensed state and hence easiest to count. Colcemid depolymerises microtubules which in turn disrupts spindle formation. Spindle fibers are essential for cells to progress from metaphase to anaphase. Disruption of the spindle will, however, not stop cells progressing from earlier stages of the cell cycle to metaphase and colcemid treatment therefore enriches the metaphase cell population [87]. Giemsa staining was then used to visualise the chromosomes and their numbers could be counted.

Typically, in the mouse a karyotype is assumed to be normal if at least 70% of cells have 40 chromosomes. Figure 3.1B is an example of how a typical chromosome spread of a single cell looks like. From these, the chromosome number was counted in 15 cells for each of the compromised cell lines. All of them had 80% or more cells with exactly 40 chromosomes (Figure 3.1A).

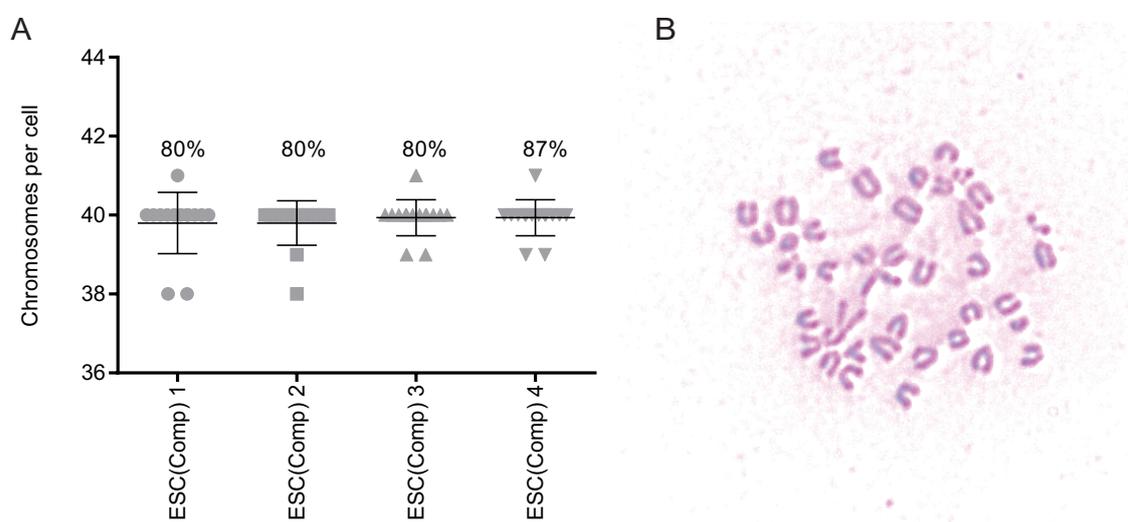


Fig. 3.1 **Chromosome counting** (A) Number of chromosomes per cell for the four ESC(Comp) clones (n=15). Percentages indicate the proportion of cells with exactly 40 chromosomes. (B) Image of a typical chromosome spread of a single cell.

3.2.2 Expression of Pluripotency Markers

The ES cell clones were analysed to see whether the cell lines still fulfilled the basic characteristics of pluripotent stem cells - self renewal and the ability to differentiate into all other cell types. Since the cell lines have been kept in culture for a seemingly infinite number of passages, they clearly still retain the ability to self renew. Another good initial test for pluripotency is to look at the expression of some of the core pluripotency genes. We chose to analyse the expression of *Nanog*, *Oct4*, *Sox2*, *Klf4*, *Esrrb* and *Stat3*. *Nanog*, *Oct4* and *Sox2* are at the heart of the pluripotency gene regulatory network and are absolutely essential for pluripotency. Several studies have shown that ES cells that do not express these genes lose their ability to self renew and will spontaneously differentiate [19, 25, 28, 31]. OCT4 and SOX2 form a heterodimer which, by binding to the *Nanog* promoter, regulates the expression of *Nanog* and a number of downstream genes that are important for the maintenance of pluripotency [88]. KLF4 is one of the original four Yamanaka factors used to reprogram mouse fibroblast cells into induced pluripotent stem cells (iPSCs) [89]. Additionally, KLF4 is a key regulator of *Nanog* expression. It can also bind to the *Nanog* promoter and thereby assist OCT4 and SOX2 in controlling the expression of *Nanog* [90]. In contrast, *Esrrb* is a target gene of NANOG and overexpression of *Esrrb* can rescue *Nanog*-null ES cells [91]. Finally, STAT3 plays an important role in ES cell self renewal and pluripotency through the LIF/JAK/STAT3 signalling pathway [92]. LIF is an essential addition to ES cell media, but it requires expression of *Stat3* in order to have an effect on pluripotency.

Figure 3.2 shows that all the ESC(Comp) clones expressed these genes at levels comparable to the controls. The only exception was the expression of *Nanog* in ESC(Comp)1 and ESC(Comp)3 as well as the control clone ESC(Ctrl)4 (One-way ANOVA, $p < 0.05$). However, they were not completely silenced and it is questionable whether such a small downregulation will have an effect on the cells' self renewal and differentiation potential.

3.2.3 Differentiation Assay

To confirm the second characteristic of pluripotency - the ability to differentiate into cells of all the three germ layers - we performed a differentiation assay. To initiate differentiation we formed embryoid bodies (EBs) in hanging drops. EBs are three dimensional aggregates of ES cells that have been shown in the past to recapitulate a number of differentiation processes that occur in the developing early embryo. In general, EBs can contain early progenitors of cells of the three germ layers (mesoderm, endoderm and ectoderm). If left to spontaneously differentiate, it is possible to eventually find a number of different specialised cell types, including cardiac muscle, skeletal muscle, adipogenic cells, neuronal cells and hepatic cells.

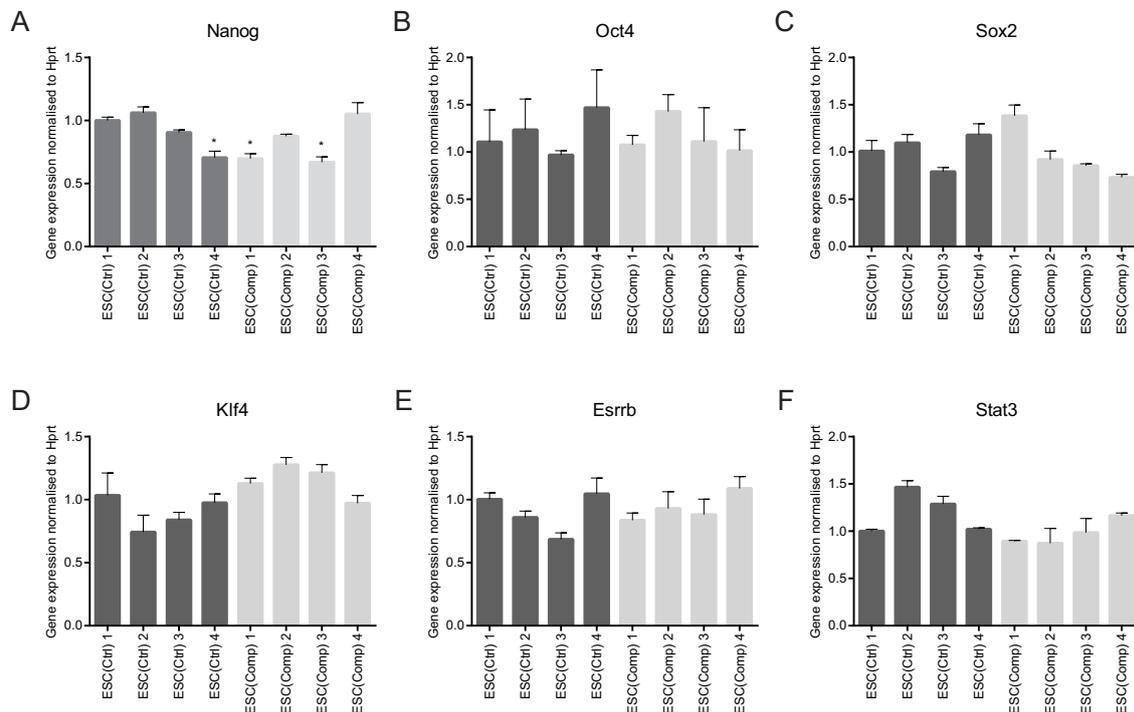


Fig. 3.2 Expression of pluripotency markers (A) *Nanog* (B) *Oct4* (C) *Sox2* (D) *Klf4* (E) *Esrrb* (F) *Stat3*. Gene expression was normalised to *Hprt* and results were normalised to the values of ESC(Ctrl)1. Error bars represent standard error of the mean. Stars represent significant differences as calculated by ANOVA ($p < 0.05$). $n = 2$ (biological replicates).

However, some cell types will arise more frequently than others, which is why the EBs were cultured in defined media that will induce differentiation down specific pathways. This can be controlled by adding growth factors or growing the cells on specific substrates. Protocols for this can be found in the literature [55, 58, 62] and the basic methodology is also outlined in Figure 3.3A.

Mesoderm - Cardiomyocyte Differentiation

Cardiac muscle is one of the cell types that frequently develops when leaving EBs to spontaneously differentiate. Therefore, for mesoderm differentiation the EBs were left to differentiate without the addition of growth factors. After EB formation in hanging drops, they were kept in suspension culture for an additional three days to increase in size, after which they were cultured on gelatin-coated plates. Only two days later, we were able to observe the first beating cell clusters, which is typical for cardiomyocyte differentiation (data not shown). Staining for

α SMA (alpha smooth muscle actin), which is an early mesoderm marker [93], was positive in cells derived from all of our eight cell lines (Figure 3.3B).

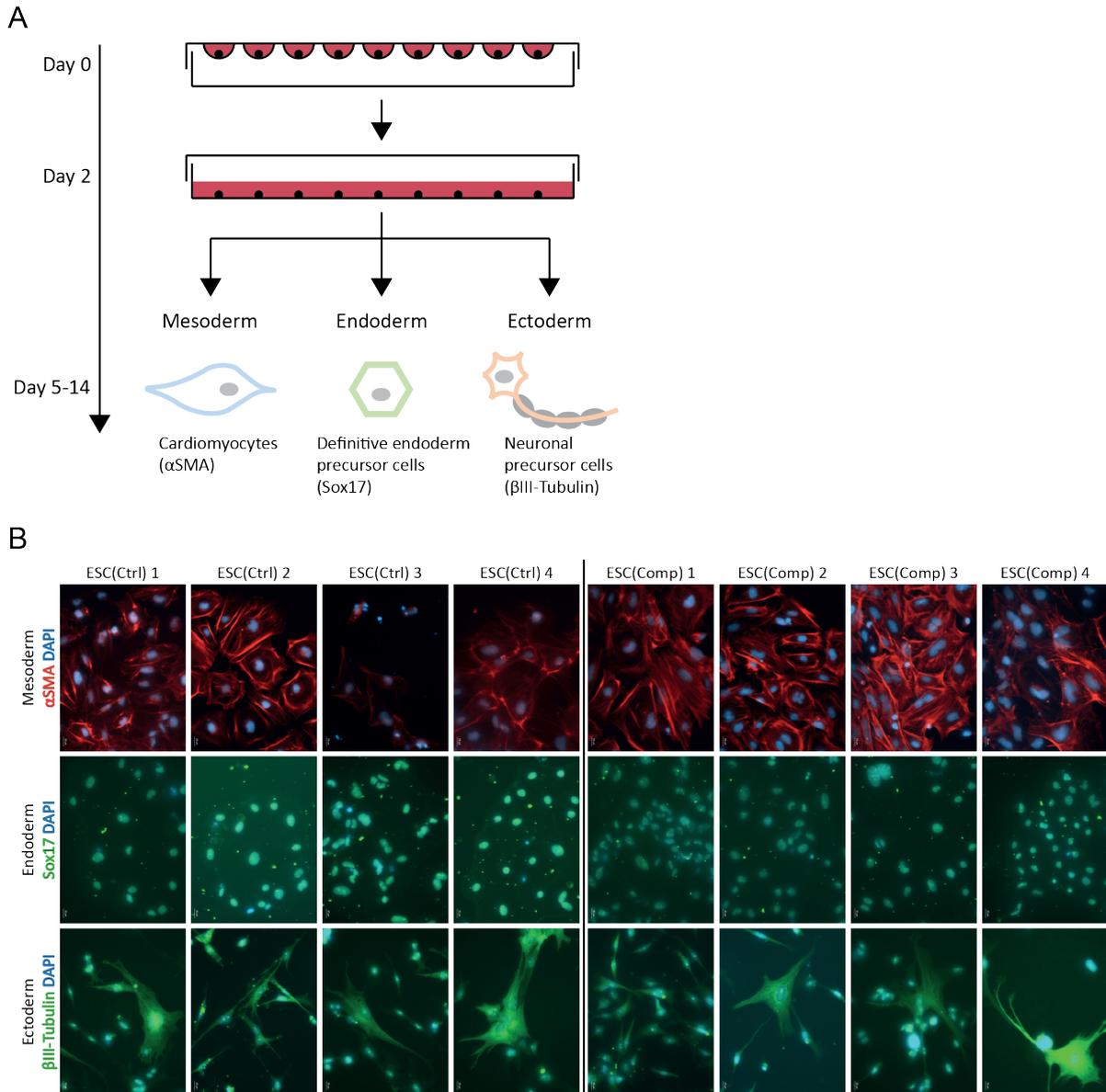


Fig. 3.3 Differentiation assay (A) Experimental set up for the differentiation assay (B) Representative images of differentiated cells from each cell line. α SMA in red stains cardiomyocytes, SOX17 in green stains definitive endoderm cells and β III-TUBULIN in green stains neuronal cells.

Definitive Endoderm Differentiation

Spontaneous endoderm differentiation in EBs occurs much less frequently. Therefore, after initiating differentiation, EBs were dissociated and plated on Poly-L-ornithine/Laminin-coated plates in differentiation media containing Nicotinamide. Nicotinamide is a form of vitamin B3 and has been shown to be important for the differentiation and proliferation of insulin-producing pancreatic progenitors. More recently, Cho et al. [94] suggest that Nicotinamide may act by sustaining *Pdx1* (Pancreatic and duodenal homeobox 1) expression, a transcription factor that is important for pancreatic development.

Nine days after replating, cells were stained for SOX17 (SRY-Box 17), a transcription factor that is a marker for definitive endoderm [95]. As Figure 3.3B shows, all cell lines have the potential to differentiate into definitive endoderm cells.

Ectoderm - Neuronal Differentiation

Similar to endoderm differentiation, EBs do not often spontaneously differentiate into neurons or neuronal progenitor cells. After initiating differentiation we again replated EBs onto Poly-L-ornithine/Laminin-coated plates but this time added EGF (Epidermal growth factor) and bFGF (basic fibroblast growth factor) to initiate neuronal differentiation. EGF and bFGF promote the proliferation, survival and self-renewal of neural stem and progenitor cells and are routinely used for the culture of neuronal cell types and for the differentiation of pluripotent stem cells into neurons [96, 97].

After six days of exposure to these growth factors, the expression of β III-TUBULIN was analysed. β III-TUBULIN is a microtubule element that is almost exclusively expressed in neurons. As can be seen in Figure 3.3B, again all eight of our cell lines were able to differentiate into neuronal cells.

3.3 Discussion

We have shown that all of the ESC(Comp) cell lines that were used for experiments presented in this thesis have a normal karyotype with >80% of cells having exactly 40 chromosomes and that all cell lines fulfil the basic characteristics of pluripotency - self renewal and the ability to differentiate into cells of the three embryonic germ layers. Additionally, all cell lines were screened for pathogens such as mycoplasma (data not shown). We are therefore very confident that the lethality and developmental defects found in chimaeric embryos resulting from blastocyst injections with these ES cell clones are due to other underlying genetic or epigenetic mutations.

3.3.1 Karyotype Analysis

The karyotyping that was performed here confirms that the cells have the right number of chromosomes. However, with this method it is not possible to identify any structural chromosomal changes like deletions, insertions, duplications, inversions or translocations. The consequences of these can be incredibly detrimental to embryonic development, often causing lethality or at least life-long disabilities [98]. Better methods to detect these changes would be extended G-banding, fluorescence *in situ* hybridisation (FISH) [99] or chromosomal microarray studies [100]. Furthermore, by simple chromosome counting we cannot rule out that there may be a trisomy of one chromosome in addition to a monosomy of another chromosome, which would also result in a total of 40 chromosomes. However, all of these occurrences are very rare and the analysis of this was beyond the scope of this project.

3.3.2 Pluripotency and Differentiation Potential

Although in general all pluripotency genes that we looked at were expressed at comparable levels between the ESC(Ctrl) and ESC(Comp) cell lines, there was a significant difference in *Nanog* expression with ESC(Comp)1, ESC(Comp)3 and ESC(Ctrl)4 only expressing it to around 70% of the level of the other clones. However, this level of silencing is unlikely to cause changes to the cells' pluripotency and their ability to contribute to chimeras. In fact, it may be attributable to transcriptional heterogeneity which is a well known phenomenon in ES cells, particularly if they have been grown in serum/LIF conditions [101, 102]. Both Chambers et al. [103] and Singh et al. [104] demonstrate that there is a small ES cell subpopulation which does not express *Nanog* or expresses it at lower levels. However, this silencing seems to be reversible and some *Nanog*-null ES cells still retain pluripotency and their differentiation potential. Similar results have also been reported for other pluripotency genes [105–108].

It therefore seems very unlikely that the ES cells will be affected by the slight difference in *Nanog* gene expression. Particularly since the differentiation assay showed that all cell lines retain the ability to differentiate into cells of the mesoderm, endoderm and ectoderm lineages. We have, however, only shown that the cells can differentiate into quite early progenitors of each of the three lineages. From this it is not clear whether there may be more specific issues in later stages of development. We have also just performed a qualitative analysis, not a quantitative one. Therefore it remains an open question whether there may be more subtle differences in the efficiency of differentiation or whether some cell lines have a "preferred" lineage that they differentiate into more frequently when left to differentiate spontaneously.

3.3.3 Chapter Conclusion

Overall, we have shown in this chapter that the ES cells used in this thesis are of good quality with a healthy karyotype, normal expression of pluripotency markers and the ability to differentiate into cell types of all three germ layers.

Chapter 4

Blastocyst Injections and Histology

4.1 Introduction

This chapter describes the embryonic lethal phenotype observed in the ESC(Comp) chimaeras in more detail. The aim of this analysis was to define the timepoint at which embryonic lethality occurs and to identify developmental defects that have occurred in these embryos. Initially, all cell lines were labelled fluorescently to identify and visualise the chimaeric embryos. These were then injected into blastocysts and the phenotype of the chimaeras was analysed at e13.5 and e17.5. While no embryonic death was observed at e13.5, we show that there was a significant decrease in embryos surviving until e17.5 and that a great number of embryos showed signs of haemorrhaging.

4.2 Results

4.2.1 Cell Labelling Using the PiggyBac System

Adult chimaeras can usually be distinguished from their wild-type littermates by differences in coat colour. In the embryo this is not possible. Therefore, all eight ES cell clones were labelled fluorescently to identify chimaeric embryos and track ES cell contribution. This was done using the PiggyBac (PB) system. This makes use of transposons which are mobile genetic elements that can transpose from one genetic location to another via a "cut and paste" mechanism [109]. Two parts are needed for this system: the transposon itself and a transposase. In the PB system, these are supplied in two separate vectors. The transposon vector contains the gene of interest flanked with transposon specific inverted terminal repeat sequences (ITRs). The second vector contains the transposase (PBase). The PBase recognises the ITRs, cuts the gene of interest from the vector and randomly moves it to TTAA sites in the chromosome. We labelled the

Table 4.1 Transfection efficiencies

	Electroporation		Nucleofection		Lipofection		-ve ctrl
	+ Venus + PBase	+ Venus - PBase	+ Venus + PBase	+ Venus - PBase	+ Venus + PBase	+ Venus - PBase	- Venus - PBase
Cells plated	3.14x10 ⁶	2x10 ⁶	1.2x10 ⁶	1.4x10 ⁶	3x10 ⁶	3x10 ⁶	5x10 ⁶
Colonies	5280	44	5332	96	1572	16	0
Efficiency	0.17%	0.0022%	0.44%	0.0069%	0.052%	0.00053%	-

Table 4.2 Copy number analysis

Sample	Ratio <i>Venus/Kiss1</i>	Number of <i>Venus</i> copies
NF 1	0.43680	1
NF 2	0.62003	1
NF 3	1.39619	3
NF 4	0.32415	1
NF 5	0.48150	1
NF 1	0.42843	1
NF 2	0.27795	1
EP 1	1.77943	3-4
EP 2	0.21199	1
EP 3	5.69712	11
EP 1	0.15177	1
EP 2	0.79565	1-2
LF 1	32.27507	64-65
LF 2	0.82803	1-2
LF 3	8.08072	16
LF 4	6.75877	13-14
LF 5	0.93101	2
LF 1	26.73544	53-54

transfection methods (electroporation, nucleofection and lipofection) were tested for their transfection efficiencies using the ESC(Ctrl)3 cell line. As Table 4.1 shows, nucleofection (NF) resulted in the best transfection efficiency with 0.44% of cells expressing *Venus* compared to 0.17% and 0.052% after electroporation (EP) and lipofection (LF), respectively. Also note the big difference in transfection efficiency between experiments with and without the PBase. Knowing that the PB system is very efficient and that it randomly integrates the gene of interest into the cells' genome, it is likely that at least a subset of cells has more than one copy of *Venus*

and it was next tested whether the type of transfection method had an effect on this. Higher copy numbers increase the probability of one or more integration sites that might cause changes in the cells' normal functions. Therefore, we wanted to minimise this. To do this, a number of individual ES cell colonies were picked for each of the different transfection methods and gDNA was extracted. Using qPCR, standard curves were plotted for *Venus* (Figure 4.1D) and *Kiss1* (Figure 4.1E) as a control (*Kiss1* was chosen randomly). We then calculated the DNA starting quantities using the qPCR results for *Venus* and *Kiss1* for each individual sample and rearranging the equation of the standard curve slope to $x=10^{(Ct-b)/m}$. Assuming that two copies of *Kiss1* are present it was then possible to predict the copy number of *Venus* by calculating the ratio of x_{Venus} to x_{Kiss1} . A ratio of approximately 0.5 implies that there is only 1 copy of *Venus*. The data for a number of ES cell colonies are summarised in Table 4.2 and show that most of the nucleofected cells only have one copy of *Venus*, whereas electroporation and especially lipofection had the tendency to introduce multiple copies of *Venus* into the genome.

Fluorescent Labelling of Cells

We therefore decided to use nucleofection as the preferred method of transfection for the eight cell lines (ESC(Ctrl)1-4 and ESC(Comp)1-4) used for the blastocyst injections. After nine days of selection in hygromycin (pPBCAG-H2BtdTomato-IH) or puromycin (pPBCAG-Venus-IP) the cells were pooled and their fluorescence was checked. Figure 4.2A shows that we managed to fluorescently label all eight cell lines. Figure 4.2B also shows fluorescent images of e13.5 chimaeric embryos, providing further prove that the cell labelling was successful.

4.2.2 Phenotypes of e13.5 and e17.5 Chimaeras

We used the labelled cells for blastocyst injections to investigate the exact phenotype of the chimaeras. All of the injections and embryo transfers into pseudopregnant recipient females were carried out by Professor W.H. Colledge. The embryos were analysed at gestational days e13.5 and e17.5. The graphs in Figure 4.3A,B show the breakdown of results for each cell line at the two time points. Overall, we observed four broad phenotypes that are also shown in Figure 4.3G: live and healthy/normal (i, ii), live with haemorrhaging (iii, iv), dead and not resorbed yet (all of these had severe haemorrhaging) (v, vi), dead and fully resorbed (vii, viii, ix). Interestingly, in most e17.5 embryos that were alive with haemorrhaging, the haemorrhages predominantly occurred at the limbs and the tail. Additionally, some embryos also showed signs of edema (Figure 4.3G(v)), which, similar to haemorrhaging, is usually caused by vascular defects.

At e13.5, no dead (not resorbed) embryos were found. At e17.5, however, we started seeing late

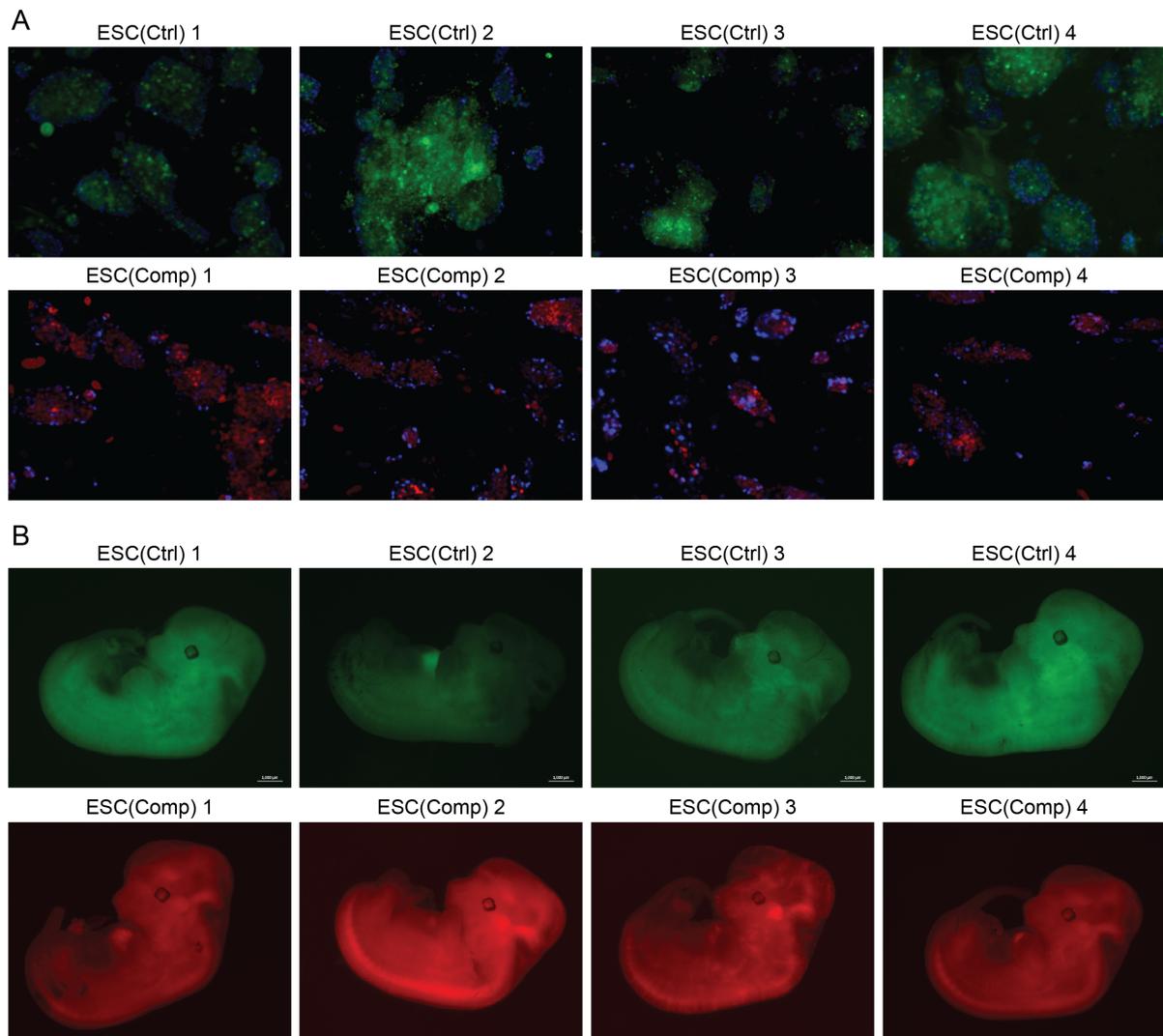


Fig. 4.2 Fluorescently labelled cells and embryos (A) Images of all eight cell lines after labelling with *Venus* or *tdTomato*. (B) Images of e13.5 chimaeras for each of the cell lines.

gestational embryonic lethality. Pooling all the data and comparing the two groups ESC(Ctrl) and ESC(Comp) with each other in terms of the proportion of embryos that were still alive at the two stages, we can see that at e13.5 there was no significant difference between the two groups, no matter whether embryos with haemorrhaging were included or not (Figure 4.3C,E). However, looking at just the healthy live embryos, the p-value is 0.09, suggesting that there is a trend, although it is not quite significant yet. At e17.5, there was a clear significant difference between the proportion of embryos that were still alive between the ESC(Ctrl) and ESC(Comp) groups, both when looking at the healthy embryos and when including haemorrhaged ones (Figure 4.3C,E).

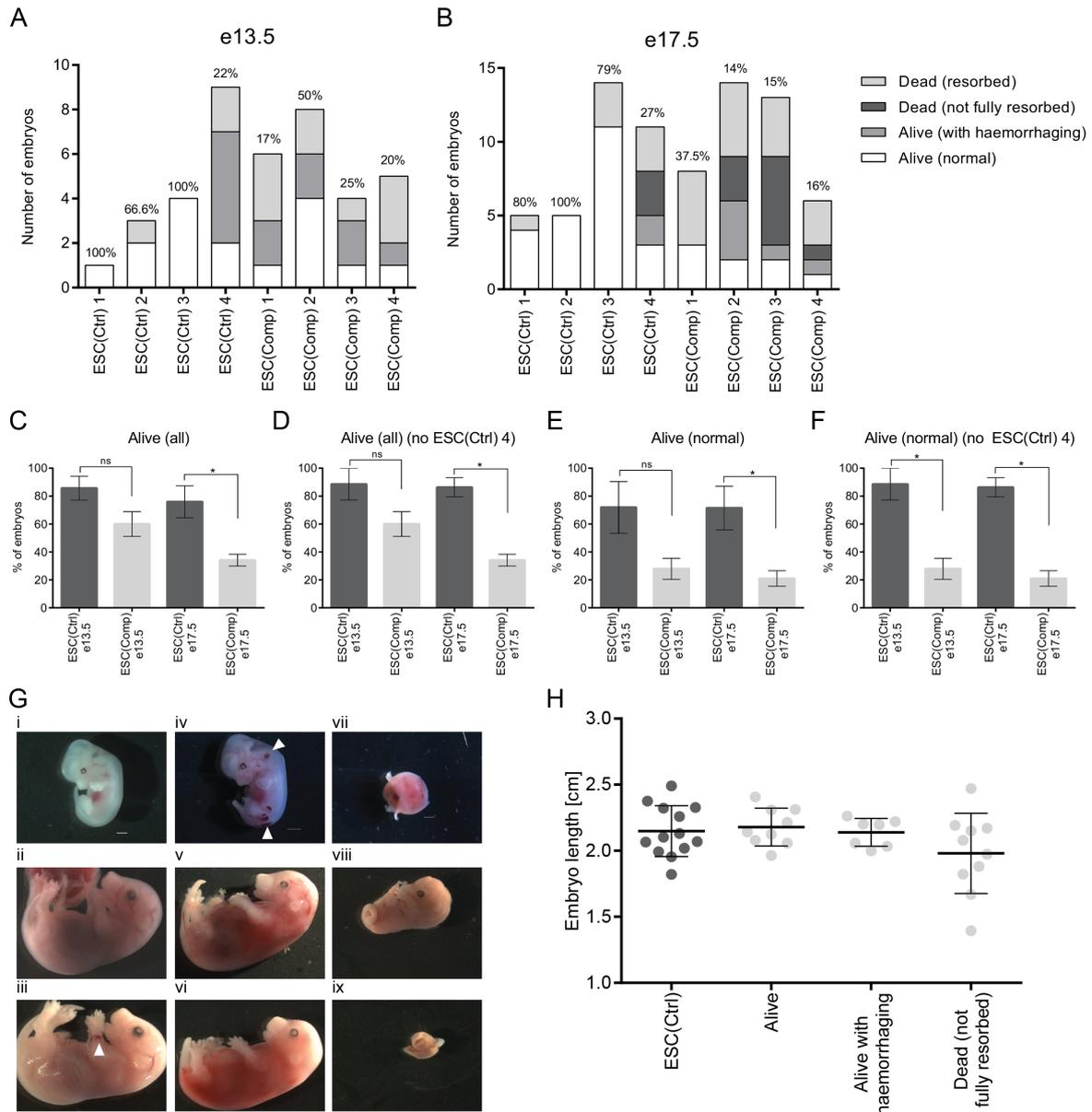


Fig. 4.3 Survival phenotype (A-B) Phenotypes of chimaeric embryos at e13.5 (A) and e17.5 (B). Numbers above bars are the percentages of live (normal) embryos. (C-F) Percentages of live embryos compared between ESC(Ctrl) and ESC(Comp) groups for both time points: including all live embryos (C), all live embryos but excluding ESC(Ctrl)4 samples (D), only healthy/normal live embryos (E), only healthy/normal live embryos and excluding ESC(Ctrl)4 samples (F). Stars represent significant difference as calculated by unpaired t-test with Welch's correction ($p < 0.05$). (G) Range of phenotypes: live (normal) e13.5 (i), live (normal) e17.5 (ii), live (haemorrhaging) e17.5 (iii), live (haemorrhaging) e13.5 (iv), dead (not fully resorbed) e17.5 (v, vi), dead (resorbed) e13.5 (vii), dead (resorbed) e17.5 (viii, ix). White arrow heads indicate small foci of haemorrhaging. Other phenotypes include edema (v) and lack of posterior development (viii). (H) Embryo crown-rump length in cm.

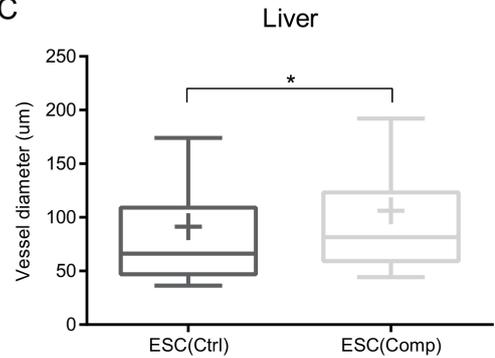
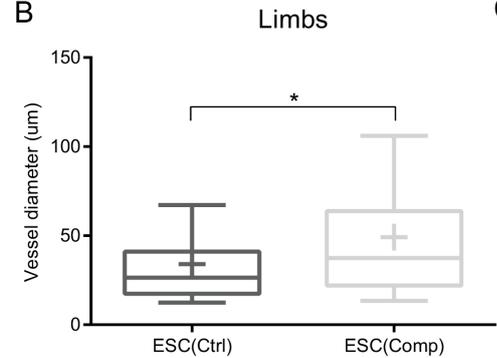
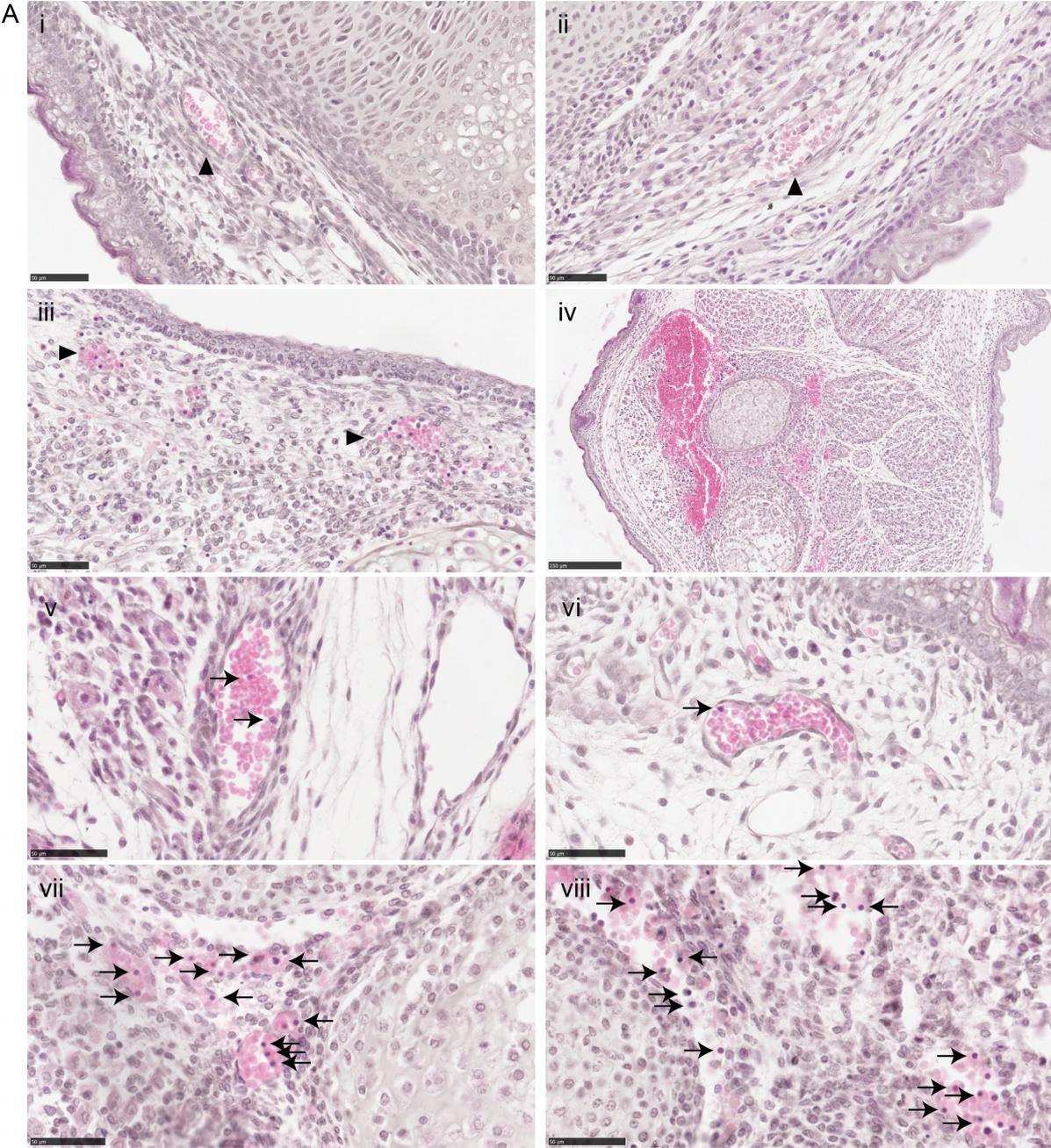
Interestingly, the phenotype of the ESC(Ctrl)4 chimaeras was more like that of the ESC(Comp) chimaeras rather than the other controls. This is intriguing since this clone has given rise to healthy chimaeric animals in the past. The proportions of each of the observed phenotypes for these chimaeras was very similar to that of ESC(Comp)1-4 for both time points (Figure 4.3A,B).

We therefore also compared the proportions of live embryos between the two groups after excluding ESC(Ctrl)4 from the analysis. Looking at the healthy and haemorrhaged embryos together, there was still no significant difference at e13.5 (Figure 4.3D), but the proportion of just the healthy looking embryos was now significantly lower in the ESC(Comp) group compared to ESC(Ctrl)1-3 (Figure 4.3F).

Furthermore, it was of interest when the embryos died. To answer this question, we looked at the size of the embryos. There was no significant difference between the ESC(Ctrl) chimaeras and any of the subsets of ESC(Comp) embryos, suggesting that their time of death was at approximately e17.5 or shortly before (Figure 4.3H).

4.2.3 Histological Analysis

To take a closer look at what developmental defects occurred in the ESC(Comp) chimaeric embryos, the e17.5 embryos were dissected and a number of the major organs were fixed, paraffin embedded, sectioned and stained with H&E (haematoxylin and eosin) for histological analysis. Since most of the bleeding was observed in the limbs, livers and some lungs, those tissues were chosen for analysis first. We cannot be completely sure whether the ESC(Ctrl)4 chimaeras serve as good controls and hence they were excluded from this analysis. Since there was a haemorrhaging phenotype, we first looked at the blood vessels. As shown in Figure 4.4B,C the blood vessels in the liver and in the limbs of the ESC(Comp) chimaeras had a greater average diameter than those of the ESC(Ctrl) chimaeras. It was also noted that the blood vessel walls were less well defined in the ESC(Comp) and it looked like their basement membrane was impaired (Figure 4.4A(ii)). Some ruptured vessels were also observed (Figure 4.4A(iii)). The severity of this varied from no or very few ruptured blood vessels observed to major haemorrhaging (Figure 4.4A(iv)), but this was not observed at all in any of the ESC(Ctrl) samples. Additionally, in the limb sections, we noticed that the number of nucleated blood cells appeared to be higher in the ESC(Comp) chimaeras than in the ESC(Ctrl) chimaeras (Figure 4.4A(v-viii)). These are likely either white blood cells or nucleated erythrocyte progenitors. This was also observed in the livers but to a lesser extent. Neither of these effects were seen in the lungs.



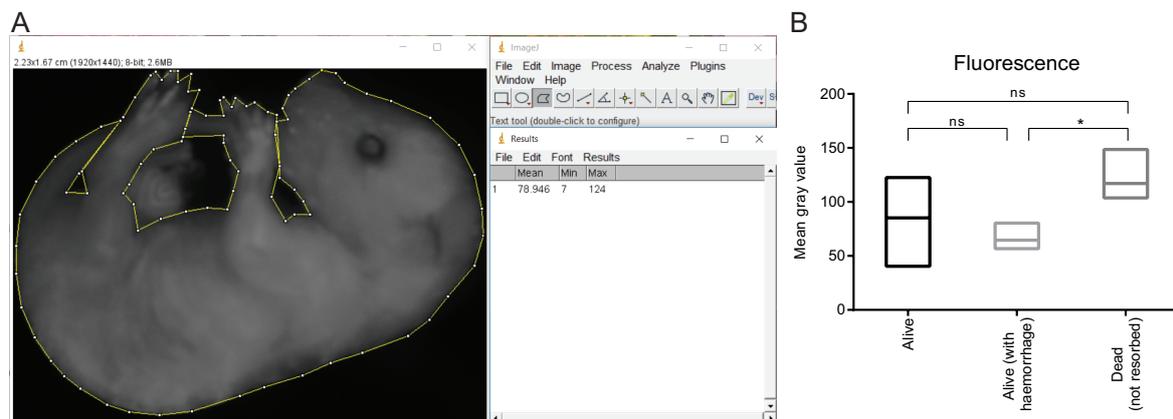


Fig. 4.5 Measuring fluorescence intensity by mean grey value (A) Fluorescent images of e17.5 embryos were analysed in ImageJ. Their outline was drawn (yellow line) and the mean grey value of the defined area was measured. (B) Mean grey values for ESC(Comp) chimaeric embryos. ns means no significant difference and stars represent significant differences as calculated by ANOVA ($p < 0.05$).

4.2.4 Correlation Between Contribution and Phenotype

There was a range of phenotypes and severity. All of the dead embryos showed severe haemorrhaging whereas the degree of this varied more for the alive embryos and was mainly restricted to smaller foci of bleeding. Interestingly, this phenotype could already be observed at e13.5 embryos, indicating that, although lethality occurs later during development, the phenotype starts to manifest during mid-gestation, if not earlier.

Given that these embryos are chimaeric and will potentially have different degrees of ES cell contribution, we wondered whether these two factors correlated. To look into this, we imaged the fluorescent embryos and used the grey-scale images to estimate ES cell contribution using the mean grey value in ImageJ as shown in Figure 4.5A. This was only done for the ESC(Comp) embryos. There was no significant difference between the live (normal) and the live (haemorrhaging) groups or the live (normal) and the dead (not resorbed) groups. However,

Fig. 4.4 (preceding page) Histological analysis at e17.5 (A) H&E images of sections of the limbs. i) ESC(Ctrl), ii-iii) ESC(Comp). Arrow heads showing blood vessels. iv) hemorrhage in ESC(Comp). v-vi) ESC(Ctrl), vii-viii) ESC(Comp). Arrows showing nucleated blood cells. Scale bar in iv is 250 μ m, all others are 50 μ m. (B-C) Box plots showing diameter of blood vessels in the limbs (B) and liver (C). Whiskers range from 10th to 90th percentile. The line inside the box represents the median, the mean is shown by a cross. Stars represent significant differences as calculated by the Mann-Whitney U test ($p < 0.05$).

the mean grey value of the fluorescent images of the dead (not resorbed) group was significantly higher than that of the live (haemorrhaging) group, indicating that there might be a correlation between contribution and phenotype (Figure 4.5B).

4.3 Discussion

This chapter describes how the eight ES cell clones were labelled fluorescently using the PiggyBac system and subsequently injected into blastocyst to further define the developmental potential by assessing the phenotype of the chimaeras derived from the compromised ESC(Comp) clones. We have shown that the ESC(Comp) clones do not give rise to viable chimaeras and that embryonic death occurs roughly around e17.5. Additionally, it was observed that there were various degrees of haemorrhaging in the ESC(Comp) chimaeras and that there are defects in blood vessel integrity.

4.3.1 Cell Labelling

To fluorescently label as many cells as possible and have a low copy number of transgene integration to avoid further detrimental mutations, three different transfection methods were tested for the efficiency and a copy number analysis in a number of labelled cells was performed. This showed that nucleofection was the most efficient transfection method. Similar results were also found previously by other groups [110]. Cells that were transfected using nucleofection also had the least copies of *Venus*. However, efficiency and copy numbers were only tested in one of the control cell lines and using the pPBCAG-Venus-IP plasmid. Although transfection efficiency is unlikely to change a great deal in the other cell lines and using the pPBCAG-H2BtdTomato-IH plasmid, it is possible that there will be variations in copy numbers between and within the eight cell lines, especially, since this was only analysed for a small subset of cells.

It is also noteworthy that the difference in the number of *Venus* copies present in cells transfected with lipofection and the other two methods is very large, which was unexpected. All three methods are routinely used by many laboratories and it would be interesting to further examine whether this effect only occurred by chance or whether this is a real phenomenon.

4.3.2 Phenotypic Differences Between ESC(Ctrl) and ESC(Comp) Cells

Using the labelled cells, we showed that there is a clear difference in the developmental potency between the ESC(Ctrl) and ESC(Comp) cells. The viability of the embryos was greatly reduced in the ESC(Comp) chimaeras at e17.5. All of the dead embryos and even some of the live

ones showed signs of haemorrhaging. Interestingly, some embryos also had edemas which are caused by abnormalities in the development of the lymphatic system [111], suggesting that there may be defects in blood vessel as well as lymphatic vessel development. Although no embryonic lethality was observed at e13.5 yet, we could already see haemorrhaging in a large number of embryos. Together with the data that there is no significant difference in embryo size at e17.5, this suggests that the phenotype starts to manifest before e13.5 but embryonic death does not occur until around e17.5. However, embryo size is not the ideal way to determine the embryo's age and developmental progress as it can be affected by a number of things such as litter size. A better way to analyse whether the embryos are at similar developmental stages would have been to weigh both the embryo and the placenta and calculate the fetal/placental weight ratio.

One problem with using fluorescence to detect chimaerism was that for the resorptions it was not possible to determine whether they had ES cell contribution or not. Apart from one embryo, fluorescence could not be detected for any of them. This is likely due to the fact that most of the tissue will have been reabsorbed or that the Venus and tdTomato proteins have been degraded. Therefore, there is a chance that some of the resorptions included in this analysis may not have had ES cell contribution. However, this is the case for both ESC(Ctrl) and ESC(Comp) chimaeras.

ESC(Ctrl)4 Clone Behaves Like the Compromised ES Cells

One ES cell clone that stuck out was ESC(Ctrl)4. This clone had previously given rise to good, high percentage chimaeras with germline transmission. However, in this study, the phenotype of the chimaeras generated using this cell line look more like those of the compromised ES cell clones. This was unexpected. Our hypothesis is that the compromised clones have acquired some genetic or epigenetic mutations during their culture that lead to the embryonic lethal phenotype. It is possible that the same or similar changes have accumulated in the cells of the ESC(Ctrl)4 cell line. This hypothesis will be explored further in the next chapters.

Histological Analysis

Histological differences were found between tissues of the ESC(Ctrl) chimaeras and the ESC(Comp) chimaeras. Due to the haemorrhaging phenotype, this analysis was mainly focussed on blood vessel development. Both blood vessels of the liver and the subcutaneous vessels of the limbs were dilated. Additionally, we saw a number of ruptured blood vessels and blood vessels with reduced vascular integrity. Similar vessel phenotypes were described in a number of mouse models with a haemorrhaging phenotype [112–114].

There is still a lot more work to do to fully define the phenotype of the ESC(Comp) chimaeras and to understand the precise developmental processes that are affected. Firstly, the H&E analysis needs to be completed for more organs to see whether the vascular phenotype persists in most of the body or whether it is confined to specific organs or tissues. Additionally, it would be of interest to look at the blood vessel structure in more detail. Antibodies for CD31 (also called PECAM1 - Platelet endothelial cell adhesion molecule) and α SMA could be used for this. CD31 is expressed by endothelial cells of all blood vessels [115] and α SMA is usually present in the walls of larger blood vessels. The combination of these two markers is often used to assess blood vessel structure by looking at α SMA coverage or by analysing the organisation of the CD31-positive and α SMA-positive cells lining the blood vessels.

An increase in the number of nucleated blood cells was also observed in tissues of ESC(Comp) chimaeras. It is not clear, however, whether these are white blood cells or nucleated erythrocyte progenitors as all of these are nucleated. If they are white blood cells, it is also unclear, whether this is part of the phenotype and a reason for embryonic lethality or whether this is part of the body's wound healing response. It is well established that white blood cells play major roles in wound healing and tissue repair [116]. Therefore, it is possible that an increase in the number of white blood cells is the body's response to the ruptured blood vessels and other potential tissue malformations. However, the nucleated blood cells could also be erythrocyte progenitors. These nucleated progenitors of red blood cells usually only persist in the blood until approximately e16.5 [117], suggesting that there may be a developmental delay in the ESC(Comp) chimaeras in terms of blood cell development.

4.3.3 Correlation Between ES Cell Contribution and Phenotype

We further showed that there is an increase in the fluorescence signal in the dead embryos compared to the haemorrhaged live ones. Interestingly, there was no significant difference between the live (normal) embryos and either the haemorrhaged or the dead ones. This is likely due to the larger variation in that set of samples. The fourth phenotype (dead and resorbed) had to be excluded from this analysis since as mentioned above the majority of the resorptions were not fluorescent.

4.3.4 Chapter Conclusion

The data presented in this chapter show that chimaeras generated from ESC(Comp) ES cells have a vascular phenotype and most of them die around e17.5. We observed a range of haemorrhaging phenotypes as well as edemas in some embryos. Histology data analysing blood vessels in the limbs and liver revealed that ESC(Comp) chimaeras have dilated vessels

and poor vascular integrity compared to the ESC(Ctrl) chimaeras. This phenotype strikingly resembles earlier steps of angiogenesis, suggesting that there may be a delay in the development of the fetal vasculature in the ESC(Comp) chimaeras. Interestingly, we also observed an increased number of nucleated blood cells in ESC(Comp) chimaeras, which could potentially be immature erythrocytes [117]. This would also indicate that there may not only be a delay in the development of the fetal vasculature but also a delay or defects in some aspects of haematopoiesis. Other tissues were not analysed and it remains to be determined whether there are developmental delays in other organs and tissues, too.

Chapter 5

Transcriptomics

5.1 Introduction

The work described in this chapter provides a detailed overview of the transcriptome profiles of the eight ES cell clones that were used for the blastocyst injections described in Chapter 4 ([Blastocyst Injections and Histology](#)). The aim of this was to identify any transcriptional changes that might account for the altered developmental phenotype of the ESC(Comp) clones and that could in the future serve as markers to distinguish between ES cells with good or compromised developmental potency.

5.2 Results

5.2.1 RNAseq Data Processing and Analysis

We performed single-end RNA sequencing of the four ESC(Ctrl) and the four ESC(Comp) clones, with three replicates (R1-R3) for each sample. The sequencing coverage was 10M reads per sample. Data processing was performed using the open, web-based bioinformatics platform Galaxy (<https://usegalaxy.org/>). Read quality was assessed using the FastQC tool. The overall quality of the reads was excellent with 94.6% of all reads having a Phred quality score (Q score) of 30 or higher (Figure 5.1). The Q score is a measure of base calling accuracy and an indicator of the probability of a base calling error (P). It is logarithmically related to P by the equation $Q = -10 \log_{10} P$, i.e. the higher the Q score, the lower the probability that a base call is wrong. At $Q=30$, there is only a 1/1000 chance of the base call being wrong, or a 99.9% chance that it is right.

To remove low quality bases (Phred score $Q \leq 30$) on the sequence read ends, as well as positional sequence bias at the start of the sequence reads [118], the first 8 and the last 4 bases

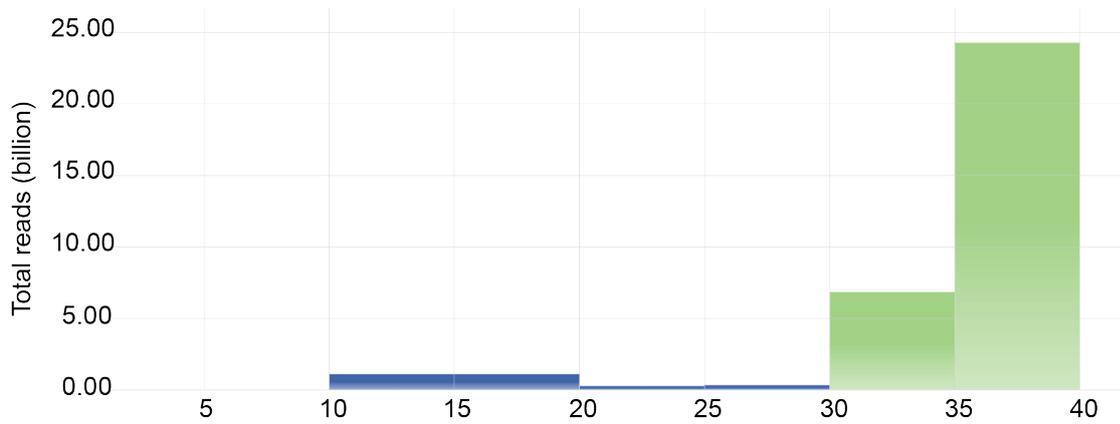


Fig. 5.1 **Q score distribution** Total Q score distribution for all reads of all samples.

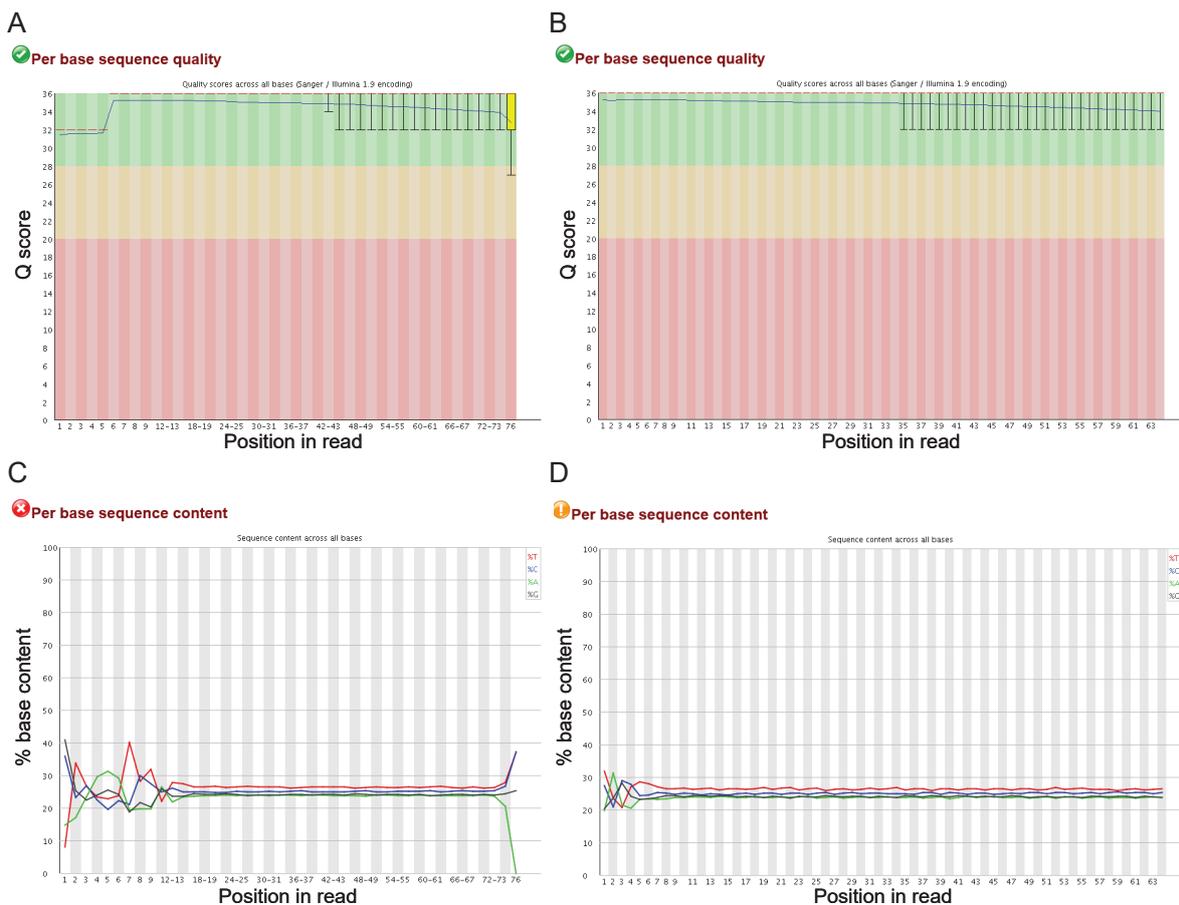


Fig. 5.2 **FastQC outputs** Representative example of (A-B) Per base sequence quality and (C-D) Per base sequence content pre (A,C) and post (B,D) trimming.

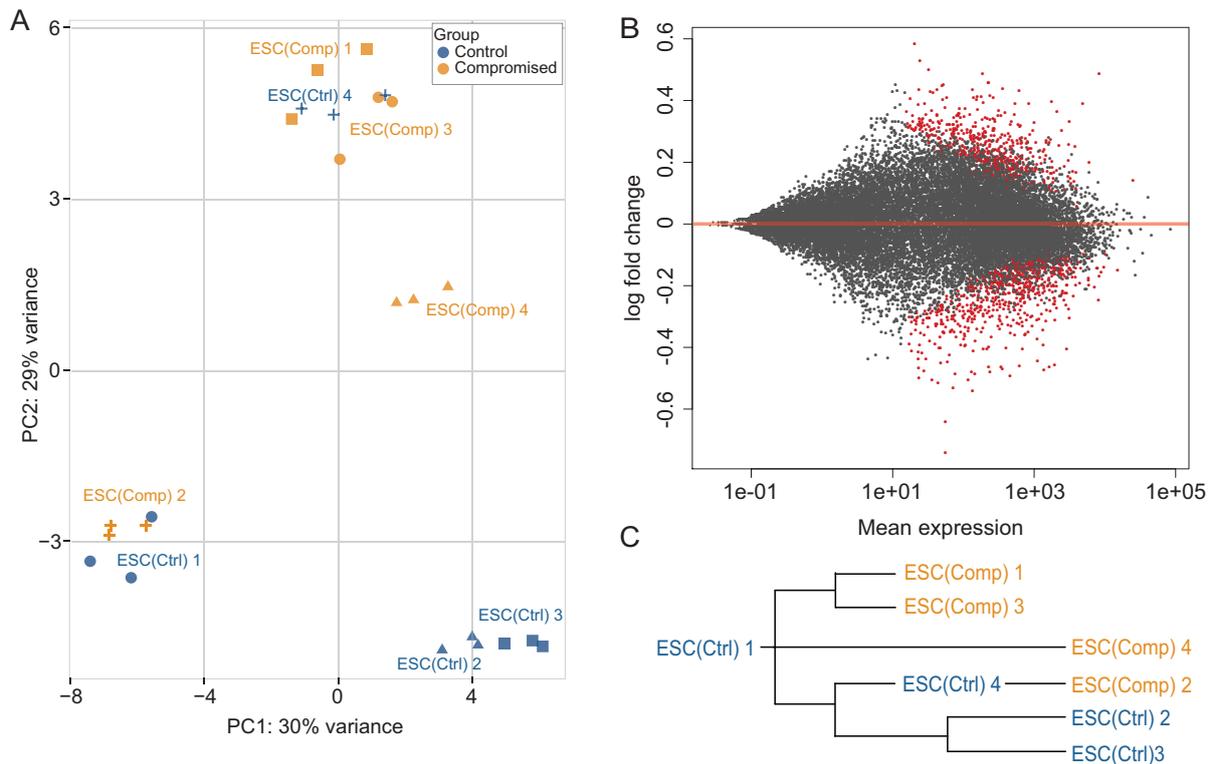


Fig. 5.3 **RNAseq results** (A) PCA plot and (B) MA-plot comparing gene expression in ESC(Comp) samples vs ESC(Ctrl) samples. (C) Pedigree of ES cell clone derivation.

were trimmed off the reads. Representative examples of FastQC outputs pre and post trimming are shown in Figure 5.2. They show that although the per base sequence quality was very good to start with (Figure 5.2A), the trimming has improved the Phred score ($Q \geq 30$) even further (Figure 5.2B). Additionally, the noise in the per base sequence content at the start of the sequence read (Figure 5.2C) has been removed (Figure 5.2D) through trimming.

The trimmed sequence reads were mapped against the GRCm38.p4 reference genome with Tophat [73]. We then calculated read counts in htseq-count [74] using the strict mode, which only counts a sequence read when the read overlaps with the whole gene and when the read is unique to a single gene. Genes were annotated with the GENCODE annotation vM7.

We next tested for differential gene expression between our ESC(Ctrl) and ESC(Comp) samples, using the DESeq2 tool. On the principal component analysis (PCA) plot (Figure 5.3A) we observed that the three replicates of each sample cluster very closely together, which is a good confirmation that the sequencing run has worked and the data are reliable. From the PCA plot it is, however, not possible to deduce whether there are differences in gene expression between the two test groups as there are no distinct clusters grouping the samples in Controls and Compromised clones. Nonetheless, we can identify four clusters of samples. The first one containing ESC(Ctrl)1 and ESC(Comp)2, the second one containing ESC(Ctrl)4, ESC(Comp)1

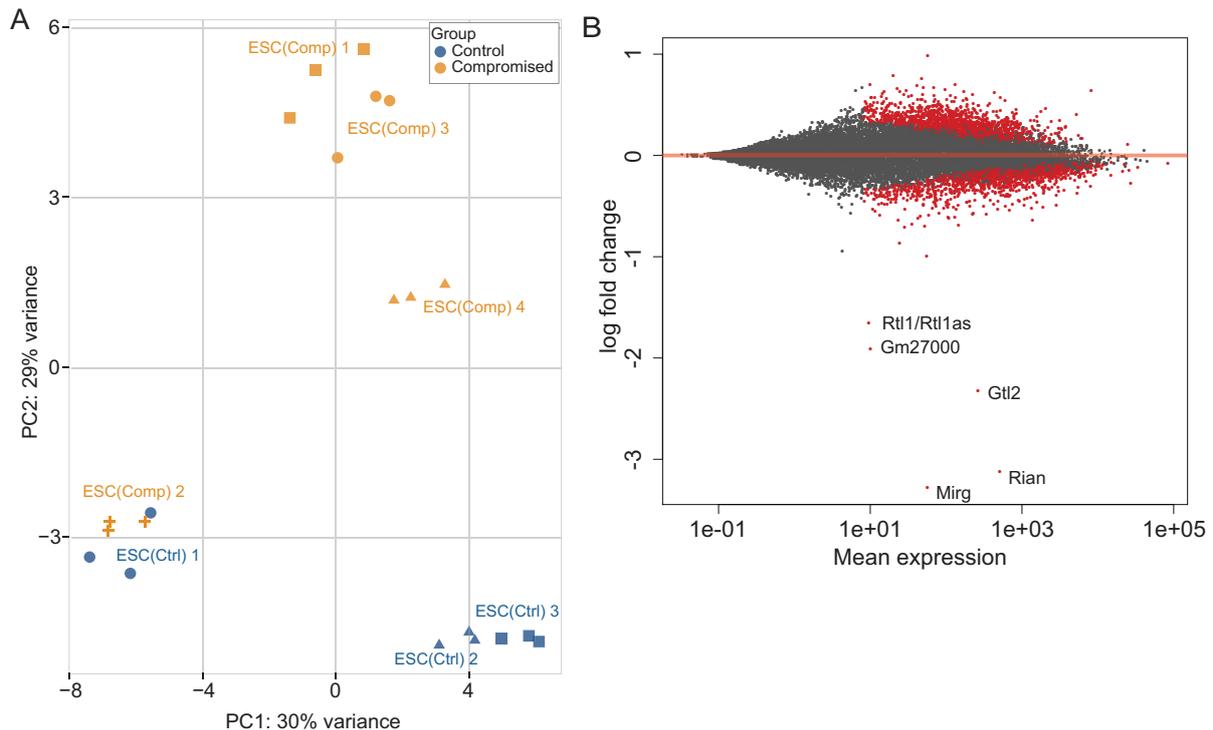


Fig. 5.4 **RNAseq results without ESC(Ctrl)4** (A) PCA plot without ESC(Ctrl)4. (B) MA-plot comparing all ESC(Comp) samples vs ESC(Ctrl)1-3.

and ESC(Comp)3, the third one containing ESC(Comp)4 and the last one containing ESC(Ctrl)2 and ESC(Ctrl)3. In general, this clustering is representative of the pedigree (Figure 5.3C) of how the samples were derived as described in Chapter 1 (Introduction). The only exception to this is sample ESC(Ctrl)4, which we would have expected to cluster closer to ESC(Comp)2 or ESC(Ctrl)2 and ESC(Ctrl)3. Similarly, the MA-plot, which visualises the differences in gene expression between the two sample groups, shows that there is not a single gene that is differentially expressed between all ESC(Comp) and all ESC(Ctrl) clones with a log₂fold change $\geq \pm 1$ (Figure 5.3B).

RNAseq Data Analysis Excluding Sample ESC(Ctrl)4

From the blastocyst injection data we have seen that in terms of its phenotype, ESC(Ctrl)4 behaves more similarly to the compromised clones than the controls. The number of chimaeras that were alive and healthy at e17.5 was significantly lower than in the other three controls and similarly to the four compromised clones, in addition to dead embryos we saw a greater number of embryos with haemorrhaging.

The PCA plot also indicates that its transcriptome may be more similar to the compromised clones. We believe that this clone may have undergone similar genetic or epigenetic changes

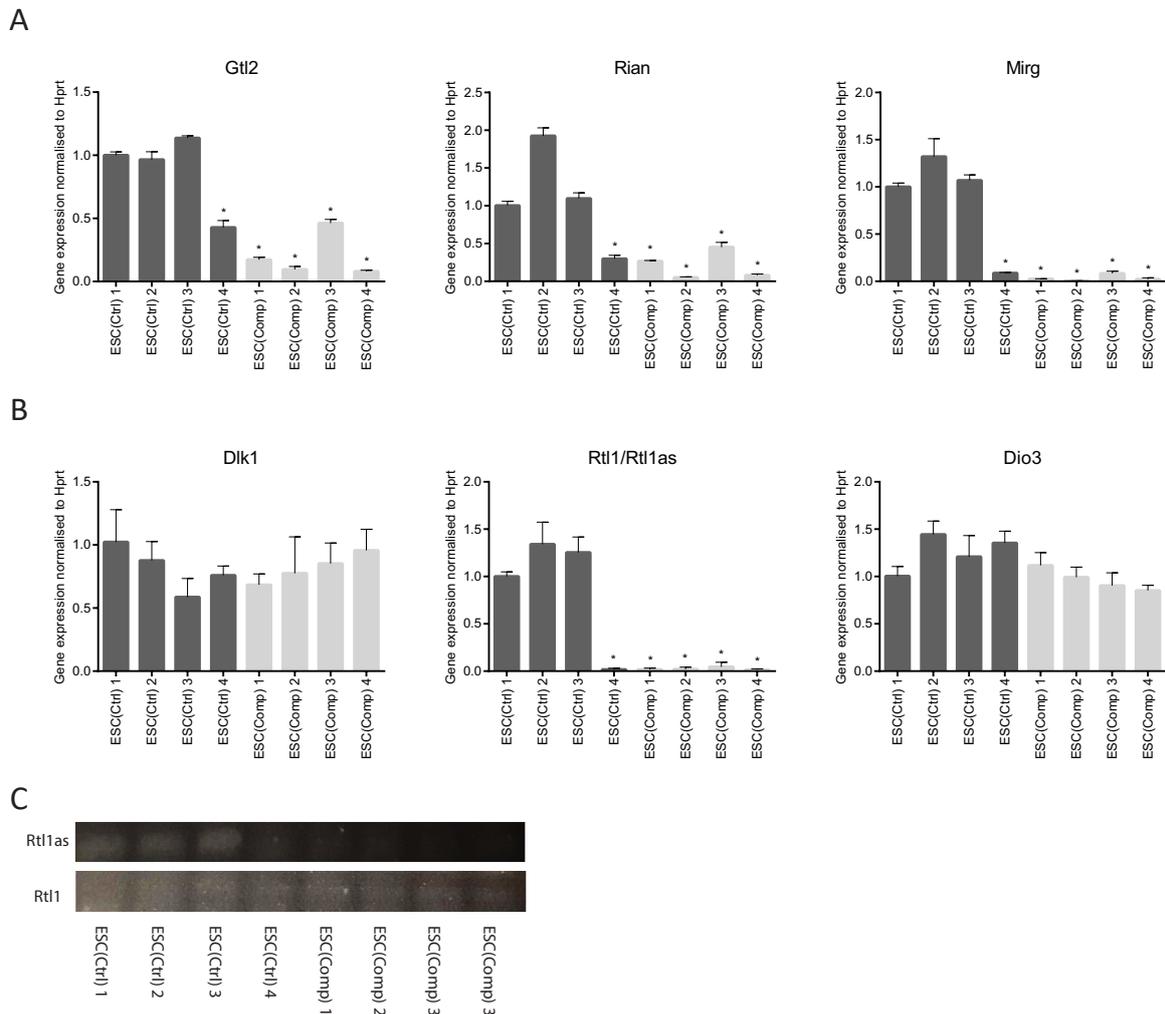


Fig. 5.5 Validation of RNaseq results (A) Maternally and (B) paternally expressed genes. Error bars represent standard error of the mean. Stars represent significant differences as calculated by ANOVA ($p < 0.05$). (C) Results of the PCR for *Rtl1as* (top lane) and *Rtl1* (bottom lane) after primer specific reverse transcription.

as the ESC(Ctrl) clones after the initial successful blastocyst injections were performed. We were therefore interested to see whether there were differences in gene expression between only ESC(Ctrl)1-3 and the ESC(Comp) clones. Indeed, as shown on the new MA-plot (Figure 5.4B), there were now five genes that are differentially expressed with a log2fold change $\geq \pm 1$ ($p < 0.05$, Benjamini-Hochberg Correction). These are *Mirg* (miRNA containing gene), *Rian* (RNA imprinted and accumulated in nucleus), *Gtl2* (also called *Meg3* - Maternally expressed 3), *Gm27000* and *Rtl1* (Retrotransposon gaglike 1). *Gm27000* is a predicted gene on the X-chromosome. Interestingly, all other genes are located in the same region on chromosome 12. This is the imprinted *Dlk1-Dio3* region (Figure 5.8), containing the paternally expressed protein

coding genes *Dlk1* (Delta like non-canonical Notch ligand 1), *Rtl1* and *Dio3* (Deiodinase, iodothyronine type III), as well the maternally expressed lncRNA *Gtl2*, *Rian*, *Mirg* and *Rtl1as*, an antisense transcript to the paternally expressed *Rtl1*. *Rtl1* and *Rtl1as* are indistinguishable by unstranded RNAseq. Therefore the lower expression of *Rtl1* in the ESC(Comp) samples identified by RNAseq really represents a lower cumulative expression of *Rtl1/Rtl1as*.

5.2.2 Validation of Expression Levels of Genes in the *Dlk1-Dio3* Region

We validated the expression levels of all these genes by qPCR using cDNA prepared from new samples of cells. The expression levels were all as expected from the RNAseq results. *Gtl2*, *Rian*, *Mirg* and *Rtl1/Rtl1as* were expressed at much lower levels in the four ESC(Comp) clones, whereas there was no significant difference between the expression of *Dlk1* and *Dio3* in the two groups (Figure 5.5A-B).

In order to distinguish between *Rtl1* and *Rtl1as*, we used primer-specific reverse transcription to make cDNA, followed by conventional PCR to amplify the product. This method has been described previously [71]. In brief, because the two transcripts are expressed in opposite directions, each of the primers of the primer pair designed to amplify the *Rtl1/Rtl1as* transcript will exclusively bind to either the *Rtl1* or the *Rtl1as* transcript. If we replace the oligo(dT)s in the reverse transcription reaction with either the forward or reverse primer, we can specifically amplify *Rtl1as* or *Rtl1*, respectively. A PCR will then show which of the transcripts was present in each sample. As shown in Figure 5.5C *Rtl1as* was expressed in ESC(Ctrl)1-3 but silenced in all ESC(Comp) samples, as well as ESC(Ctrl)4. *Rtl1* was not detectable in any of our samples.

5.2.3 Additional Compromised Clones Have a Similar Expression Profile at the *Dlk1-Dio3* Region

To further confirm the relationship between the expression levels of the maternally expressed genes of the *Dlk1-Dio3* region and the embryonic lethal phenotype that we observed after blastocyst injections of the ESC(Comp) clones, four more control clones (ESC(Ctrl)5-8) and three more compromised clones (ESC(Comp)5-7) were analysed. In Table 5.1 we show that the control clones have previously given rise to a high number of excellent chimeras, whereas we did not get many pups (and no chimeras) born with the compromised clones.

qPCR analysis of expression of *Gtl2*, *Rian* and *Mirg* revealed that while the four additional controls had normal expression levels, all three ESC(comp) clones showed silencing of *Rian* and *Mirg* and two of them also had lower expression of *Gtl2* (Figure 5.6).

Table 5.1 Blastocyst injection results of additional clones

Cell line	Number of pups	Of which chimera
ESC(Ctrl) 5	22	20
ESC(Ctrl) 6	18	18
ESC(Ctrl) 7	20	18
ESC(Ctrl) 8	14	14
ESC(Comp) 5	1	0
ESC(Comp) 6	0	0
ESC(Comp) 7	0	0

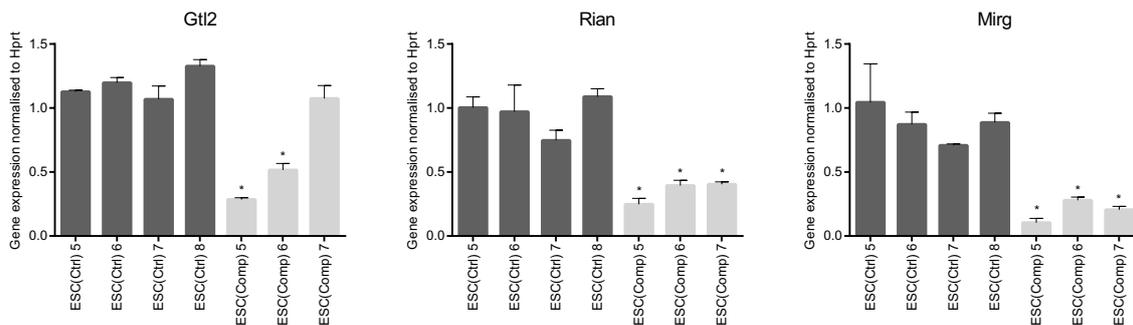


Fig. 5.6 Expression of *Gtl2*, *Rian* and *Mirg* in additional clones Error bars represent standard error of the mean. Stars represent significant differences as calculated by ANOVA ($p < 0.05$).

5.2.4 *Gtl2* Expression in the Embryos

We have identified differences in gene expression by using ES cells. However, the phenotype we observe occurs late during gestation around embryonic day e17.5. This poses the question of whether these genes are still silenced in the embryo.

To answer this question, we collected limbs from e17.5 chimeras from ESC(Ctrl)1-4 and ESC(Comp)1-4 and looked at the expression of *Gtl2* as a marker for the whole region. We chose limbs because *Gtl2* expression has been shown to be high in fetal skeletal muscle [119]. To account for the variation in the amount of muscle in each sample, we normalised gene expression to the muscle marker *Myf5* (Myogenic factor 5). In Figure 5.7A we show that as expected chimeras from ESC(Ctrl)1-3 ES cell clones expressed *Gtl2* in their limbs, whereas *Gtl2* expression in chimeras from ESC(Comp)1-3 and ESC(Ctrl)4 was very low. In chimeras

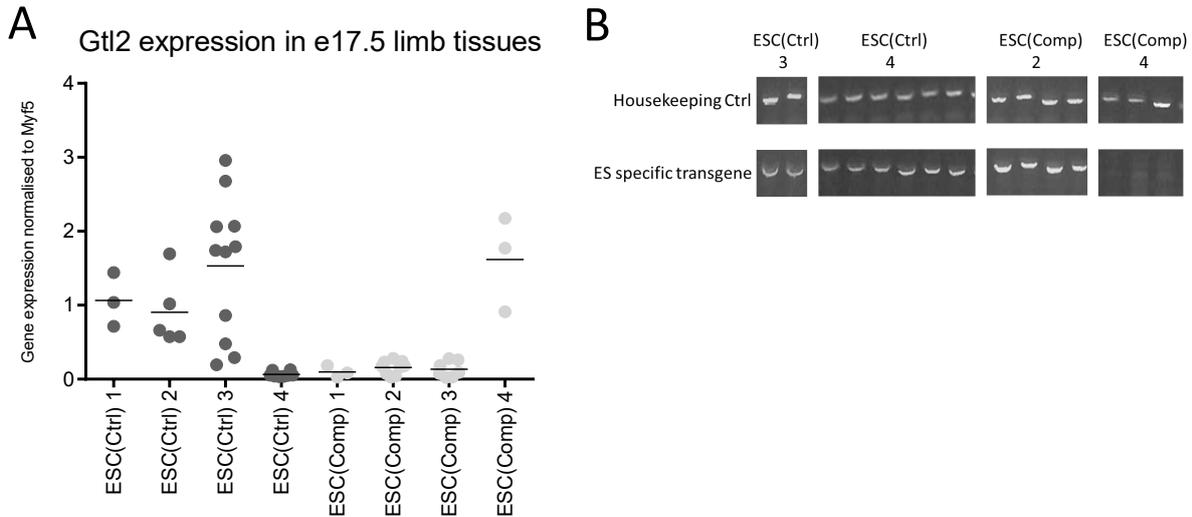


Fig. 5.7 *Gtl2* tissue expression (A) *Gtl2* expression in limbs from e17.5 chimeras. Each dot represents a tissue sample. Horizontal bars represent the means for each group of samples. (B) Genotyping results for the three ESC(Comp)4 chimaeras and representative chimaeras generated from other ESC(Ctrl) and ESC(Comp) clones. Top row is the PCR result for a housekeeping control, bottom row is the PCR results for primers specific to each ES cell clone.

from the ESC(Comp)4 clone, however, *Gtl2* levels were similar to those of the controls. Interestingly, ES cell contribution in these chimaeras was also much lower than in all other chimaeras (Figure 5.7B)

5.3 Discussion

We have shown that there is a relationship between the expression of the maternally expressed genes in the *Dlk1-Dio3* region on Chromosome 12 and the developmental potency of ES cells. Initially, this was identified using RNAseq technology to analyse the transcriptome of the four ESC(Ctrl) and the four ESC(Comp) clones that were used for the blastocyst injections described in Chapter 4 ([Blastocyst Injections and Histology](#)). We further confirmed this relationship with additional control and compromised ES cell clones and also showed that the expression of these genes remains silenced in ES cell derived embryonic tissues.

5.3.1 Structure and Control of the *Dlk1-Dio3* Region

The *Dlk1-Dio3* region is located on the distal part of chromosome 12 and is an imprinted region, meaning that some genes are expressed exclusively from the paternally inherited copy, whereas others are expressed exclusively from the maternally inherited copy. This is usually controlled

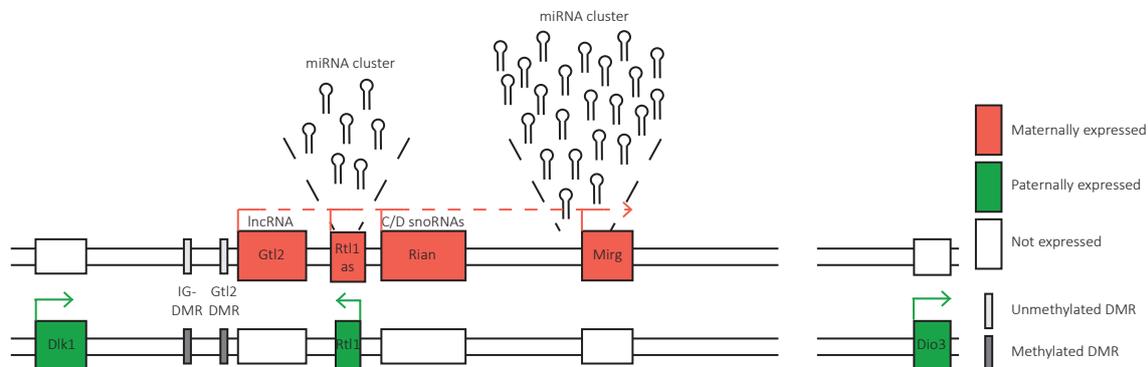


Fig. 5.8 *Dlk1-Dio3* locus *Dlk1-Dio3* imprinted region on the distal part of chromosome 12. The paternal allele encodes the protein coding genes *Dlk1*, *Rtl1* and *Dio3*. The maternal allele encodes the non-coding RNAs *Gtl2*, *Rtl1as*, *Rian* and *Mirg*. *Rian* encodes a number of C/D box snoRNAs. The region also contains two miRNA clusters: one within *Rtl1as* and the other one within *Mirg*. There are two DMRs: the IG-DMR and the *Gtl2*-DMR which are methylated on the paternal but unmethylated on the maternal chromosome. Expression of maternally expressed genes is mainly controlled by the IG-DMR.

through differential DNA methylation of the two alleles at known regulatory regions.

The *Dlk1-Dio3* region (Figure 5.8) contains three protein coding genes expressed from the paternal allele as well as a number of noncoding RNAs expressed from the maternal allele. Figure 5.8 gives an overview of the region. The paternally expressed protein coding genes are *Dlk1*, *Rtl1* and the most distal gene of the region *Dio3* which is located 800kb downstream of *Dlk1*. Expressed from the maternal allele are the long noncoding RNA (lncRNA) *Gtl2*, the three genes *Rian*, *Rtl1as*, an antisense transcript to *Rtl1*, and *Mirg*, as well as additional intergenic transcripts such as microRNAs (miRNAs). *Rian* harbours a large cluster of C/D box small nucleolar RNAs (snoRNAs) and *Rtl1as* and *Mirg* also produce a number of miRNAs. Given that these transcripts are all expressed in the same direction and the lack of an obvious promoter apart from the *Gtl2* promoter, it is believed that the maternal genes are all expressed as one long polycistronic transcription unit. Their gene expression profile is remarkably similar during embryonic development, but the function of most of these maternally expressed genes is largely unknown.

There are several differentially methylated regions (DMRs) in this region that are known to regulate expression, but the most important one for the control of *Gtl2*, *Rian*, *Rtl1as* and *Mirg* is the IG-DMR (intergenic DMR), located between *Dlk1* and *Gtl2*, approximately 13kb upstream

of the *Gtl2* promoter. It inherits a methylation mark during spermatogenesis and therefore, only the paternal copy is methylated whereas the maternal copy remains unmethylated. DNA methylation is associated with repressed gene expression. Hence, in a normal situation *Gtl2*, *Rian*, *Rtl1as* and *Mirg* are exclusively expressed from the maternal allele.

5.3.2 Literature Review - Previous Publications Confirm Our Results

Uniparental Disomy and Knockout Mouse Models

The *Dlk1-Dio3* region has previously been implicated with an embryonic lethal phenotype and developmental defects. In 2000, Georgiades et al. [120] created a mouse model for maternal or paternal uniparental disomy of chromosome 12 (mUPD12 and pUPD12 respectively). The pUPD12 embryos would be expected to have low expression of the usually maternally expressed genes *Gtl2*, *Rian* and *Mirg* as they inherit two paternal alleles. pUPD12 embryos are alive until embryonic day e15.5 but the number of alive embryos decreases for later stages and none survive to birth. mUPD12 embryos have a slightly less severe phenotype with embryos dying perinatally. pUPD12 embryos show defects in muscle and skeletal development. Another study [121] looked at the effect of maternal or paternal duplication of just the distal part of chromosome 12 (which is where the *Dlk1-Dio3* region is located). Their results are similar to the above study, yet slightly more severe with embryos dying prior to e18.5. Since the *Dlk1-Dio3* region is the only known imprinted region on chromosome 12, these studies suggest that the genes in this region play important roles during embryonic development. This was further confirmed with a study by Lin et al. in 2007 [122]. They created a mouse model with a deletion in the IG-DMR. Maternal transmission of this deletion causes paternalisation of the maternal chromosome at this locus, meaning that the genes that are normally expressed from the maternal chromosome are silenced and instead *Dlk1*, *Rtl1* and *Dio3* are expressed from both chromosomes. The phenotype they observe is similar to the previous two studies [120, 121] - embryos die pre- or perinatally starting from e16.5 and they also show defects in muscle and skeletal development. Table 5.2 sums up the results of these publications.

Findings From iPSCs and ES Cells

In 2010, Stadtfeld et al. [123] observed that iPSCs that do not give rise to high percentage chimeras by 2n blastocyst injection and that do not give rise to embryos through tetraploid complementation do not express *Gtl2*, *Rian*, *Mirg* and many of the miRNAs encoded in that region. These results show striking similarities to our findings. Unlike our study, however, they did have chimeras born after 2n blastocyst injection, although they had very low iPSC

Table 5.2 Mouse models with defects at the *Dlk1-Dio3* region

Model	mUPD12/ mUPD(dist12)	pUPD12/ pUPD(dist12)	matIG-DMR deletion
Paternal gene expression	0X	<i>Dlk1/Dio3</i> 2X <i>Rtl1</i> 4X	<i>Dlk1/Dio3</i> 2X <i>Rtl1</i> 4X
Maternal gene expression	2X	0X	0X
Lethality	perinatal	prenatal pUPD12: from e15.5 pUPD(dist12) before e18.5	pre-/perinatal beginning at e16.5
Developmental defects	Muscle Skeletal/Bone Placenta	Muscle Skeletal/Bone Placenta	Muscle Skeletal/Bone
Citations	[120][121]	[120][121]	[122]

contribution.

More recently, similar results have also been published for ES cells. Stelzer et al. have developed a reporter system that can track methylation at imprinted loci *in vivo* [124]. They have used this system to analyse methylation and gene expression dynamics at the *Dlk1-Dio3* locus in ES cells as well as embryos and adult mice [125]. They show that in a culture of ES cells, a small population of cells will emerge over time that has either activated (IG-DMR hypomethylation) or silenced (IG-DMR hypermethylation) the usually maternally expressed genes on both alleles, the latter one likely being equivalent to our ESC(Comp) cells in terms of gene expression. Using tetraploid complementation assays with these cells they show that disruption of normal gene expression at the *Dlk1-Dio3* locus causes severe developmental defects. While 4n complementation embryos with both alleles active (twice the normal dosage of maternal genes) die prior to implantation, 4n complementation embryos with both alleles silenced (no expression of maternal genes) survive until midgestation but then die with severe developmental defects such as brain malformation and muscle defects. The findings from this study are similar to our findings. However, the exact phenotype they observe is different to our results described in Chapter 4 ([Blastocyst Injections and Histology](#)). In this study they find that the embryos die between e12.5 and e15.5, whereas we observe lethality later during gestation around e17.5. Importantly, they also do not report any haemorrhaging. There are several potential explanations for this. The first one being the difference in blastocyst injections. They performed tetraploid complementation which results in embryos with 100% ES cell contribution, whereas we injected 2n blastocysts. Although embryos from 2n blastocyst

injection can have close to 90-95% ES cell contribution, there will almost always be at least some host cell contribution. This will potentially have an effect on the severity of the phenotype. Additionally, there is a difference in genetic background of the ES cells that were used. Stelzer et al. used 129XCAST F1 ES cells, whereas all of our ES cells were 129XB6 F1 ES cells. It is possible that the precise phenotype will also depend on the genetic background.

5.3.3 Silencing of Maternally Expressed Genes From The *Dlk1-Dio3* Region in ESC(Ctrl)4

Our RNAseq experiment and the gene expression validations showed that ESC(Ctrl)4 has the same pattern of gene expression of *Gtl2*, *Rian*, *Mirg* and *Rtl1as* as the ESC(Comp) clones. Together with the blastocyst injection data, which showed that the chimeras derived from ESC(Ctrl)4 cells have the same phenotype as the chimeras derived from the ESC(Comp) clones, this strengthens our hypothesis that this clone may have undergone similar genetic or epigenetic changes as the compromised clones, which ultimately means that it may not be a good control. In Chapter 6 ([Epigenetics](#)) we will comment on why we think it has contributed to good chimeras in the past but seems to have lost that ability. We will continue to include it in our experiments and analyses but it will be treated as a separate, independent sample.

5.3.4 Impact of Other Genes

Our data in combination with the evidence from the literature review, indicate that silencing of the maternally expressed genes from the *Dlk1-Dio3* region is likely causing the embryonic lethal phenotype that we observe. However, we did also find that *Gm27000* expression is significantly lower in ESC(Comp) clones compared to the ESC(Ctrl) clones. It is of course a possibility that this is also playing a role in the differences in developmental potency that we observe between the two groups.

Additionally, given that the phenotype that we observe occurs relatively late during gestation, we cannot rule out the possibility of there being differences in gene expression at later stages during development. Our RNAseq analysis only looked at differential gene expression at the ES cell state. If there were additional genetic or epigenetic changes affecting loci that are naturally not expressed at the ES cell level, these changes would remain hidden.

5.3.5 *Gtl2* Expression in e17.5 Tissues

By looking at the expression levels of *Gtl2* in limbs from e17.5 embryos we confirmed that on the whole the expression of *Gtl2* remained silenced during development. The only exception

here were the three chimeras obtained from the ESC(Comp) 4 clone. This was very unexpected, but can potentially be explained by relating tissue *Gtl2* levels with contribution of ES cells to the chimeras. The ESC(Comp)4 chimeras all had much lower contribution of ES cell-derived cells than the chimeras obtained from the other cell lines. This means that there must have been a high contribution from the host cells, which we would expect to have normal *Gtl2* expression levels and therefore they would increase total *Gtl2* expression levels.

We looked at the expression in the whole limb because they contain skeletal muscle and *Gtl2* is known to be expressed at high levels in skeletal muscle [119]. To account for the differences in amount of muscle in every sample we normalised *Gtl2* expression to *Myf5* which is expressed in muscle and involved in myogenesis during embryonic development. This method has two main caveats. Firstly, we cannot exclude *Gtl2* expression from other tissues from the limb such as skin, fat or bone. This is unlikely, as previous work on *Gtl2* expression pattern during embryonic development indicates [119], but it cannot be ruled out completely. Secondly, knockout and uniparental disomy studies have shown that lack of expression of the maternally expressed genes from the *Dlk1-Dio3* region causes a muscle phenotype. Using a muscle marker to normalise against could therefore skew the data. However, the work from Stadtfeld et al. [123] show that while gene expression of other muscle markers such as *Myogenin* are affected there is no difference in *Myf5* expression between embryonic tissues derived after 4n blastocyst injections of iPSCs that do or do not express *Gtl2*. Nevertheless, a more accurate way to assess *Gtl2* expression in ES cell derived embryonic tissues might be to make use of the fact that our ES cells are fluorescently labelled and any ES cell derived cells in the embryo should therefore be fluorescent, too. We could dissect muscle tissue, then dissociate it (for example using enzymatic dissociation with trypsin) and finally sort or manually pick the fluorescent cells. Looking at gene expression using qPCR on those cells only would make the results more accurate and likely reduce variation between samples as we would exclude host cell contribution. Furthermore, just using muscle tissue would eliminate the need to use a muscle marker as a housekeeping gene and it would also reduce the possibility of *Gtl2* contamination from other tissues.

5.3.6 Chapter Conclusion

In this chapter we described that RNA sequencing of ESC(Ctrl) and ESC(Comp) ES cells only identified 5 genes that are differentially expressed between the two groups. Four of these, *Gtl2*, *Rtl1as*, *Rian* and *Mirg* are all expressed from the maternal allele of the *Dlk1-Dio3* imprinted region on chromosome 12 and silenced in the ESC(Comp) clones. Growing evidence from the literature also suggests that silencing of this locus causes embryonic lethality in mice. This lack of expression is also seen not just seen in the ES cells but also in tissues from e17.5 chimaeric

embryos.

Not much is known so far about the precise role and mode of action of these genes during development. However, this locus has been implicated in the regulation of angiogenesis in tumours. Interestingly, as presented in Chapter 4 ([Blastocyst Injections and Histology](#)), we also observed a blood vessel phenotype that resembles a delay in angiogenesis or vessel maturation. This will be discussed further in the overall thesis discussion.

Chapter 6

Epigenetics

6.1 Introduction

In Chapter 5 ([Transcriptomics](#)) we described that the maternally expressed genes of the *Dlk1-Dio3* region, *Gtl2*, *Rian*, *Mirg* and *Rtl1as*, are aberrantly silenced in the compromised ES cell clones. Gene expression in this region is mainly regulated by differential methylation of the IG-DMR, located approximately 13kb upstream of the *Gtl2* promoter. We therefore hypothesise that the differences in gene expression between the ESC(Ctrl) and the ESC(Comp) clones is caused by abnormal DNA methylation of the *Dlk1-Dio3* IG-DMR in the ESC(Comp) clones. In this chapter, we present a methylation analysis of this region and show that the IG-DMR is indeed hypermethylated in the compromised ES cell clones. We additionally show that gene expression can be rescued with the DNMT inhibitor 5-azacytidine.

6.2 Results

6.2.1 Bisulfite Sequencing of the *Dlk1-Dio3* IG-DMR

The IG-DMR is the main regulatory element controlling expression of the maternal genes at the *Dlk1-Dio3* region and we performed targeted bisulfite sequencing of the IG-DMR, since we were interested to find out whether hypermethylation of this region on the maternal chromosome may be causing the silencing of the maternally expressed genes in our ESC(Comp) samples. Bisulfite treatment of DNA will convert all cytosine residues into uracil. Methylated cytosines, however, are protected from this and bisulfite treated DNA will therefore only retain those cytosine residues that are methylated. Using sequencing methods, we can then identify and analyse DNA methylation. This experiment was carried out under the guidance of Dr M. Eckersley-Maslin at the Babraham Institute Cambridge. We first bisulfite converted gDNA that

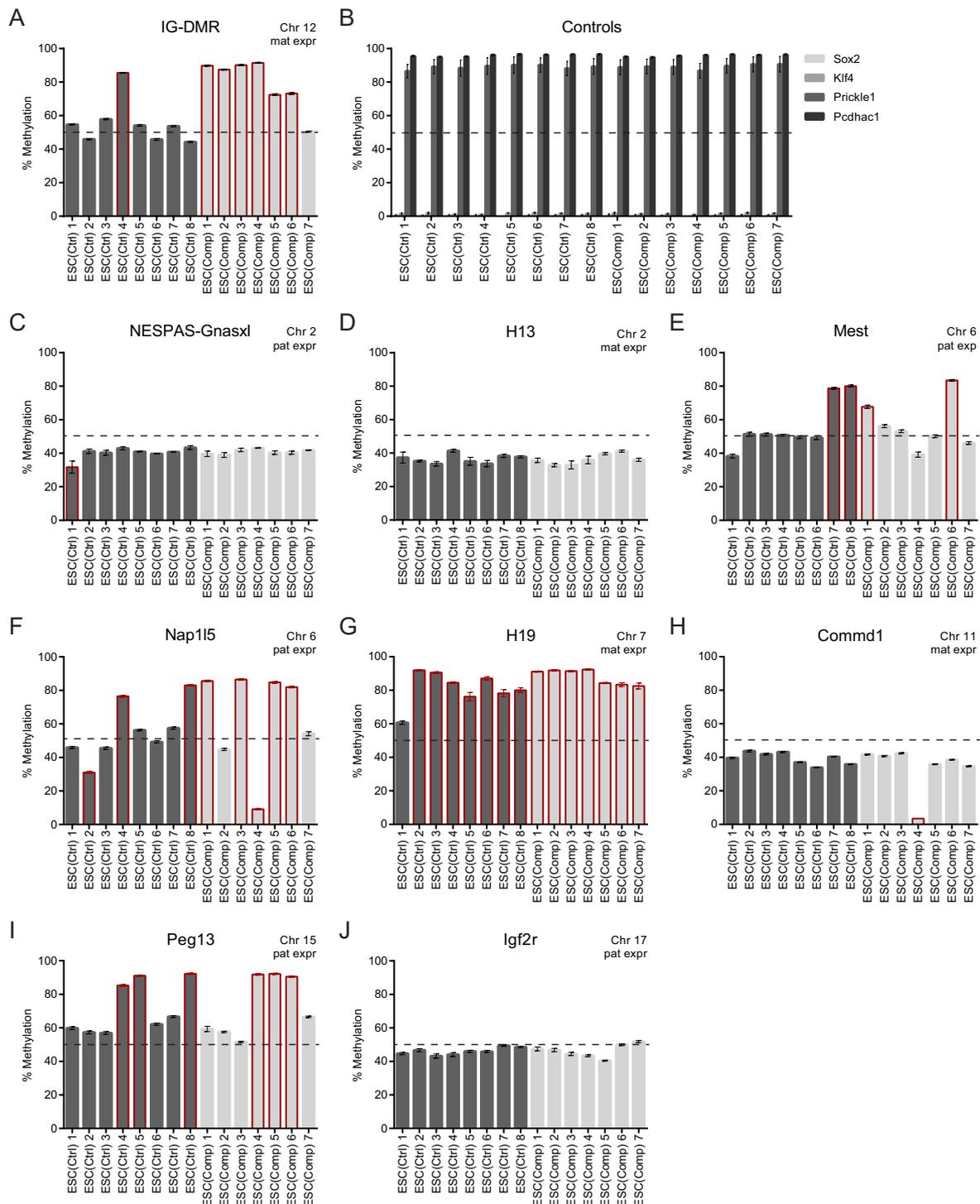


Fig. 6.1 Bisulfite sequencing results (A) IG-DMR of the *Dkl1-Dio3* region (B) Control regions (C) NESPAS-Gnasxl (D) H13 (E) Mest (F) Nap115 (G) H19 (H) Commd1 (I) Peg13 (J) Igf2r. Red outlines indicate aberrant methylation levels. The dotted line indicates 50% methylation.

was extracted from the cell lines, followed by two PCR steps to amplify the imprinted region and to attach barcoded sequencing adapters. Finally, the samples were sequenced in a MiSeq PE150 run.

In the normal situation, the *Dlk1-Dio3* IG-DMR is only methylated on the paternal allele to allow transcription from the maternal allele. We therefore expected the control cell lines to have a methylation level of approximately 50% overall, whereas any clones where gene expression is silenced due to hypermethylation of the IG-DMR were expected to have methylation levels close to 100%. As Figure 6.1A shows, all ESC(Ctrl) clones (apart from ESC(Ctrl)4) were indeed 50% methylated whereas ESC(Ctrl)4 and ESC(Comp)1-6 were almost fully methylated. The only exception was ESC(Comp)7 which also had normal methylation levels. Importantly, however, we also show that all cell lines had the expected methylation levels of the four control genes: *Sox2* and *Klf4* are active in ES cells and should therefore not be methylated, whereas *Prickle1* and *Pcdhac1* are usually fully methylated and not expressed at the ES cell state (Figure 6.1B). We did not distinguish between DNA methylation on the paternal and the maternal allele, but only took into account total DNA methylation.

6.2.2 Bisulfite Sequencing and RNAseq Results of Other Imprinted Regions

To see whether the abnormal methylation of imprinted genes (loss of imprinting) was restricted to the *Dlk1-Dio3* locus or whether it was a more global phenomenon, we analysed methylation levels of known DMRs of additional imprinted genes by Bisulfite sequencing as before. The data presented in Figure 6.1C-J show that there was indeed clone-to-clone variation in methylation levels at other imprinted genes. However, no other region showed such a consistent difference between the controls and the compromised ES cell clones as the IG-DMR of the *Dlk1-Dio3* region.

We then went back to our RNAseq data to see whether this pattern of clone-to-clone variation was also reflected in the expression levels of other imprinted genes. Similar to the methylation data, there was silencing or overexpression in individual clones but again, there was no obvious pattern like for the maternally expressed genes from the *Dlk1-Dio3* region (data not shown).

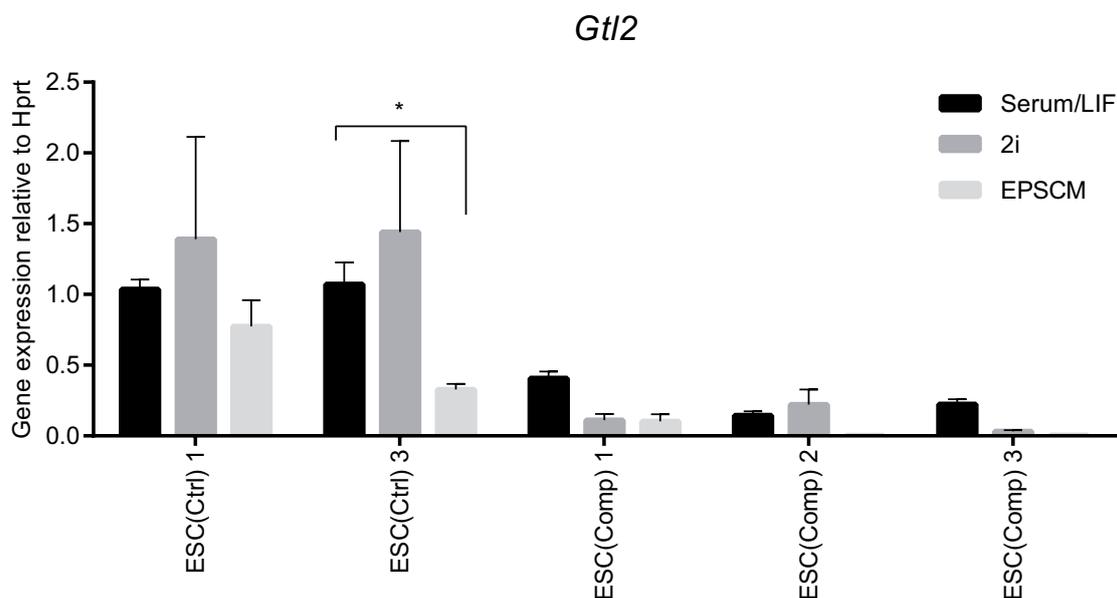


Fig. 6.2 **ES cell culture in different media conditions** *Gtl2* expression in five ES cell clones after culture in standard Serum/LIF conditions as well as 2i and EPSCM. Error bars represent standard error of the mean. Stars represent significant differences as calculated by 2-way-ANOVA ($p < 0.05$).

6.2.3 Rescuing Expression of the Maternally Expressed Genes of the *Dlk1-Dio3* Region

Culture in 2i or EPSCM Did Not Affect *Gtl2* Expression Levels

It is a well established fact that methylation is a very dynamic and reversible DNA modification. We therefore wondered, whether it was possible to reduce methylation of the IG-DMR and thereby reactivate expression of the maternally expressed genes. To do this, we cultured some of the ES cell clones in 2i and EPSCM (Expanded potential stem cell media) for at least 3 passages. Both of these are routinely used to culture ES cells and have been shown previously to reduce global DNA methylation levels [126, 127]. We have assessed only *Gtl2* expression as an indicator of whether there is any effect of these conditions on the whole region. However, there was no significant increase in *Gtl2* expression after culturing the cells in 2i or EPSCM compared to our standard culture conditions. If at all, there was a tendency for lower *Gtl2* expression after culture in EPSCM, but this was only significant for ESC(Ctrl)3 (Figure 6.2).

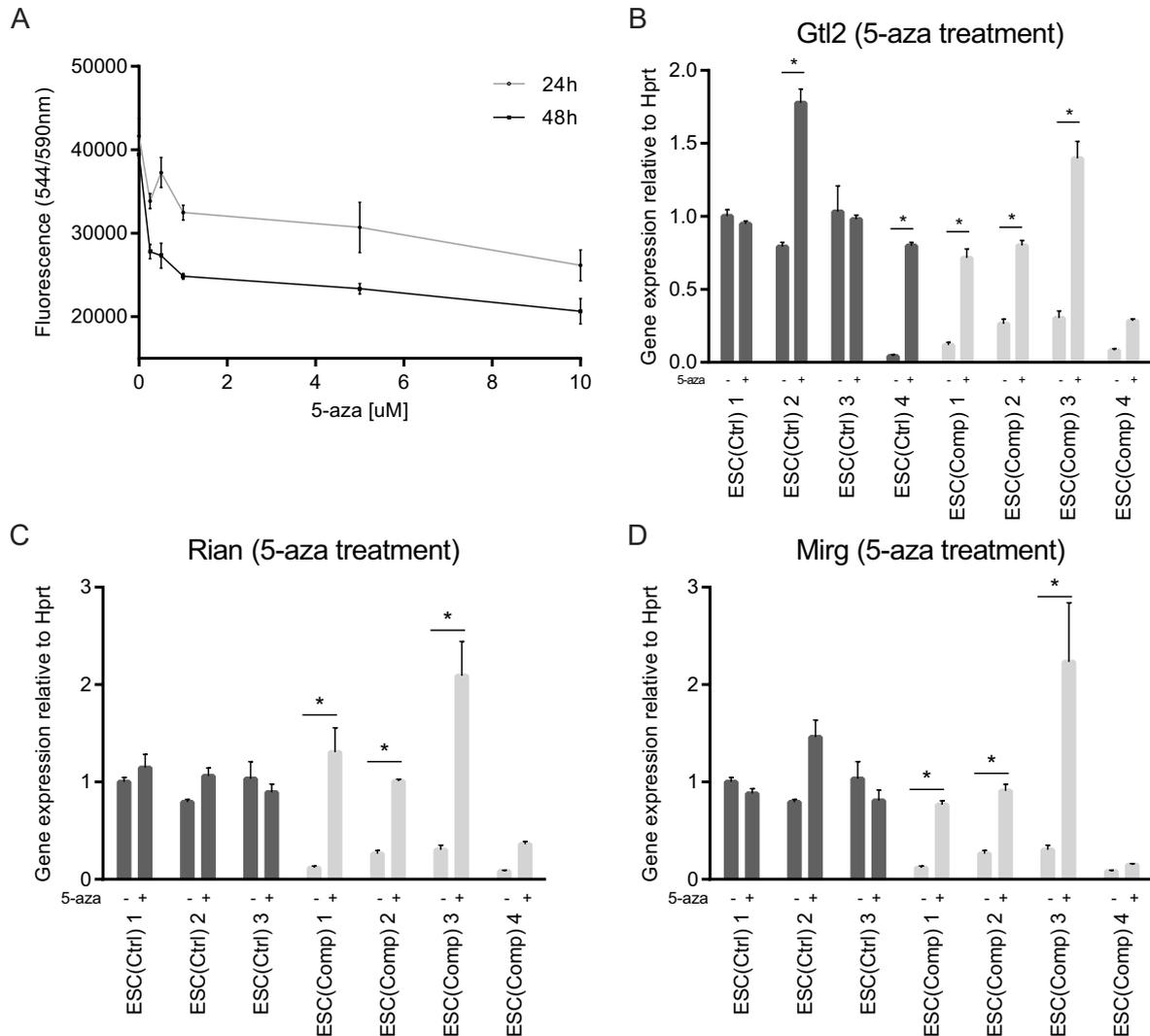


Fig. 6.3 Treatment with 5-azacytidine (A) Viability assay for 5-azacytidine concentrations ranging from 0-10 uM and for 24 h (light grey line) or 48 h (dark grey line). (B-D) *Gtl2*, *Rian* and *Mirg* expression after 48 h of 0.5 uM 5-azacytidine treatment. Error bars represent standard error of the mean. Stars represent significant differences as calculated by 2-way-ANOVA ($p < 0.05$).

5-azacytidine Treatment Reactivated *Gtl2* Expression

Since we did not achieve reactivation by using different culture conditions, we were keen to investigate whether treatment of the cells with 5-azacytidine (5-aza) would work. 5-aza is a DNA methyltransferase (DNMT) inhibitor and decreases global DNA methylation by trapping DNMTs and marking them for degradation. This obviously causes 5-aza to be quite toxic for cells in high concentrations or if it is administered for too long. Hence we performed

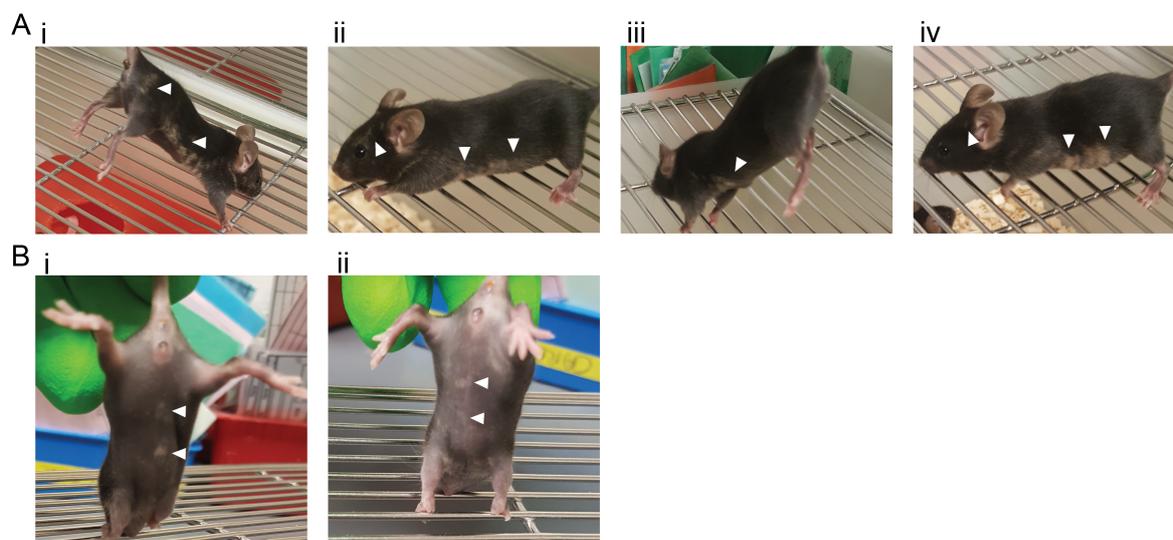


Fig. 6.4 Chimaeric offspring after 5-aza injections (A) One female (i) and three males (ii-iv) born after injection of 5-aza treated, unselected cells. Photographs were taken at 1 month of age (B) Two very low percentage chimaeras born after injection of a 5-aza treated subclone (ESC(Comp)3-AZA28). Photographs were taken at 2 weeks of age. White arrows show differences in coat colour.

a cell viability assay first to ascertain the optimal treatment conditions (Figure 6.3A). From this we chose to treat the cells with 0.5 μ M 5-aza for 48 h. RNA was extracted immediately after stopping the treatment. qPCR analysis for *Gtl2*, *Rian* and *Mirg* expression showed that we could indeed increase their expression with 5-aza treatment (Figure 6.3B-D). Two of the compromised clones expressed these genes to levels comparable to the untreated controls. There was no significant difference in ESC(Comp)4 after 5-aza treatment and the expression of all three genes in ESC(Comp)3 after 5-aza treatment was even higher than in the untreated controls. Interestingly, one of the control clones (ESC(Ctrl)2) had doubled its *Gtl2* expression levels after 5-aza treatment. This was potentially due to activation of the paternal allele. We decided to inject some of the 5-aza treated cells into blastocysts to see whether this was sufficient to rescue their embryonic lethal phenotype and whether any chimaeras survive. After the 48 h 5-aza treatment, the cells were left to recover for another 48 h before being injected. As shown in Figure 6.4A four chimaeras were born, but they were all quite low percentage chimaeras. As before, blastocyst injections were performed by Professor W.H. Colledge.

Subcloning of 5-azacytidine Treated Cells

So far we had only looked at the population average of *Gtl2*, *Rian* and *Mirg* expression directly after the 5-aza treatment. However, it is unlikely that all cells will have the exact same

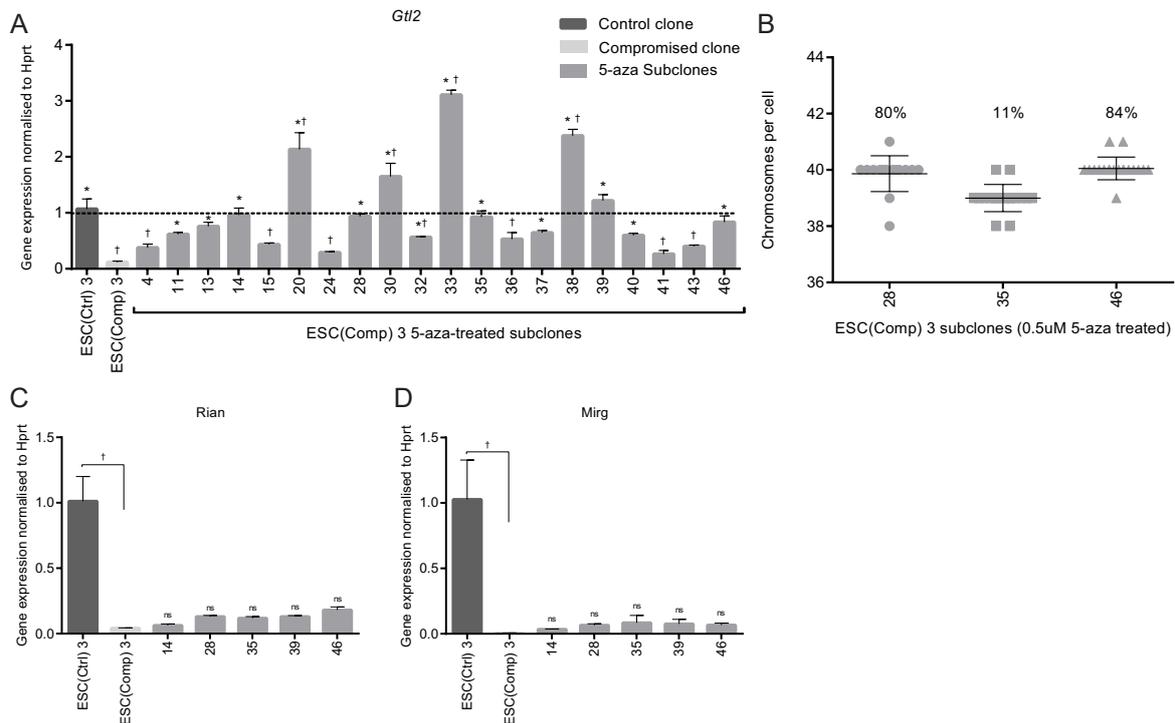


Fig. 6.5 Analysis of 5-azacytidine treated subclones of ESC(Comp)3 (A) *Gtl2* expression in 20 subclones of ESC(Comp)3 after 48h of 0.5uM 5-azacytidine treatment, compared to ESC(Ctrl)3 and the parental untreated ESC(Comp)3 cell line. Error bars represent standard error of the mean. Crosses and stars represent significant differences to ESC(Ctrl)3 and ESC(Comp)3 respectively as calculated by ANOVA ($p < 0.05$). (B) Karyotype analysis of three subclones with rescued *Gtl2* expression ($n \geq 15$). (C-D) *Rian* (C) and *Mirg* (D) expression in the *Gtl2* expressing ESC(Comp)3 subclones. ns means not significantly different from ESC(Comp)3 parental clone.

methylation changes. In some cells 5-aza may not have had an effect on the methylation of the *Dlk1-Dio3* IG-DMR at all, in others it may have demethylated one allele and in yet another group of cells it may have demethylated both the maternal and the paternal allele causing a double dose of gene expression. Additionally, previous publications have shown that after a few days without treatment, 5-aza treated cells partially recover their DNA methylation levels. To investigate this and to find out whether this may have been the reason for the low contribution of the 5-aza rescued cells to chimaeras, we strove to maintain a pure population of cells that all have stably reactivated the maternally expressed genes. For this, ESC(Comp)3 ES cells were treated with 0.5 uM 5-aza for 48 h like before and then the cells were let to recover for a week. After passaging, a number of ES cell colonies were picked to grow up and perform qPCR for *Gtl2*, *Rian* and *Mirg* on. Overall, the cells had been off treatment for 18 days at the time of RNA extraction. There was a lot of variation between the subclones in their level of

Gtl2 reactivation, confirming our hypothesis that the effect of 5-aza varies between cells. While the *Gtl2* expression in some colonies was just increased slightly, others had levels three times as high as the untreated ESC(Ctrl)3 clone. Out of 20 subclones of ESC(Comp)3 that were analysed, 14 had significantly increased *Gtl2* levels. Of these, six (13, 14, 28, 35, 39, 46) had *Gtl2* levels comparable to the untreated ESC(Ctrl)3 clone (Figure 6.5A). Unexpectedly, the levels of *Rian* and *Mirg* did not change (Figure 6.5C,D).

From the six partially rescued subclones, three were chosen for karyotyping. This was performed as described in Chapter 3 (Initial Characterisation of the ES Cell Clones). Two of those, subclones 28 and 48, had a normal karyotype with 80% and 84% of chromosome spreads that were counted respectively having 40 chromosomes (Figure 6.5B).

One subclone(ESC(Comp)3-AZA28) was chosen for blastocyst injections. Although only *Gtl2* was successfully reactivated, we were interested to see whether this was sufficient to at least partially rescue the phenotype. However only two very lower percentage chimaeras were born as well as 11 pups with no contribution (Figure 6.4B).

6.3 Discussion

In this chapter we have demonstrated that the IG-DMR of the *Dlk1-Dio3* region is hypermethylated in six out of seven of our compromised clones as well as the abnormal ESC(Ctrl)4 clone. Other imprinted regions also showed loss of imprinting, but there were no consistent differences between ESC(Ctrl) and ESC(Comp) clones. We have further shown that we can rescue *Gtl2*, *Rian* and *Mirg* expression using 5-azacytidine treatment.

6.3.1 Bisulfite Sequencing of the IG-DMR

Using targeted bisulfite sequencing of the IG-DMR of the *Dlk1-Dio3* region, we have shown that in the control clones the level of methylation is approximately 50% which is what would be expected for a correctly imprinted locus. In the compromised clones as well as the ESC(Ctrl)4 clone, however, the level of methylation is close to 100%. Since DNA methylation is associated with gene silencing this hypermethylation of the IG-DMR, which is the main regulatory element of expression of *Gtl2*, *Rian*, *Rtl1as* and *Mirg*, is very likely responsible for the lack of expression of these genes in the compromised clones.

The only exception to this was ESC(Comp)7 which has approximately 50% methylation at the IG-DMR, similar to the controls. In Chapter 5 (Transcriptomics) we have seen that this clone fails to give rise to viable chimaeras and also did not express *Rian* and *Mirg*, but *Gtl2* expression is not significantly different from the ESC(Ctrl) clones. Together these results suggest that for

this clone there may be a different cause of aberrant silencing of *Rian* and *Mirg*. It is possible that there is a deletion of the region containing *Rian* and *Mirg*, or another mutation that is for example causing a stop codon at the wrong position. This would be interesting to analyse but it is beyond the scope of this thesis.

Despite this, our hypothesis that this region is important for proper embryonic development still holds.

6.3.2 Loss of Imprinting at Additional Imprinted Loci

We have found that additional imprinted genes are also hyper- or hypomethylated and/or their expression is increased or decreased in a number of our ES cell clones. This is not just restricted to the compromised clones. However, this result is not particularly surprising as it has been shown previously that the epigenome in ES cells is extremely unstable [86, 128] and loss of imprinting is a very common phenomenon in ES cells. Changes in the expression of imprinted genes can often be observed in cloned mice even if they were derived from the same ES cell clone. However, most of these changes do not seem to affect viability and Humpherys et al. [128] argue that embryonic development is very tolerant to changes in the epigenome. This would also explain why we see these kinds of changes in our ESC(Ctrl) clones without observing an abnormal phenotype.

Despite this, misregulation of the *Dlk1-Dio3* region seems to be detrimental for development. It is likely that hypermethylation of the IG-DMR was caused by loss of imprinting during ES cell culture. This has also been shown by Stelzer et al. [125]. They observe a small population of cells in their ES cell cultures that is either hyper- or hypomethylated at the *Dlk1-Dio3* IG-DMR. All of our cell lines (apart from the parental ESC(Ctrl)1 cell line) were derived by subcloning. It is possible that our compromised cell lines were derived from cells that happened to be hypermethylated at the *Dlk1-Dio3* IG-DMR.

Loss of Imprinting in ESC(Ctrl)4

The ESC(Ctrl) 4 clone is unusual. It was selected as a control cell line because in the past it had given rise to excellent chimaeras. However, in our present study this ability seems to have gotten lost and its phenotype is more like that of the compromised clones in that we observed embryonic lethality in ESC(Ctrl)4 chimaeras. Additionally, its transcriptome and methylation profile of the *Dlk1-Dio3* region is almost identical to that of the compromised cell lines. We believe that this is also due to loss of imprinting at the *Dlk1-Dio3* locus. This cell line most likely used to be a good control, but over passages compromised cells may have accumulated to a point where most of the cells have this change. It is possible that a number

of those hypermethylated cells were already present at the time of the successful blastocyst injections. However, their proportion may have been low enough that only "normal" cells were injected by chance.

6.3.3 *Gtl2* Expression Rescue

As described in Chapter 1 ([Introduction](#)), correct DNA methylation is extremely important during mammalian development. It is part of the cells' mechanism to dynamically regulate gene expression profiles and developmental processes. Most importantly, we also know that DNA methylation is a reversible modification of the genome. We therefore argued that it should be possible to reverse the hypermethylation and hence rescue gene expression at the *Dlk1-Dio3* region in the ESC(Comp) clones.

Culture in 2i or EPSCM

Initially, we tried rescuing expression by culturing the cells in 2i media [129] and EPSCM (Expanded potential stem cell media) [127]. 2i is routinely used in a number of labs to culture ES cells as it has been shown that 2i-cultured ES cells are more representative of inner cell mass cells of the blastocyst. EPSCM was introduced very recently as a means to establish cultures of so called expanded potential stem cells that have the ability to differentiate into and contribute to not only the embryonic lineages but also extraembryonic tissues. These can be established from individual blastomeres, ES cells or iPS cells. Both of these media have previously been shown to decrease global DNA methylation levels in ES cells [126, 127].

Culturing the ESC(Comp) clones in 2i media or EPSCM was unsuccessful in rescuing the expression of *Gtl2*. In hindsight, this result was not very surprising since Leicht et al. [126] also found that although global DNA methylation decreases with 2i culture, imprints are not usually affected by this. However, it was later also shown that prolonged culture in 2i actually does indeed affect and reduces DNA methylation at imprinting control regions [130]. So perhaps, three passages were not a long enough exposure to reduce methylation of the *Dlk1-Dio3* IG-DMR and reactivate *Gtl2* expression. In their work on expanded potential stem cells, Yang et al. [127] have only focussed on global DNA methylation and do not comment on DNA methylation levels of imprinting control regions.

Treatment With 5-azacytidine

We successfully managed to reactivate expression of *Gtl2*, *Rian* and *Mirg* in the compromised clones using 5-azacytidine. This is a DNMT inhibitor that reduces global methylation levels. After uptake into the cell, it becomes metabolised to 5-aza-dCTP which can be incorporated

into the DNA by the DNA replication machinery instead of cytosine. DNMTs recognise the forming azacytosine-guanine dinucleotides and start a methylation reaction. Unlike the bonds between DNMTs and cytosine, however, the covalent bonds formed between DNMTs and azacytosine are irreversible and the DNMTs become trapped and will eventually be depleted. This leads to global reduction in DNA methylation.

Interestingly, none of the ESC(Comp)3 subclones that were derived after 5-aza treatment, expressed all three genes *Gtl2*, *Rian* and *Mirg*. Several were identified that had reactivated *Gtl2* expression but not that of *Rian* and *Mirg*. This is particularly interesting since they are all believed to be regulated as one large polycistronic transcription unit. The cause for this result remains to be elucidated.

When the 5-aza "batch-treated" cells were injected into blastocysts, a number of quite low percentage chimaeras were born. However, the subcloned cells mostly gave rise to pups with no contribution from the ES cells and two extremely low percentage chimaeras. This suggests that *Gtl2* is likely necessary for proper embryonic development but is not sufficient on its own without *Rian* and *Mirg*.

An explanation for the low contributions could also be the 5-aza treatment itself. Although it did reactivate gene expression, there are issues with this method. 5-azacytidine is very toxic for the cells since it interferes with the DNA replication machinery. It is therefore possible that the cells were damaged and although this may not be noticeable immediately it could lead to long-term effects on viability and cell function. This could lead to the cells dying or unhealthy cells being excluded from the inner cell mass of the developing embryo.

Alternative Ways to Rescue Expression of Maternally Expressed Genes From the *Dlk1-Dio3* Region

Although there is very strong evidence for a correlation between aberrant regulation of the *Dlk1-Dio3* region and the embryonic lethal phenotype that we observe, as mentioned previously we cannot exclude other or additional causes entirely. To prove this - and to avoid toxic effects of 5-aza treatment - we would need to decrease methylation and rescue expression exclusively at the *Dlk1-Dio3* locus. One way to do this could be using CRISPR-mediated DNA demethylation by TET (Tet methylcytosine dioxygenase) enzymes. This method was recently reported by several groups [131, 132]. They make use of the CRISPR-Cas9 genome editing machinery as well as the DNA demethylation action of TET enzymes. Specifically, they have fused the TET catalytic domain (TET-CD) to a catalytically deactivated Cas9 protein and designed appropriate guide RNAs to target this complex to specific genes that are known to be hypermethylated in their cell type of interest. The TET-CD then acts to specifically

demethylate neighbouring CpGs which subsequently leads to activation of gene expression at that locus. This system could be utilised to alter gene expression at the *Dlk1-Dio3* region by using guide RNA that will chaperone the Cas9 with the fused TET-CD to the IG-DMR. A potential problem with this, however, is that this mechanism is not able to distinguish between the maternal and the paternal strand and would therefore likely demethylate both alleles, most certainly resulting in higher than normal gene expression. And as Stelzer et al. [125] show, 4n complementation embryos from ES cells with hypomethylation at the *Dlk1-Dio3* IG-DMR are not viable. However, this region is rich in single nucleotide polymorphisms (SNPs) and since the ES cells are from a mixed genetic background it may be possible to design a guide RNA that will preferentially bind to the maternal allele. Allele specific genome editing has been shown to be successful [133], but may be more complicated in this case as the fused TET-CD would potentially still be able to reach and demethylate CpGs on both alleles.

6.3.4 Chapter Conclusion

In this chapter, we showed that the IG-DMR in the *Dlk1-Dio3* imprinted region, which controls expression of the maternally expressed genes, is hypermethylated in the ESC(Comp) ES cells. Hypermethylation of the IG-DMR is likely the cause for the silencing of *Gtl2*, *Rtl1as*, *Rian* and *Mirg*. This is also supported by our data that shows that treatment of the cells with the DNMT inhibitor 5-aza can rescue expression of *Gtl2*, *Rian* and *Mirg*. Although gene expression was increased in ESC(Comp) clones to levels comparable to that in ESC(Ctrl) cells, blastocyst injection of the 5-aza treated cells did not yield any high percentage chimaeras. More work is required for this experiment to specifically demethylate the maternal allele of the IG-DMR only and/or to reduce toxicity effects of 5-aza, in order to rescue not only the *Gtl2*, *Rtl1as*, *Rian* and *Mirg* gene expression levels in the ESC(Comp) cells but also their developmental potency.

Chapter 7

Phenotypic Screening by Phage Display

7.1 Introduction

The work presented in this chapter was carried out during my industrial placement at Medimmune Cambridge which was part of the funding body requirements. Phenotypic screening using phage display technology was performed to look at differences in cell surface markers between the ESC(Ctrl) and ESC(Comp) cells. Similarly to the transcriptomics experiments described in Chapter 5 ([Transcriptomics](#)) the goal was to potentially identify a diagnostic marker to distinguish between ESC(Ctrl) and ESC(Comp) cells. Unlike RNA sequencing, however, this method has the potential to identify changes in surface protein composition as well changes in other cell surface molecules such as lipids, carbohydrates or post-translational modifications of cell surface proteins. These play an important role in cell adhesion, cell signalling and interaction between cells.

7.2 Introduction to Phage Display

Phage display was first demonstrated by George Smith in 1985 [134]. He successfully managed to display foreign peptides on the surface of filamentous bacteriophages by fusing them to one of their coat proteins. This discovery opened the door to a wide range of applications for phage display, including the production of large quantities of antibodies and peptides, drug discovery, medical diagnosis and many more.

7.2.1 Filamentous Bacteriophages

The most commonly used phage display systems are the *E. coli* filamentous bacteriophage (FB) strains M13 and f1. These FBs belong to a group of bacterial viruses that infect gram-negative

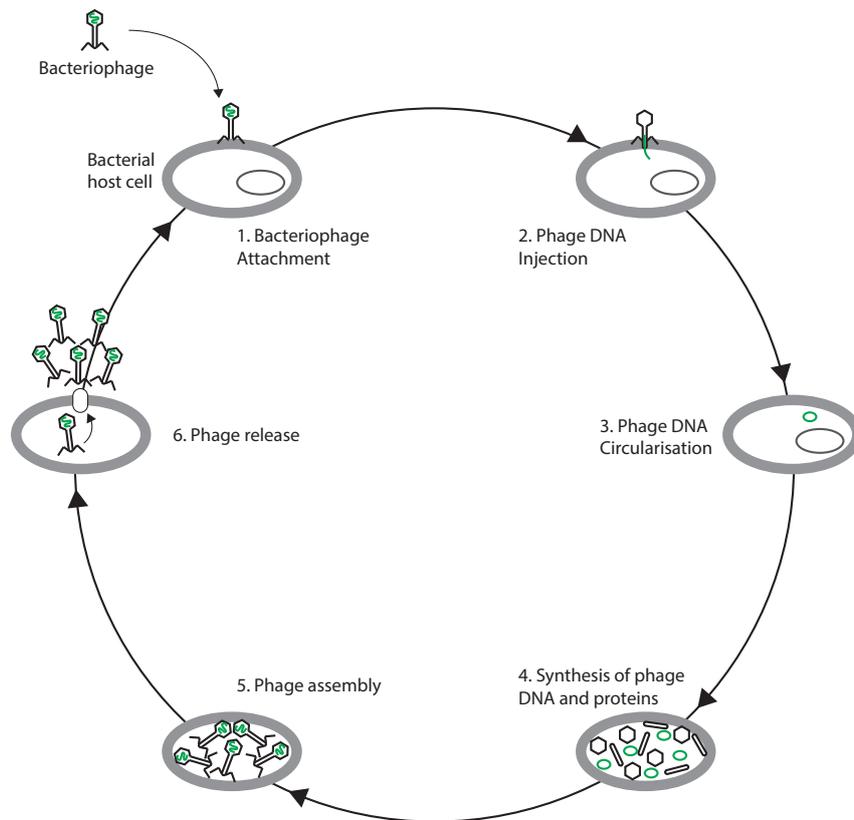


Fig. 7.1 **Bacteriophage life cycle** After a bacteriophage attaches to a bacterial host cell, it injects its DNA which is then replicated and used for phage protein synthesis using the host's DNA replication and protein synthesis machinery. The different phage proteins are then assembled within the host cell and released through a channel in the outer cell membrane that is encoded by one of the phage genes.

bacteria. Instead of inducing a lytic infection, they use their host's replication machinery to produce and secrete new phage particles. Upon infection with circular single-stranded DNA, the host polymerase converts it to the replicative double stranded form which then serves as the template for DNA replication as well as transcription and translation of the 11 phage genes. Capsid and assembly proteins are located in the bacterial inner cell membrane where assembly will take place, whereas the replication proteins remain in the cytoplasm. Through a channel in the outer membrane, the assembled phage particles are secreted out of the cell, ready to infect another bacterial cell. This bacteriophage life cycle is illustrated in Figure 7.1.

7.2.2 Antibody Structure

Antibodies are a vital part of the body's immune system as they have the ability to identify, bind to and neutralise foreign objects, called antigens. Each antibody has binding specificity to

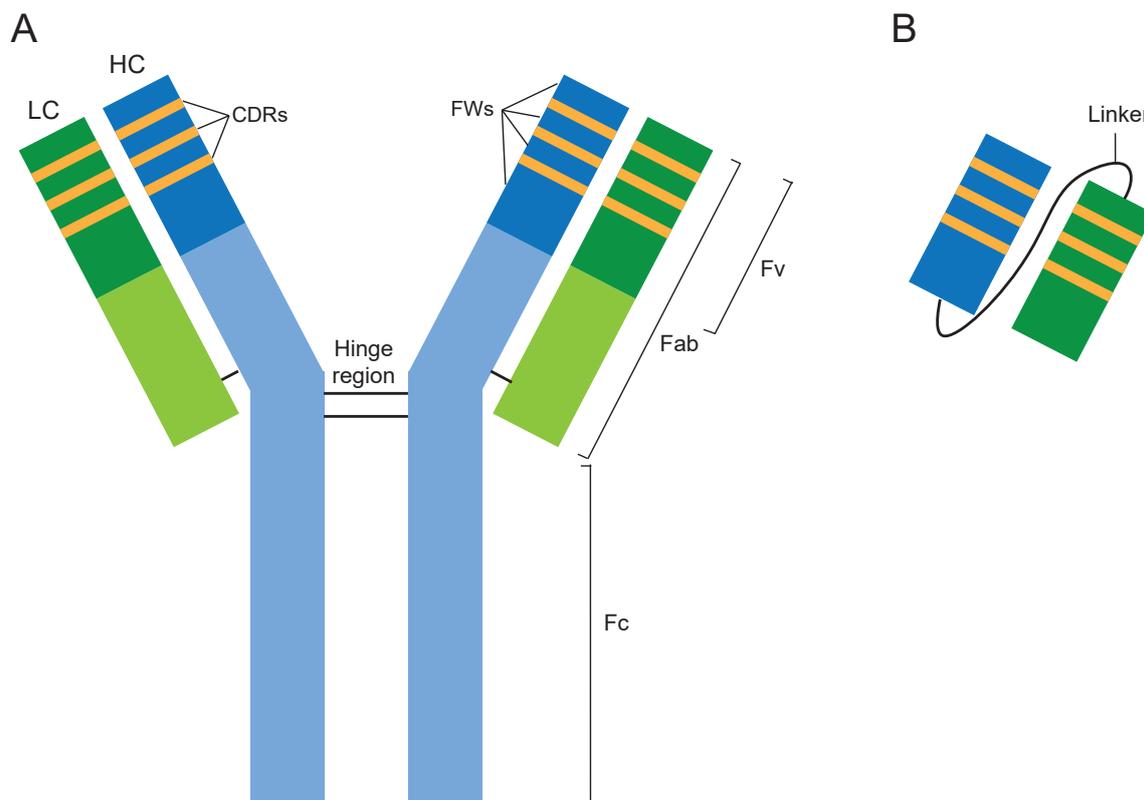


Fig. 7.2 **Antibody structure** (A) Structure of a full length antibody. The two heavy chains (HC) are shown in blue and the two shorter light chains (LC) are shown in green. The darker shaded ends are the variable regions (Fv) of each the HCs and LCs. Each variable region has three complementarity determining regions (CDRs) and four framework regions (FWs). The two heavy chains are held together by a disulfide bond in the hinge region. (B) Structure of a single chain variable fragment (scFv) containing only the variable regions of the heavy and light chains, held together with a polypeptide linker.

a unique antigen. This binding specificity as well as their versatility and biological role can be explained by looking at the antibody structure.

Antibodies all share the same basic structure which was discovered by Rodney Porter and Gerald Edelman in the 1950s and 1960s [135–137]. Using dithiothreitol and iodoacetamide, chemical agents that break and inhibit the formation of disulphite bonds, as well as denaturing agents to disrupt noncovalent interactions, Edelman found two types of subunits with different sizes which he called heavy and light chain, according to their molecular weight. Porter used the proteolytic enzyme papain to fragment rabbit immunoglobulin G (IgG). This produced three fragments, two of which had the same size and charge. The two identical fragments still retained their antigen binding capacity and he therefore called them fragments of antigen

binding (Fab). The other fragment did not bind antigen and due to its crystallising nature was called fragment crystallisable (Fc). Porter later discovered that the Fab fragments are composed of parts of the heavy chain as well as the entire light chain, whereas Fc fragments only contain heavy chain. From this he concluded that antibodies must have a Y-shaped structure as shown in Figure 7.2A.

We now know that both the heavy and the light chain can be further divided into a constant (C) region at the C-terminal end and a variable (V) region, consisting of 110-130 amino acids at the N-terminal end. Differences in the C region mainly determine the immunoglobulin isotype, whereas antigen specificity is determined in the V region, specifically three hypervariable regions called complementarity determining regions (CDRs). There are three CDRs (CDR1, CDR2 and CDR3) on each chain and they are separated by so called framework residues. These are responsible for the correct positioning of the CDRs at the surface of the antibody's 3D structure, thereby enabling interaction of CDRs and antigens.

7.2.3 Antibody Fragments for Modern Applications

In the past few decades it has become clear that the Fc region is not needed for a lot of applications and the use of fragments lacking the Fc domain has become more and more common. This has been facilitated by recombinant antibody technologies and genetic engineering which make it possible to produce several different types of antibody fragments.

The first known antibody fragments were Fabs. Since their discovery, they have been engineered into even smaller fragments. The most commonly used ones are single chain variable fragments (scFvs). scFvs only consist of the variable regions of the heavy and light chains, held together with a polypeptide linker (Figure 7.2B). These are very easy to express in *E.coli* and are therefore widely used in applications such as phage display.

7.2.4 Construction of scFv Phage Libraries

In order to construct a high quality phage library, a large number of very diverse antibody genes is required. These come from B-lymphocytes that can be found in the spleen, bone marrow or blood. It is generally distinguished between immune [138], naïve [139] and synthetic [140] phage libraries. Immune libraries employ the immune system's ability to produce specific antibodies in large quantities when presented with antigens, and thereby to skew the diversity of its own antibody repertoire. They are made from B-lymphocytes of immunised animals or human donors, meaning that they have previously been presented with an antigen of interest or that the patient has been affected by a disease. This can have the advantage that the resulting phage library is naturally enriched for antibodies that are specific to the antigen of interest.

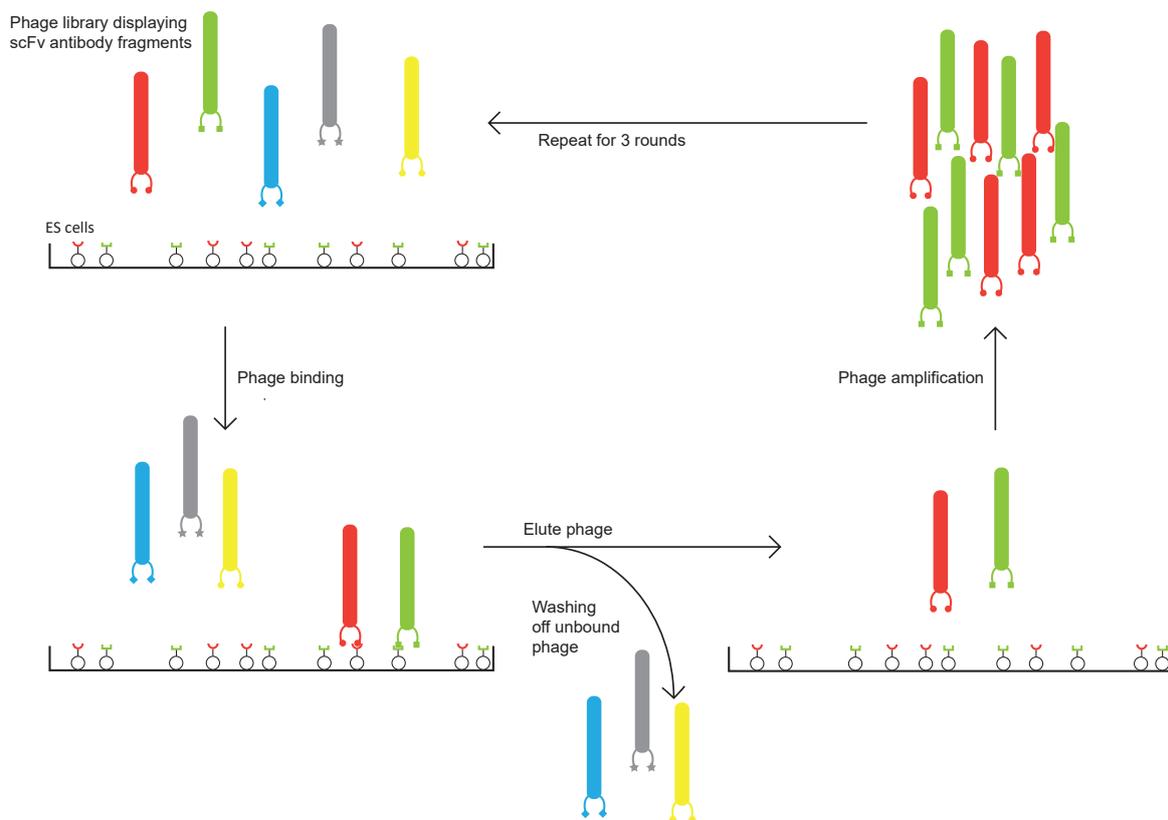


Fig. 7.3 **Biopanning** Diverse phage library is added to the cells of interest. Some of the phage will bind to the cells, the rest is washed off. Bound phage is eluted, amplified and used for another round of selection. After several rounds the phage outputs will be enriched for phage binding to the cells of interest.

Naïve libraries are derived from B-lymphocytes of non-immunised donors. They are not biased towards a limited number of different antibodies, but in fact are designed to have a very high diversity. Naïve libraries are constructed by isolating RNA from B-lymphocytes and reverse transcribing mRNA into cDNA. Using different primers, VH and VL chain genes are amplified separately and then combinatorially assembled into scFv fragments using a DNA linker to create very large arrays of diverse antibody fragments. They can then be further amplified and ligated into phagemid vectors, for transformation and growth in *E. coli*.

Similar to naïve libraries, synthetic libraries can in theory contain antibody fragments to any possible antigen. However, they are created outside of the immune system. They have the advantage that it is possible to control the phage contents as well as the variability and diversity of the library.

7.2.5 Biopanning

Phage libraries can contain up to 10^{12} phage particles, each of them displaying a unique antibody fragment. During a process called biopanning which is illustrated in Figure 7.3 the number of unique phage particles can be greatly reduced and the pool is instead enriched for clones that bind to antigens of interest.

The phage library is added to a given target such as immobilised antigens or cells. After removing any unbound phage, all bound phage is eluted, amplified in *E. coli* and used for another round of biopanning. This process is usually repeated two to four times and leads to an enrichment of the phage library for clones specific to the target.

7.2.6 Applications of Phage Display

Phage display has a number of diverse applications. Particularly notable, however, is the major role that phage display of antibody fragments has had in the latest advances in drug discovery. Up until recently, drug discovery heavily relied on the identification of target genes through molecular biology techniques. However, finding suitable antibodies or small molecules that bind to these and have the desired effect, can be very laborious. With the invention of antibody phage display, this process was simplified and sped up as this technique allows for the screening of a vast number of antibody fragments based on their functional activity rather than their ability to bind to a specific target. Only after finding antibody fragments with the desired phenotype, their binding partners will be identified.

7.3 Phage Display and Why It is Useful in This Project

Cell signalling plays an important role during embryonic development and is necessary for the proper regulation of developmental processes [141]. Signalling events between cells often involve the engagement of cell surface proteins. With the RNA sequencing experiment described in Chapter 5 (Transcriptomics) we did not identify any changes in RNA coding for cell surface proteins. However, with phage display we are able to identify changes not only in cell surface proteins but also in lipids, carbohydrates and post-translational modifications of proteins such as glycosylation or phosphorylation (a summary of examples of this can be found in the 2014 review by Chan et al. [142]). These also play important roles in cell signalling, cell adhesion and generally in the interaction between cells. We therefore hypothesise that there may be a difference in any of those cell surface molecules between the ESC(Ctrl) and the ESC(Comp) cell populations that causes the embryonic lethal phenotype in the ESC(Comp) chimeras, assuming that it is continuously expressed after ES cell differentiation.

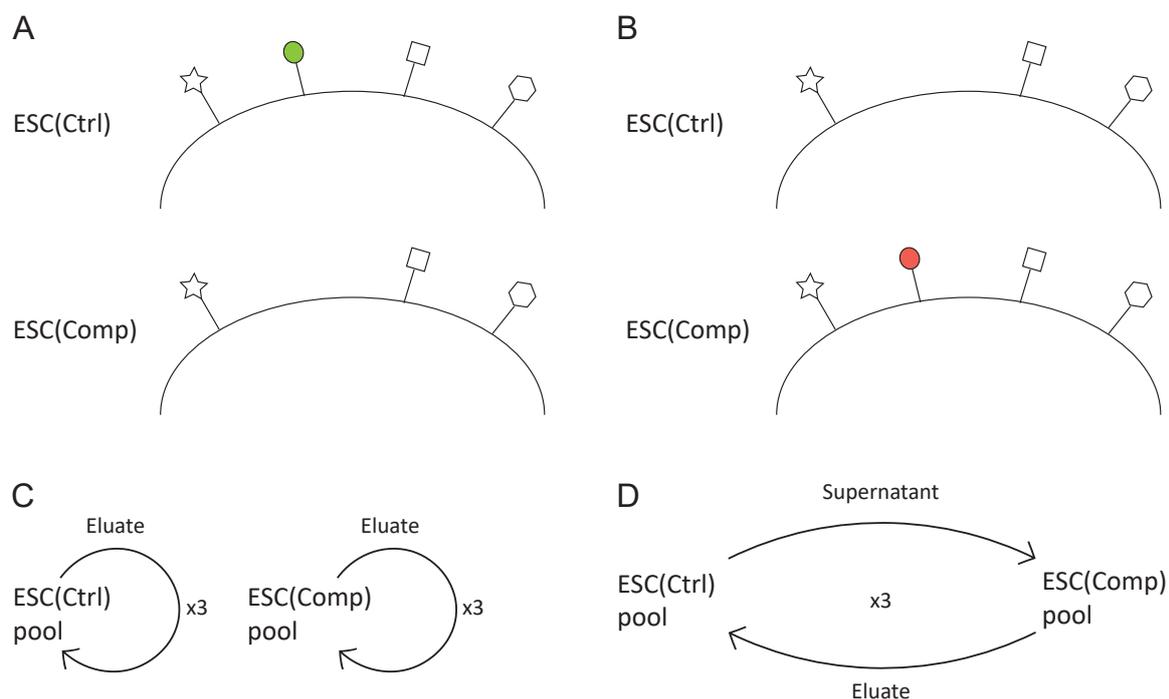


Fig. 7.4 **Experimental setup** (A) Hypothesis 1: The ESC(Comp) cells are missing a cell surface marker (shown in green) that is vital for embryonic development. (B) Hypothesis 2: The ESC(Comp) cells overexpress a cell surface marker (shown in red) that is causing the embryonic lethal phenotype. (C) Experimental setup 1: three rounds of selection on the pool of ESC(Ctrl) cells and the pool of ESC(Comp) cells. (D) Experimental setup 2: three rounds of deselection to select for phage binding specifically to the ESC(Comp) cells.

7.4 Experimental Setup

We were interested in differences between the two ES cell populations - ESC(Ctrl) and ESC(Comp) - in terms of their cell surface and we had two hypotheses what the nature of these differences could be: firstly, there may be a cell surface marker that is vital for proper cell function but that is missing from the ESC(Comp) clones (Figure 7.4A), or secondly there may be a cell surface marker that is expressed on the ESC(Comp) but not the ESC(Ctrl) cells that is contributing to the embryonic lethal phenotype (Figure 7.4B). To investigate these, two different experimental setups were used. For both of these, the cells were pooled into the two groups (ESC(Ctrl) and ESC(Comp)) so that we were dealing with only two cell populations rather than four ESC(Ctrl) and four ESC(Comp) cell lines. In the first experiment, we performed three rounds of selection on the two pools of cells separately (Figure 7.4C). The other one was a deselection experiment that allowed us to specifically select for phage that binds only to the ESC(Comp) cells. Deselection involves taking the supernatant (anything that did not

bind) of ESC(Ctrl) cells, adding it to the ESC(Comp) cells and then selecting for those phage that bind to the ESC(Comp) cells (Figure 7.4D). Three rounds of deselection were carried out. From now on we will call the different setups Ctrl, Comp and DS (Deselection).

The deselection experiment could have been done the opposite way to specifically select for phage that binds exclusively to the ESC(Ctrl) cells in order to identify surface markers that are missing from the ESC(Comp) clones. However, this requires all the ESC(Comp) cells to be missing the same surface marker. As soon as one of the ESC(Comp) clones is different to the other three, this difference could be masked. Since the cells were pooled and we could not be entirely certain that the ESC(Comp) cells are all compromised in the same way, we would potentially not be able to pick up any changes like that and decided to focus on identifying overexpression changes in the ESC(Comp) cell pool as shown in Figure 7.4B.

Unfortunately, the Ctrl pool also contains the ESC(Ctrl)₄ cell line since these experiments were run in parallel with the RNAseq study. Therefore, any results coming out of this study need to be carefully validated as the results will likely be skewed and subtle differences between the two groups might be masked.

7.5 Results

7.5.1 Three Rounds of Cell Surface Selection and Selection Rescue

We performed three rounds of selection and selection rescue for both the two normal selections and the deselection, using a combination of the CS (Combined Spleen [143]) and BMV (Bone Marrow Vaughan [144]) phage libraries, in order to enrich for phage that binds to the different cell populations. For each round, the selection output was monitored and after the second and the third rounds a number of phage containing *E. coli* colonies were picked and sequenced to check the quality of the outputs. Table 7.1 shows that with subsequent rounds the output size (number of colony forming units (cfu)) increased. This is a good sign that the selections have worked and we have indeed enriched for phage that binds to the respective cells of interest. Interestingly, the output of the deselection was smaller than that of Ctrl and Comp, especially after the first round. This might be because by the time the phage is added to the ESC(Comp) cells, a lot of it will already have bound to the ESC(Ctrl) cells in the previous step, and therefore the effective size of the phage library was smaller than for the Ctrl and Comp selections.

In order to check the quality of the outputs and to determine which output to take forward to some of the next experiments, DNA sequencing was performed on a number of colonies and the sequences analysed using Blaze2. Blaze2 is a software package that was developed by Medimmune and that automatically recognises specific antibody sequences and annotates

CDR1-3 and framework regions. It is also able to detect small changes in the sequences, such as frameshift mutations and stop codons, as well as large truncations or missing sequences. We looked at the number of unique functional sequences for each group after rounds two and three, as well as the number of sequences with stop codons, frameshift mutations and truncations (Table 7.2). While the number of sequences with frameshift mutations and stop codons were quite constant from round two to round three, there was a clear increase in the number of truncated sequences, particularly during deselection, and a subsequent decrease in the number of unique functional sequences. We therefore decided to use the round two outputs for the reformatting to IgG format described later in this chapter.

Table 7.1 **Output sizes of the three selection rounds**

	Controls	Compromised	Deselection
Round 1	2×10^3 cfu/ml	2×10^3 cfu/ml	1×10^3 cfu/ml
Round 2	7×10^4 cfu/ml	6×10^4 cfu/ml	7×10^4 cfu/ml
Round 3	3×10^7 cfu/ml	3×10^7 cfu/ml	2×10^7 cfu/ml

Table 7.2 **Blaze2 results of round two and round three outputs** The numbers show absolute numbers (and %) of sequences.

	Controls		Compromised		Deselection	
	Round 2	Round 3	Round 2	Round 3	Round 2	Round 3
Missing sequences	2 (4.2%)	1 (2.1%)	0	0	1 (2.1%)	3 (6.25%)
No/truncated heavy chain	3 (6.25%)	8 (16.6%)	6 (12.5%)	9 (18.75%)	7 (14.6%)	17 (35.4%)
Frameshift	8 (16.6%)	6 (12.5%)	12 (25%)	9 (18.75%)	4 (8.3%)	7 (14.6%)
Stop codon	5 (10.4%)	8 (16.6%)	5 (10.4%)	2 (4.2%)	1 (2.1%)	2 (4.2%)
Functional sequences	30 (62.5%)	25 (52.1%)	25 (52.1%)	28 (58.3%)	35 (72.9%)	19 (39.5%)
Unique sequences	23 (47.9%)	16 (33.3%)	18 (37.5%)	14 (29.2%)	26 (54.2%)	13 (27.1%)

7.5.2 Next Generation Sequencing of Selection Outputs

The sequencing of a number of phage colonies gave us a good initial overview of the quality of the selection outputs. However, we were interested to analyse both the initial phage library

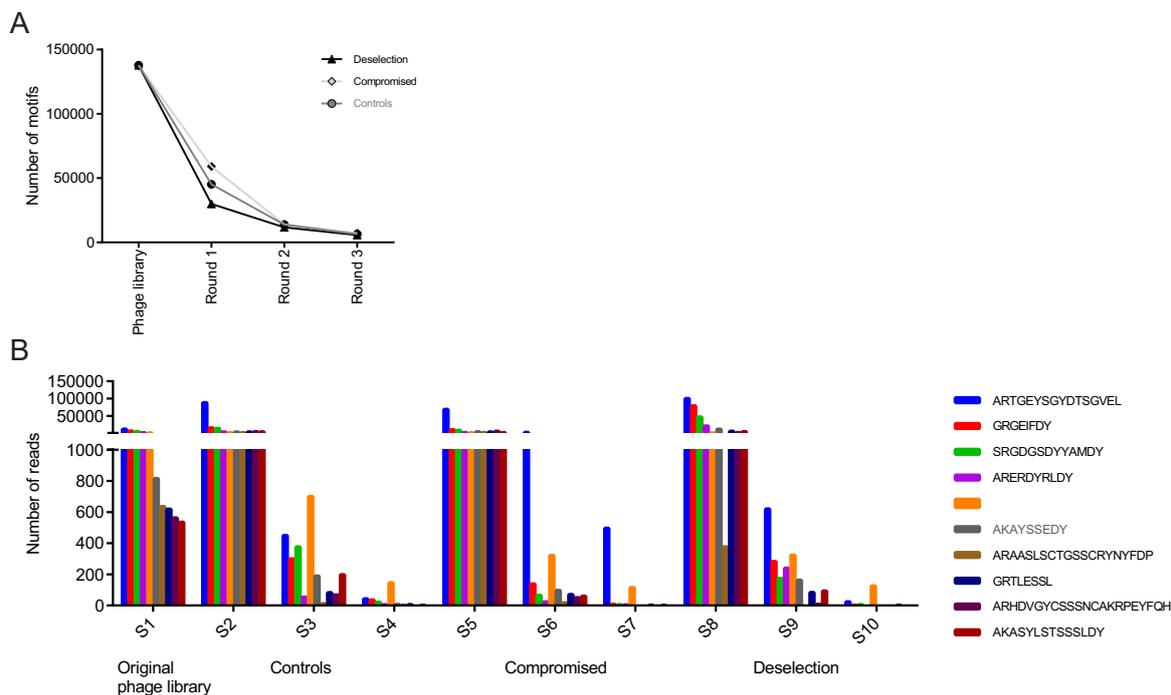


Fig. 7.5 Library diversity and dominant clones (A) The diversity of the original phage library drops each round (B) Graph displaying the ten motifs with the highest read count in the original library and their read counts in the subsequent round one to three outputs. There are a few dominant clones in the original library. These persist in the round one outputs but disappear in rounds two and three.

as well as all the selection outputs in more detail and performed next generation sequencing (NGS) on them. To do this, plasmid DNA was extracted from glycerol stocks of all round one, two and three outputs as well as the initial phage library. The scFv region of the phagemid populations was PCR amplified and sequencing was performed in the Department of Biochemistry, University of Cambridge. The raw data was analysed by Dr M. Samborsky. In brief, the sequences were mapped against a scFv processing pipeline that automatically annotates CDR1-3 and framework regions. After exclusion of out of frame sequences, the occurrence of each CDR3 motif was counted and its relative frequency was calculated.

Firstly, we were interested in the diversity of the samples. Figure 7.5A shows that - as we expected - the number of different motifs dropped significantly from the original phage library to round one and also to the subsequent rounds. This shows that the selection has worked and we have indeed enriched for specific phage during the three selection rounds. Additionally, we wanted to find out whether there were any dominant clones (phagemids that are overrepresented) that were already present in the original library and what happened to them during

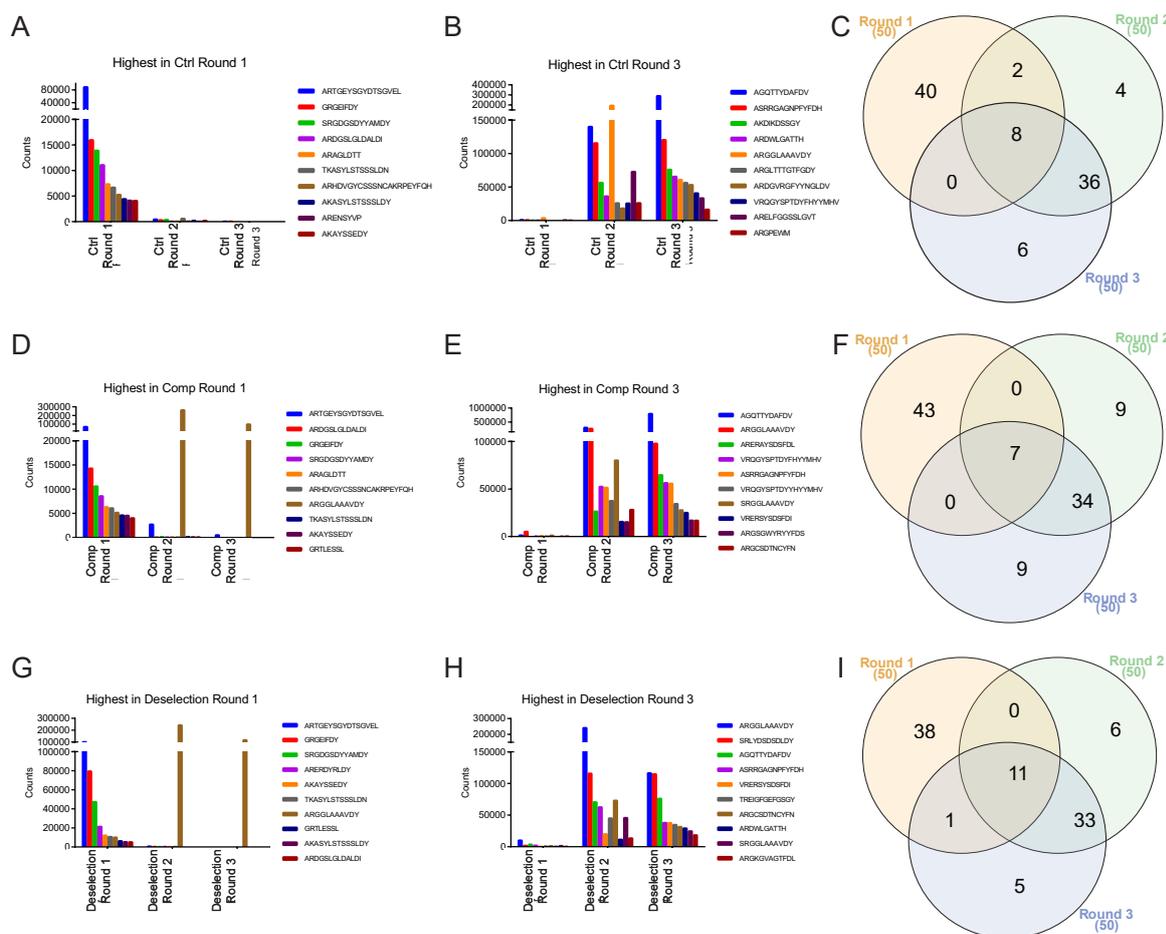


Fig. 7.6 Evolution over selection rounds Top ten motifs in rounds one and three for Ctrl (A, B), Comp (D, E) and Deselection (G, H) outputs and their counts per round. (C, F, I) Venn diagrams of the top 50 motifs each round for Ctrl (C), Comp (F) and Deselection (I).

selection rounds as the presence of dominant clones could potentially skew the results. As can be seen in Figure 7.5B there were indeed a number of sequence motifs that occur repeatedly in the original library (the majority of motifs occurs only once). They still seemed to persist in the round one outputs, but the occurrence dropped in rounds two and three and therefore the presence of dominant clones does not appear to be problematic for the further analysis.

To analyse what the "evolution" was from round one to round three, we looked at the top 50 motifs from each output and compared them to each other. The Venn diagrams in Figure 7.6 show that round two and three outputs were much more similar to each other than either of them were to the round one outputs. This was the case for all selections (Ctrl, Comp and Deselection). Approximately 35 of the top 50 motifs were shared only between the round two and the round three output, whereas almost none were shared only between the round one output and either the round two or round three outputs. The top 10 motifs of rounds one and three and their

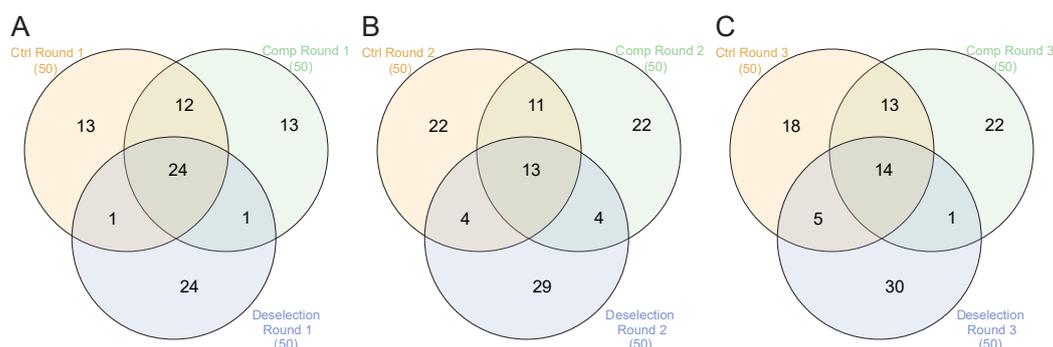


Fig. 7.7 Differences between Ctrl, Comp and Deselection outputs Venn diagrams comparing the top 50 motifs of Ctrl, Comp and Deselection outputs after round one (A), round two (B) and round three (C).

count numbers for each round are also displayed in Figure 7.6. Most interesting, however, was to investigate how similar or different the three outputs of each round were to each other and whether we can identify one or more sequence motifs that can distinguish between ESC(Ctrl) and ESC(Comp) cells. These data are displayed in Figure 7.7. We again looked at the top 50 motifs for each output. In the round three outputs, 14 motifs occurred in all three, five were shared only between Ctrl(3) and Deselection(3), 13 were shared only between Ctrl(3) and Comp(3), and only one motif was common to both Comp(3) and Deselection(3) but not Ctrl(3). Interestingly, Comp(3) was more similar to Ctrl(3) than to Deselection(3). The results for the round two outputs were very similar with 13 motifs shared between all three outputs, four shared between only Ctrl(2) and Deselection(2), eleven shared between only Ctrl(2) and Comp(2) and four shared between only Comp(2) and Deselection(2). After only one round of selection the outputs were still more similar to each other. However, we can start to see that the deselection outputs were more distinct.

7.5.3 scFv Reformating to IgG Format

The NGS data showed that the round two and round 3 outputs were very similar and therefore justifies using the round 2 outputs for further characterisation. In order to confirm that any of the phage that were enriched during the rounds of cell surface selection differentially bind to one cell type or the other, the scFv fragments need to be reformatted into full length IgG antibodies. A group at Medimmune has recently developed a high throughput method for scFv to IgG reformatting [76]. In brief, this is done in two main steps: Firstly, the scFV phagemid pool is linearised by PCR amplification and by In-Fusion cloning a donor vector supplies the heavy chain constant region, the hinge region and the light chain control elements. In a second

step the incomplete IgG cassette of this intermediate pool is again amplified in a PCR reaction and cloned into an IgG vector containing the heavy chain control elements as well as the λ and κ light chain constant regions. This results in a complete IgG cassette coding for the entire heavy chain and light chain (Figure 7.8).

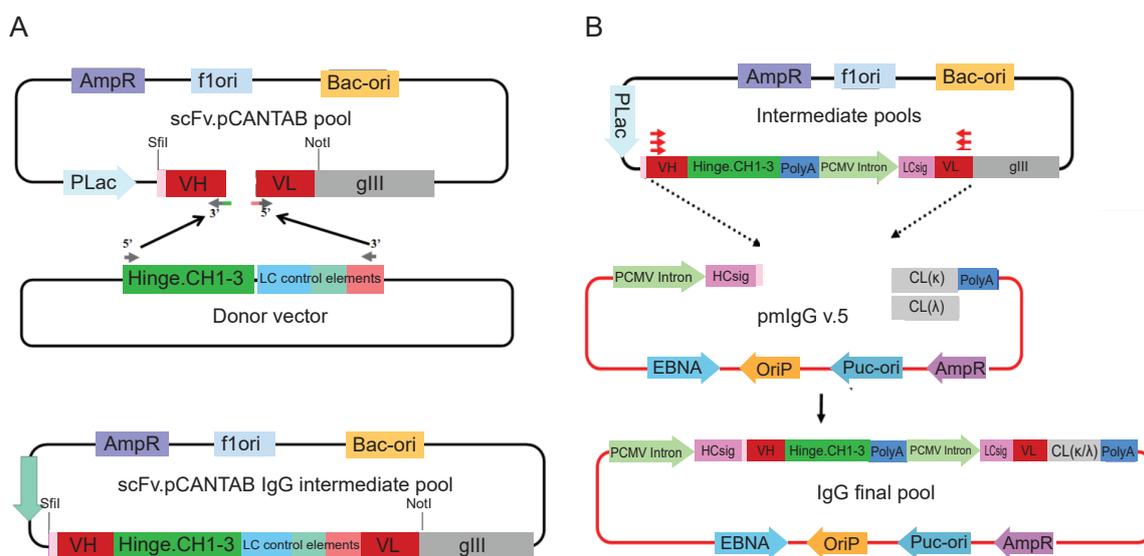


Fig. 7.8 scFv to IgG reformatting (A) To create the intermediate pool, the phage display vector is linearised by emulsion PCR amplification and fused with a donor fragment containing the hinge region, the heavy chain constant region and the light chain control elements. (B) The incomplete antibody cassettes from the intermediate pool are again amplified by emulsion PCR and cloned into a pmIgG v.5 vector containing the missing heavy chain control elements and light chain constant regions (λ/κ)

Intermediate Pool

Using plasmid DNA from round two outputs we performed an emulsion PCR to linearise the scFv.pCANTAB phagemid pool. Emulsion PCR has the advantage that individual sequences are separated into small water-in-oil droplets and therefore the chance of recombination between antibody sequences is reduced and correct VH-VL pairing can be maintained. This was done in two separate reactions for each population, in order to amplify both λ and κ light chains. In-Fusion cloning was then performed using the linearised phagemid vector and the donor vector supplying the heavy chain constant region, the hinge region and the light chain control elements. This formed the intermediate pool which was transformed into Stellar competent cells and colonies were grown over night. To check the quality of the intermediate pool we

again analysed the output size (Table 7.3) as well as DNA sequenced a number of colonies and analysed them using Blaze2 as described before (Table 7.4). The intermediate pools were in the range of $1\text{-}5 \times 10^4$ cfu/ml, which was in the expected range. Blaze2 analysis of a number of sequences from individual colonies showed that between 58-75% of sequences were unique and functional. We also checked that the cloning sites were in frame.

Final Expression Pool

Plasmid DNA was prepared from the intermediate pools and the incomplete IgG cassette was amplified using emulsion PCR. To prepare the final expression vector, the incomplete antibody cassettes from the linearised intermediate pools were cloned into linearised pmIgG donor vectors containing the heavy chain control elements and the λ and κ light chain constant regions. This final expression vector was transformed into Stellar competent cells and as before the output sizes (Table 7.3) and the quality of the final expression pool sequences were assessed (Table 7.4). The outputs were in the range of $4 \times 10^2\text{-}3 \times 10^3$ cfu/ml. This was on the low side but was sufficient for picking individual colonies for IgG expression. Blaze2 analysis showed that there was a reduction in the number of unique and functional sequences. However, the number was still sufficient to proceed to IgG expression.

Table 7.3 Output sizes during IgG reformatting

	Controls		Compromised		Deselection	
	λ	κ	λ	κ	λ	κ
Intermediate	3×10^4 cfu/ml	4×10^4 cfu/ml	2.5×10^4 cfu/ml	3×10^4 cfu/ml	3×10^4 cfu/ml	1.5×10^4 cfu/ml
Final	6×10^2 cfu/ml	3×10^3 cfu/ml	1.5×10^3 cfu/ml	4×10^2 cfu/ml	8×10^2 cfu/ml	8×10^2 cfu/ml

Table 7.4 Blaze2 results of intermediate and final pools

	Controls		Compromised		Deselection	
	Intermediate	Final	Intermediate	Final	Intermediate	Final
Missing sequences	3 (3.4%)	12 (13.6%)	7 (8.0%)	13 (14.8%)	3 (3.4%)	10 (11.4%)
No/truncated heavy chain	15 (17.0%)	21 (23.9%)	1 (1.1%)	47 (53.4%)	2 (2.3%)	34 (38.6%)
Frameshift	3 (3.4%)	0	6 (6.8%)	3 (3.4%)	7 (8.0%)	4 (4.5%)
Stop codon	13 (14.8%)	9 (10.2%)	10 (11.4%)	5 (5.7%)	5 (5.7%)	4 (4.5%)
Functional sequences	56 (63.6%)	49 (55.7%)	70 (79.5%)	25 (28.4%)	73 (82.9%)	40 (45.5%)
Unique sequences	51 (57.9%)	38 (43.2%)	66 (75.0%)	23 (26.1%)	65 (73.8%)	33 (37.5%)

7.5.4 High Throughput IgG Expression

For each population (Ctrl, Comp, Deselection) 400 individual colonies were picked from the final expression pool for high throughput IgG expression. Theoretically, each of those colonies should contain a unique IgG plasmid. These were submitted to the scFv-HTE team at Medimmune who performed IgG expression and quantification as described previously [76]. Next steps to analyse differential binding of those IgGs to either ESC(Ctrl) or ESC(Comp) clones are in progress.

7.6 Discussion

In this chapter we attempted to identify differences in cell surface markers between the ESC(Ctrl) and the ESC(Comp) cells using phage display. This technology can identify not just changes in protein composition but also changes in lipids, carbohydrates and post-translational modifications to proteins on the cell surface. The NGS data indicate that there may be differences between the two cell populations since the deselection output is quite distinct from the two normal selections. However, as described below, more work needs to be done to confirm whether there really are differences between the outputs in terms of their binding capacity and to identify what the cellular changes are.

Furthermore, there is one major flaw with the experimental design. In the previous chapters we have argued that the ESC(Ctrl)4 clone is unlikely to be a good control and should be excluded. However, since the selection experiments were performed in parallel with some of the work presented in previous chapters, it was still included in the ESC(Ctrl) pool. If there are indeed differences between the two cell populations in terms of their cell surface, this may be masked by the inclusion of ESC(Ctrl)4. To get a more precise analysis, the three rounds of selection would ideally be repeated with just ESC(Ctrl)1-3. However, the data may still be able to give us a good indication of whether there are any differences between the ESC(Ctrl) and ESC(Comp) cells.

7.6.1 NGS of Selection Outputs

The NGS data has confirmed that over the three rounds, selection for specific phage is happening as there is a reduction in the diversity of sequence motifs. The two biggest selection events seem to have occurred from the original phage library to round one and in round two. However, the outputs from rounds two and three are quite similar, which suggests that after two rounds we have already enriched for phage binding to the respective cell types and also justifies using the round two output for scFv to IgG reformatting.

Dominant clones in the original phage library that could have a growth advantage and therefore skew the results did not seem to be a concern for the analysis.

Comparing the three outputs of each round with each other, we would have expected the Comp and the Deselection outputs to be the most similar. However, the Ctrl and Comp outputs were more alike in terms of the top 50 motifs. This could be due to the type of selection. It is likely that just running three rounds of selection on both the ESC(Ctrl) cell pool and the ESC(Comp) cell pool will enrich for phage that is specific for ES cells in general, whereas deselection is more likely to pick up differences between the two cell populations and enrich for phage that binds specifically to the compromised ES cells. However, another explanation could also be that during deselection, a lot of the phage that will bind to the common surface markers will bind to the ESC(Ctrl) cells during the first part of the selection round and the pool of phage that is added to the ESC(Comp) cells is therefore depleted of those phage. This would increase the likelihood to enrich for phage that binds to less common surface epitopes during the second step of each deselection round and not necessarily just to those that are specific to the ESC(Comp) cells.

Nevertheless, we identified one CDR3 motif in the round three outputs and four CDR3 motifs in the round two outputs that exclusively occurred in the list of top 50 motifs of the Comp and Deselection outputs. It would be interesting to determine, whether these really do preferentially bind to the ESC(Comp) cells or whether this was by chance. Similarly, if we assume that deselection is better at identifying surface markers specific for ESC(Comp) cells than normal selection, all the sequence motifs only occurring in the top 50 list of deselection could potentially be of interest.

7.6.2 Next Steps and Future Work

To test the significance of these candidate motifs, a number of them could be synthesised and that sequence could then be cloned into the appropriate IgG vector similar to how it was described in section 7.5.3 of this chapter. These would then be used for a binding screen to test whether any of them differentially bind to either the ESC(Ctrl) or ESC(Comp) cells.

Additionally, we have reformatted the scFv phagemids from the round two outputs into full length IgG vectors and a number of these have been expressed by the HTE team at Medimmune. The next step here would also be a binding screen and this work is in progress.

Binding Screen

This is typically done by combining the expressed IgGs with a secondary fluorescent antibody and adding them to the two cell pools to allow for binding to occur. Using the high through-

put Mirrorball system (www.ttplabtech.com/products/Detection-Instruments/mirrorball/), the plates can then be imaged and analysed to potentially identify a small number of IgGs which differentially bind to either the ESC(Ctrl) or the ESC(Comp) cells and which can ultimately be used as a marker of ES cells with compromised (or normal) developmental potency.

Target Identification

If any IgGs that bind differentially to one group of cells are identified, target identification can be performed. Knowing that a phage or a full length IgG has the required phenotypic effect (in our case cell binding) does not tell us anything about the target that it is binding to. Target identification can be a lengthy process and is usually done using affinity purification followed by mass spectrometry.

7.6.3 Chapter Conclusion

In this chapter we describe the progress to date of the phage display phenotypic screen that was done with the aim to identify differences in cell surface markers between ESC(Ctrl) and ESC(Comp) cells. Two approaches were chosen, normal selection of both the ESC(Ctrl) and the ESC(Comp) cell pools and deselection which selects phage that preferentially binds to the ESC(Comp) cells. NGS results suggests that there may be a difference in terms of cell surface markers between ESC(Ctrl) and ESC(Comp) cells. However, the design of this experiment was flawed due to the inclusion of the potentially compromised ESC(Ctrl)⁴ clone and the results are therefore likely to be skewed. Hence, a number of validation steps are necessary to confirm or refute these results.

Chapter 8

Conclusion and Future Work

8.1 Overview of Findings

In this work we identified a number of ES cell clones with compromised developmental potency. When they were injected into e3.5 blastocysts, these cells mainly generated chimaeras that were not viable beyond mid to late gestation. At e13.5, there was no difference in viability between the ESC(Ctrl) and ESC(Comp) chimaeras but we already observed varying degrees of haemorrhaging in the ESC(Comp) chimaeras. At e17.5 their viability was significantly decreased and all of the dead chimaeric embryos were severely haemorrhaged. Some haemorrhaging was also observed in the live ESC(Comp) chimaeras. Histological analysis revealed defects in blood vessel integrity.

Initially, we analysed whether the ESC(Comp) cells still fulfilled all the characteristics of pluripotent stem cells - self renewal and the ability to differentiate into cells of the three germ layers. We were able to culture the cells for a large number of passages and they expressed a number of the core pluripotency markers, suggesting that they are able to maintain the pluripotent, self renewing state. Additionally, in a differentiation assay we showed that under the appropriate culture conditions the ESC(Comp) cells differentiated into cells of the mesoderm, endoderm and ectoderm lineages.

To investigate potential causes for the phenotype that we observed and also to identify a marker that can distinguish between ES cells with normal or compromised developmental potency, RNA sequencing was performed. This identified five genes that were silenced in the ESC(Comp) clones. Four of these genes - *Gtl2*, *Rtl1/Rtl1as*, *Rian* and *Mirg* - are located in the *Dlk1-Dio3* imprinted region of mouse chromosome 12 and are all maternally expressed. This finding was confirmed with additional ESC(Ctrl) and ESC(Comp) cell lines. We also

showed that *Gtl2* expression remained silent in tissues of e17.5 embryos. The fifth gene that was identified, *Gm27000*, is a predicted gene with unknown function. The role of this was not examined further.

Since gene expression at the *Dlk1-Dio3* locus is mainly regulated by a differentially methylated region (the IG-DMR) upstream of the *Gtl2* promoter, we performed bisulfite sequencing to analyse whether aberrant methylation was the cause for the gene silencing observed at this locus. Indeed, we found that the *Dlk1-Dio3* IG-DMR was hypermethylated. Other imprinted regions that were analysed also showed hyper- or hypomethylation of their respective DMRs, but for none of them there was such a clear correlation between phenotype and methylation profile. The expression of *Gtl2*, *Rian* and *Mirg* could be partially rescued using the DNMT inhibitor 5-azacytidine. Blastocyst injections of the 5-aza-treated cells gave rise to a small number of very low percentage chimaeras that survived to adulthood.

Finally, we also looked at the cell surface proteome using a phenotypic screen by phage display. This work is still in progress but NGS analysis of phage outputs indicated that there may be additional differences between the ESC(Ctrl) and ESC(Comp) cells. However, the data remains to be validated in order to determine whether these findings are real or whether they have occurred by chance.

8.2 General Discussion and Future Work - Relationship Between Phenotype and Genotype

8.2.1 Possible Role of Angiogenesis via VEGF Signalling

Not much is known about the precise role and mechanism of the maternally expressed genes of the *Dlk1-Dio3* region, but in recent years they have been linked with inhibition of angiogenesis and tumour suppression. A large number of human tumours show hypermethylation of the *Dlk1-Dio3* IG-DMR and consequently express very low levels of *MEG3* (the human homologue to *Gtl2*) including brain [145, 146], breast [147], liver [148], lung [149], bone [150] and kidney cancers [146]. This suggests a role of *MEG3* loss in the onset and progression of cancer. Indeed, it has been shown that *MEG3* interacts with and activates the tumour suppressor p53 (Tumour protein 53) [151]. When transfecting *MEG3* into a colon cancer cell line that does not express *MEG3*, functional p53 protein levels were significantly increased.

Gtl2/MEG3 downregulation or silencing has also been associated with increased angiogenesis in human tumours (angiogenesis is one of the hallmarks of cancer [152]) [147] but also in

diabetes [153], osteoarthritis [154], *Gtl2* knockout mouse models [155] as well as a rat model of tissue repair after ischaemic brain injury [156]. Angiogenesis is a multistep process that starts with the secretion of pro-angiogenic factors including vascular endothelial growth factor (VEGF), fibroblast growth factors (FGFs) or tumour necrosis factor α (TNF α). This leads to dilation of existing vessels, breakdown of the basement membrane and pericyte detachment, allowing existing vascular endothelial cells (VECs) to sprout from the original vessel and start forming new functional blood vessels. Pericytes are cells that wrap around the VECs to provide physical structure as well as important signalling molecules. After formation of a lumen, the vessel matures by synthesis of a new basement membrane and incorporation of pericytes and smooth muscle cells. A number of publications highlight the role of *Gtl2/MEG3* in these processes via a variety of different pathways. Gordon et al. [155] investigated the gene expression profiles of brains of *Gtl2* knockout mouse embryos. They found an increase in expression of angiogenesis related genes including *Vegf* and a VEGF receptor *Vegfr1* as well as genes in the Notch signalling pathway. In accordance with this finding, increased microvessel formation in the brain was also observed. Similarly, Zhang et al. [147] showed that proliferation and angiogenesis in breast cancers with low levels of *MEG3* can be inhibited by overexpressing *MEG3*, both *in vitro* and *in vivo*. The level of angiogenesis related genes were significantly decreased in a *MEG3*-low breast cancer cell lines after overexpression of *MEG3*. These did not only include the direct modulators of angiogenesis, *Vegf* and *bFGF*, but also *TGF- β 1* (Transforming growth factor β 1). *TGF- β 1* is an upstream regulator of VEGF, which suggests that *MEG3* acts through the *TGF- β 1* signalling pathway to modulate angiogenesis. Studies in diabetic mice have shown that diabetic retinopathies are often accompanied by microvascular dysfunction caused by a reduction in *MEG3* levels [153]. This was found to be regulated through the PI3K/AKT pathway, which is also upstream of *Vegf* expression. Overall, these studies suggest a role of *Gtl2/MEG3* in the inhibition of VEGF-mediated angiogenesis through a number of different pathways including the *TGF- β 1* and PI3K/AKT signalling pathways. In addition, p53 has also previously been shown to be able to inhibit expression of *Vegf* [157].

Interestingly, our ESC(Comp) chimaeras that do not express *Gtl2* and the other maternally expressed genes of the *Dlk1-Dio3* region, show haemorrhaging and a vascular phenotype which in part resembles some of the early steps of angiogenesis. We observed dilated vessels and poor vascular integrity with seemingly impaired basement membrane. Given that loss of *Gtl2/MEG3* expression led to enhanced angiogenesis via *Vegf* activation in a variety of pathologies and tissues, including mouse tissues and cell lines, we hypothesise that aberrant signalling via VEGF could at least in part have caused the phenotype that we observed. Increased angiogenesis can

lead to vascular leakage and subsequent haemorrhaging. More intriguingly, however, over-expression of *Vegf* has also been linked with increased occurrence of haemorrhaging [158–160].

We also noticed a striking resemblance between the phenotypes of the ESC(Comp) chimaeras and *Tie1* (Tyrosine kinase with immunoglobulin-like and EGF-like domains 1) mutant embryos. Both have severe haemorrhaging at e17.5 or 16.5 respectively and first signs of haemorrhaging can be observed around e13.5 [161]. TIE1 is a receptor that is part of the Angiopoietin/TIE signalling axis which is involved in vascular maturation and stabilisation during later stages of angiogenesis [162]. Apart from its general importance for vascular integrity, the precise role and mechanism of action of TIE1 are not fully understood [114, 163].

Given the similarity between our ESC(Comp) chimaeras and mouse embryos, which lack the expression of a gene that is so important for vascular maturation, we further hypothesise that aberrant silencing of *Gtl2* could lead to defects in vascular maturation and integrity.

These two hypotheses are also supported by the fact that vascular endothelial cells in the human express high levels of *MEG3* and thereby may be able to regulate angiogenesis related signalling cascades and processes. To investigate the two hypotheses, immunohistochemistry stainings could initially be performed for a number of proteins that are known to be involved in angiogenesis. These proteins should include VEGF, VEGFR1, members of the TGF- β 1 and PI3K/AKT pathways as well as TIE1 and other members of the Angiopoietin/TIE family. Other interesting targets might include EPH receptors (Erythroprotein-producing human hepatocellular carcinoma receptors) as well as their ligands Ephrins (Eph family receptor-interacting proteins) since they have also been shown to play an important role in vessel remodeling, organisation and maturation [164]. Since haemorrhaging is seen as early as e13.5, it would be interesting to analyse vessel formation and maturation at various stages during development to not only analyse whether angiogenesis pathways are misregulated in the ESC(Comp) chimaeras but also to detect the developmental timepoint at which the changes are first occurring.

8.2.2 Possible Effect of *Gtl2* Silencing on Vasculature via DLK1-Notch signalling

Gtl2 is known to negatively control the expression of *Dlk1* (Delta like non-canonical Notch ligand 1), one of the paternally expressed genes of the *Dlk1-Dio3* region, [165, 166] by targeting the polycomb repressive complex 2 (PRC2) to the *Dlk1* promoter. In a number of mouse models with aberrantly low expression of *Gtl2*, *Dlk1* expression is significantly in-

creased. DLK1 inhibits Notch signalling by mimicking the other Notch ligands. However, it lacks an important domain that is required for activation of the Notch signalling pathway and therefore DLK1 is inhibitory. Notch signalling is involved in a variety of processes related to angiogenesis and interestingly, knockouts of some Notch proteins and a number of its ligands cause vascular phenotypes including defects in vascular remodelling and morphogenesis and severe haemorrhaging. *NOTCH1/NOTCH4* double knockouts display severe defects in vascular morphogenesis including malformation of the large blood vessels. These embryos are not viable and die at e10 [167]. Deletion of the Notch ligand *JAGGED* results in abnormal vasculature and haemorrhaging. These embryos are also not viable beyond e10 [168]. Similarly, a double knockout of *HEY1* and *HEY2* (Hairy/enhancer-of-split related with YRPW motif protein 1/2) also results in severe haemorrhaging from e9.5 and defects in vascular patterning [169].

It is not very likely that disturbed signalling of the Notch signalling pathway via DLK1 is causing the phenotype in our ESC(Comp) chimaeras. Firstly, *Dlk1* gene expression levels in the ES cells were not significantly increased compared to ESC(Ctrl) cells and secondly, lethality in these Notch signalling mutants occurs much earlier than in our chimaeras. However, it is possible that although *Dlk1* expression was normal in the ES cells it was upregulated at a later developmental stage or that the chimaeric environment decreases the severity of the phenotype, leading to lethality at a later timepoint. Therefore this possibility should certainly be considered for future experiments.

8.2.3 Possible Role of a Placental Phenotype

Most of the Notch mutants also have a strong vascular placental phenotype that is largely responsible for the embryonic lethality [170]. This poses the question to what extent a placental phenotype could be responsible for the lethality observed in the ESC(Comp) chimaeras. Since ES cells only contribute to the embryonic tissues, they do not contribute to the placenta. However, they do contribute to the fetal blood vessels which form part of the developing placenta. It would be very interesting to examine this further, specifically focussing on the development of fetal placental blood vessels, as well as the labyrinth zone where the interchange between fetal and maternal circulations occurs.

8.2.4 Possible Role of a Musculoskeletal Phenotype

Previous studies on mice with mutations in the *Dlk1-Dio3* region that result in the loss of expression of the maternally expressed genes, show that those embryos have musculoskeletal defects. Georgiades et al. [120] show that embryos lacking *Gtl2* expression have defects

in skeletal muscle maturation and present with skeletal abnormalities. Similar results were also published by other groups [121, 122, 171]. These studies suggest, that the *Dlk1-Dio3* region may play more than one role during embryonic development. However, whether this is regulated through similar pathways remains unclear.

We did not analyse musculoskeletal phenotypes in the ESC(Comp) chimaeras in this present study, but it is possible that there are such muscular defects as well as other developmental defects or delays that have contributed to embryonic lethality. It is unlikely that a muscular defect alone is causing embryonic lethality, but understanding the full extent of the *Dlk1-Dio3* knockout/silencing phenotype could potentially shed light on the underlying cellular mechanisms that cause these developmental defects.

It would also be interesting to generate conditional knockouts of the maternally expressed genes of the *Dlk1-Dio3* region to analyse the role of them in different tissues or cell types, including vascular endothelial cells and muscle. This may enable us to better understand its precise role during development.

8.3 Concluding remarks

To conclude, we have identified the maternally expressed genes of the *Dlk1-Dio3* region - *Gtl2*, *Rtl1/Rtl1as*, *Rian* and *Mirg* - as markers that can distinguish between ES cells with normal or compromised developmental potency. This finding is in accordance with previous studies in the literature, but the precise roles and mechanisms of action of these genes during development and their exact contribution to the phenotype that we observed remains to be determined. Nevertheless, this finding is important especially for researchers working in the field of transgenics. It is common practise to screen targeted ES cells for the correct gene targeting event, karyotype and pathogens before performing blastocyst injections and investing time and money for the generation of genetically modified mice. We suggest to also include analysing gene expression of *Gtl2*, *Rtl1/Rtl1as*, *Rian* and *Mirg* to the pre-injection screening routine. Loss of imprinting is not an uncommon phenomenon in ES cell culture and cells with aberrant silencing of the maternally expressed genes of the *Dlk1-Dio3* imprinted region will most certainly not be able to contribute to viable chimaeras.

References

- [1] E. Haeckel, *Natürliche Schöpfungsgeschichte: Gemeinverständliche wissenschaftliche Vorträge über die Entwicklungslehre im Allgemeinen und diejenige von Darwin, Goethe und Lamarck im Besonderen*. Berlin: George Reimer, 1889.
- [2] V. Haecker, "Die Kerntheilungsvorgänge bei der Mesoderm- und Entodermbildung von Cyclops," *Archiv für Mikroskopische Anatomie*, vol. 39, pp. 556–581, 1892.
- [3] T. Boveri, "Ueber die Entstehung des Gegensatzes zwischen den Geschlechtszellen und den somatischen Zellen bei *Ascaris megalcephala*, nebst Bemerkungen zur Entwicklungsgeschichte der Nematoden," *Ges Morphol Physiol*, vol. 8, pp. 114–125, 1892.
- [4] L. C. Stevens and C. C. Little, "Spontaneous Testicular Teratomas in an Inbred Strain of Mice," *Proceedings of the National Academy of Sciences*, vol. 40, no. 11, pp. 1080–1087, 1954.
- [5] L. C. Stevens, "Experimental production of testicular teratomas in mice," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 52, no. 3, pp. 54–59, 1964.
- [6] L. J. Kleinsmith and G. B. Pierce, "Multipotentiality of Single Embryonal Carcinoma Cells," *Cancer Research*, vol. 24, pp. 1544–1551, 1964.
- [7] L. C. Stevens, "The Development of Transplantable Teratocarcinomas from Intratesticular Grafts of Pre- and Postimplantation Mouse Embryos," *Developmental Biology*, vol. 21, no. 3, pp. 364–382, 1970.
- [8] D. Solter, N. Škreb, and I. Damjanov, "Extrauterine Growth of Mouse Egg-cylinders results in Malignant Teratoma," *Nature*, vol. 227, no. 5257, pp. 503–504, 1970.
- [9] M. J. Evans, "The isolation and properties of a clonal tissue culture strain of pluripotent mouse teratoma cells," *Journal of Embryology and Experimental Morphology*, vol. 28, no. 1, pp. 163–176, 1972.
- [10] R. L. Brinster, "The effect of cells transferred into the mouse blastocyst on subsequent development," *Journal of Experimental Medicine*, vol. 140, no. 4, pp. 1049–1056, 1974.
- [11] V. E. Papaioannou, M. W. Mcburney, R. L. Gardner, and M. J. Evans, "Fate of teratocarcinoma cells injected into early mouse embryos," *Nature*, vol. 258, no. 5530, pp. 70–73, 1975.

- [12] B. Mintz and K. Illmensee, "Normal genetically mosaic mice produced from malignant teratocarcinoma cells," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 72, no. 9, pp. 3585–3589, 1975.
- [13] M. J. Evans and M. H. Kaufman, "Establishment in culture of pluripotential cells from mouse embryos," *Nature*, vol. 292, no. 5819, pp. 154–156, 1981.
- [14] G. R. Martin, "Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells," *Proceedings of the National Academy of Sciences*, vol. 78, no. 12, pp. 7634–7638, 1981.
- [15] A. Bradley, M. Evans, M. H. Kaufman, and E. Robertson, "Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines," *Nature*, vol. 309, no. 5965, pp. 255–256, 1984.
- [16] A. K. Tarkowski and J. Wróblewska, "Development of blastomeres of mouse eggs isolated at the 4- and 8-cell stage," *Journal of Embryology and Experimental Morphology*, vol. 18, no. 1, pp. 155–180, 1967.
- [17] D. Strumpf, C.-A. Mao, Y. Yamanaka, A. Ralston, K. Chawengsaksophak, F. Beck, and J. Rossant, "Cdx2 is required for correct cell fate specification and differentiation of trophoctoderm in the mouse blastocyst," *Development*, vol. 132, no. 9, pp. 2093–2102, 2005.
- [18] S. L. Palmieri, W. Peter, H. Hess, and H. R. Schöler, "Oct-4 Transcription Factor is Differentially Expressed in the Mouse Embryo during Establishment of the First Two Extraembryonic Cell Lineages Involved in Implantation," *Developmental Biology*, vol. 166, no. 1, pp. 259–267, 1994.
- [19] J. Nichols, B. Zevnik, K. Anastasiadis, H. Niwa, D. Klewe-Nebenius, I. Chambers, H. Scholer, and A. Smith, "Formation of Pluripotent Stem Cells in the Mammalian Embryo Depends on the POU Transcription Factor Oct4," *Cell*, vol. 95, no. 3, pp. 379–391, 1998.
- [20] I. Chambers, D. Colby, M. Robertson, J. Nichols, S. Lee, S. Tweedie, and A. Smith, "Functional Expression Cloning of Nanog, a Pluripotency Sustaining Factor in Embryonic Stem Cells," *Cell*, vol. 113, no. 5, pp. 643–655, 2003.
- [21] R. L. Gardner and J. Rossant, "Investigation of the fate of 4.5 day post-coitum mouse inner cell mass cells by blastocyst injection," *Journal of Embryology and Experimental Morphology*, vol. 52, pp. 141–152, 1979.
- [22] R. L. Gardner, "Investigation of cell lineage and differentiation in the extraembryonic endoderm of the mouse embryo," *Journal of Embryology and Experimental Morphology*, vol. 68, pp. 175–198, 1982.
- [23] B. Plusa, A. Piliszek, S. Frankenberg, J. Artus, and A.-K. Hadjantonakis, "Distinct sequential cell behaviours direct primitive endoderm formation in the mouse blastocyst," *Development*, vol. 135, no. 18, pp. 3081–3091, 2008.

- [24] C. Chazaud, Y. Yamanaka, T. Pawson, and J. Rossant, "Early Lineage Segregation between Epiblast and Primitive Endoderm in Mouse Blastocysts through the Grb2-MAPK Pathway," *Developmental Cell*, vol. 10, no. 5, pp. 615–624, 2006.
- [25] H. Niwa, J.-i. Miyazaki, and A. G. Smith, "Quantitative expression of Oct3/4 defines differentiation, dedifferentiation or self-renewal of ES cells," *Nature Genetics*, vol. 24, no. 4, pp. 372–376, 2000.
- [26] M. Keramari, J. Razavi, K. A. Ingman, C. Patsch, F. Edenhofer, M. Christopher, and S. J. Kimber, "Sox2 Is Essential for Formation of Trophectoderm in the Preimplantation Embryo," *PLoS One*, vol. 5, no. 11, p. e13952, 2010.
- [27] A. A. Avilion, S. K. Nicolis, L. H. Pevny, L. Perez, N. Vivian, and R. Lovell-badge, "Multipotent cell lineages in early mouse development depend on SOX2 function," *Genes and Development*, vol. 17, no. 1, pp. 126–140, 2003.
- [28] S. Masui, Y. Nakatake, Y. Toyooka, D. Shimosato, R. Yagi, K. Takahashi, H. Okochi, A. Okuda, R. Matoba, A. A. Sharov, M. S. H. Ko, and H. Niwa, "Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells," *Nature Cell Biology*, vol. 9, no. 6, pp. 625–635, 2007.
- [29] H. Yuan, N. Corbi, C. Basilico, and L. Dailey, "Developmental-specific activity of the FGF-4 enhancer requires the synergistic action of Sox2 and Oct-3," *Genes and Development*, vol. 9, no. 21, pp. 2635–2645, 1995.
- [30] D.-C. Ambrosetti, C. Basilico, and L. Dailey, "Synergistic activation of the fibroblast growth factor 4 enhancer by Sox2 and Oct-3 depends on protein-protein interactions facilitated by a specific spatial arrangement of factor binding sites," *Molecular and Cellular Biology*, vol. 17, no. 11, pp. 6321–6329, 1997.
- [31] K. Mitsui, Y. Tokuzawa, H. Itoh, K. Segawa, M. Murakami, K. Takahashi, M. Maruyama, M. Maeda, and S. Yamanaka, "The Homeoprotein Nanog Is Required for Maintenance of Pluripotency in Mouse Epiblast and ES Cells," *Cell*, vol. 113, no. 5, pp. 631–642, 2003.
- [32] J. Silva, J. Nichols, T. W. Theunissen, G. Guo, A. L. Van Oosten, O. Barrandon, J. Wray, S. Yamanaka, I. Chambers, and A. Smith, "Nanog Is the Gateway to the Pluripotent Ground State," *Cell*, vol. 138, no. 4, pp. 722–737, 2009.
- [33] A. G. Smith, J. K. Heath, D. D. Donaldson, G. G. Wong, J. Moreau, M. Stahl, and D. Rogers, "Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides," *Nature*, vol. 336, no. 6200, pp. 688–690, 1988.
- [34] R. L. Williams, D. J. Hilton, S. Pease, T. A. Willson, C. L. Stewart, D. P. Gearing, E. F. Wagner, D. Metcalf, N. A. Nicola, and N. M. Gough, "Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells," *Nature*, vol. 336, no. 6200, pp. 684–687, 1988.
- [35] Q.-L. Ying, J. Nichols, I. Chambers, and A. Smith, "BMP Induction of Id Proteins Suppresses Differentiation and Sustains Embryonic Stem Cell Self-Renewal in Collaboration with STAT3," *Cell*, vol. 115, no. 3, pp. 281–292, 2003.

- [36] T. Kunath, M. K. Saba-El-Leil, M. Almousailleakh, J. Wray, S. Meloche, and A. Smith, "FGF stimulation of the Erk1/2 signalling cascade triggers transition of pluripotent embryonic stem cells from self-renewal to lineage commitment," *Development*, vol. 134, no. 16, pp. 2895–2902, 2007.
- [37] Q.-L. Ying, J. Wray, J. Nichols, L. Batlle-Morera, B. Doble, J. Woodgett, P. Cohen, and A. Smith, "The ground state of embryonic stem cell self-renewal," *Nature*, vol. 453, no. 7194, pp. 519–523, 2008.
- [38] R. Anton, H. A. Kestler, and M. Kühl, " β -Catenin signaling contributes to stemness and regulates early differentiation in murine embryonic stem cells," *FEBS Letters*, vol. 581, no. 27, pp. 5247–5254, 2007.
- [39] L. Haegeler, B. Ingold, H. Naumann, G. Tabatabai, B. Ledermann, and S. Brandner, "Wnt signalling inhibits neural differentiation of embryonic stem cells by controlling bone morphogenetic protein expression," *Molecular and Cellular Neuroscience*, vol. 24, no. 3, pp. 696–708, 2003.
- [40] M. F. Kielman, M. Rindapää, C. Gaspar, N. Van Poppel, C. Breukell, S. Van Leeuwen, M. M. Taketo, S. Roberts, R. Smits, and R. Fodde, "Apc modulates embryonic stem-cell differentiation by controlling the dosage of β -catenin signaling," *Nature Genetics*, vol. 32, no. 4, pp. 594–605, 2002.
- [41] C.-I. Wu, J. A. Hoffman, B. R. Shy, E. M. Ford, E. Fuchs, H. Nguyen, and B. J. Merrill, "Function of Wnt/ β -catenin in counteracting Tcf3 repression through the Tcf3- β -catenin interaction," *Development*, vol. 139, no. 12, pp. 2118–2129, 2012.
- [42] M. Gardiner-Garden and M. Frommer, "CpG Islands in Vertebrate Genomes," *Journal of Molecular Biology*, vol. 196, no. 2, pp. 261–282, 1987.
- [43] F. Fuks, W. A. Burgers, A. Brehm, L. Hughes-Davies, and T. Kouzarides, "DNA methyltransferase Dnmt1 associates with histone deacetylase activity," *Nature Genetics*, vol. 24, no. 1, pp. 88–91, 2000.
- [44] W. Reik, W. Dean, and J. Walter, "Epigenetic Reprogramming in Mammalian Development," *Science*, vol. 293, no. 5532, pp. 1089–1093, 2001.
- [45] A. Tsumura, T. Hayakawa, Y. Kumaki, S.-I. Takebayashi, M. Sakaue, C. Matsuoka, K. Shimotohno, F. Ishikawa, E. Li, H. R. Ueda, J.-I. Nakayama, and M. Okano, "Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b," *Genes to Cells*, vol. 11, no. 7, pp. 805–814, 2006.
- [46] M. Jackson, A. Krassowska, N. Gilbert, L. Forrester, J. Ansell, T. Chevassut, and B. Ramsahoye, "Severe Global DNA Hypomethylation Blocks Differentiation and Induces Histone Hyperacetylation in Embryonic Stem Cells," *Molecular and Cellular Biology*, vol. 24, no. 20, pp. 8862–8871, 2004.
- [47] B. E. Bernstein, T. S. Mikkelsen, X. Xie, M. Kamal, D. J. Huebert, J. Cuff, B. Fry, A. Meissner, M. Wernig, K. Plath, R. Jaenisch, A. Wagschal, R. Feil, S. L. Schreiber, and E. S. Lander, "A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells," *Cell*, vol. 125, no. 2, pp. 315–326, 2006.

- [48] T. Nakamura, Y. Arai, H. Umehara, M. Masuhara, T. Kimura, H. Taniguchi, T. Sekimoto, M. Ikawa, Y. Yoneda, M. Okabe, S. Tanaka, K. Shiota, and T. Nakano, "PGC7/Stella protects against DNA demethylation in early embryogenesis," *Nature Cell Biology*, vol. 9, no. 1, pp. 64–71, 2007.
- [49] X. Li, M. Ito, F. Zhou, N. Youngson, X. Zuo, P. Leder, and A. C. Ferguson-Smith, "A Maternal-Zygotic Effect Gene, *Zfp57*, Maintains Both Maternal and Paternal Imprints," *Developmental Cell*, vol. 15, no. 4, pp. 547–557, 2008.
- [50] J. McGrath and D. Solter, "Completion of Mouse Embryogenesis Requires Both the Maternal and Paternal Genomes," *Cell*, vol. 37, no. 1, pp. 179–183, 1984.
- [51] S. C. Barton, M. A. H. Surani, and M. L. Norris, "Role of paternal and maternal genomes in mouse development," *Nature*, vol. 311, no. 5984, pp. 374–376, 1984.
- [52] T. M. DeChiara, E. J. Robertson, and A. Efstratiadis, "Parental Imprinting of the Mouse Insulin-like Growth Factor II Gene," *Cell*, vol. 64, no. 4, pp. 849–859, 1991.
- [53] A. C. Ferguson-Smith, B. M. Cattanach, S. C. Barton, C. V. Beechey, and M. A. Surani, "Embryological and molecular investigations of parental imprinting on mouse chromosome 7," *Nature*, vol. 351, no. 6328, pp. 667–670, 1991.
- [54] A. L. Fowden, C. Sibley, W. Reik, and M. Constancia, "The Biology of Fetal Growth Imprinted Genes, Placental Development and Fetal Growth," *Hormone Research*, vol. 65, no. suppl 3, pp. 50–58, 2006.
- [55] A. M. Wobus, K. Guan, H. T. Yang, and K. R. Boheler, "Embryonic Stem Vells as a Model to Study Cardiac, Skeletal Muscle, and Vascular Smooth Muscle Cell Differentiation," *Methods in Molecular Biology*, vol. 185, no. 3, pp. 127–156, 2002.
- [56] S. J. Kattman, A. D. Witty, M. Gagliardi, N. C. Dubois, M. Niapour, A. Hotta, J. Ellis, and G. Keller, "Stage-Specific Optimization of Activin/Nodal and BMP Signaling Promotes Cardiac Differentiation of Mouse and Human Pluripotent Stem Cell Lines," *Cell Stem Cell*, vol. 8, no. 2, pp. 228–240, 2011.
- [57] I.-N. E. Wang, X. Wang, X. Ge, J. Anderson, M. Ho, E. Ashley, J. Liu, M. J. Butte, M. Yazawa, R. E. Dolmetsch, T. Quertermous, and P. C. Yang, "Apelin Enhances Directed Cardiac Differentiation of Mouse and Human Embryonic Stem Cells," *PLoS One*, vol. 7, no. 6, p. e38328, 2012.
- [58] A. Rolletschek, H. Chang, K. Guan, J. Czyz, M. Meyer, and A. M. Wobus, "Differentiation of embryonic stem cell-derived dopaminergic neurons is enhanced by survival-promoting factors," *Mechanisms of Development*, vol. 105, pp. 93–104, 2001.
- [59] G. B. Miles, D. C. Yohn, H. Wichterle, T. M. Jessell, V. F. Rafuse, and R. M. Brownstone, "Functional Properties of Motoneurons Derived from Mouse Embryonic Stem Cells," *Journal of Neuroscience*, vol. 24, no. 36, pp. 7848–7858, 2004.
- [60] K. Moriyasu, H. Yamazoe, and H. Iwata, "Induction dopamine releasing cells from mouse embryonic stem cells and their long-term culture," *Journal of Biomedical Materials Research - Part A*, vol. 77, no. 1, pp. 136–147, 2006.

- [61] J. Sui, M. Mehta, B. Shi, G. Morahan, and F.-X. Jiang, "Directed Differentiation of Embryonic Stem Cells Allows Exploration of Novel Transcription Factor Genes for Pancreas Development," *Stem Cell Reviews and Reports*, vol. 8, no. 3, pp. 803–812, 2012.
- [62] P. Blyszczuk, C. Asbrand, A. Rozzo, G. Kania, L. St-Onge, M. Rupnik, and A. M. Wobus, "Embryonic stem cells differentiate into insulin-producing cells without selection of nestin-expressing cells," *International Journal of Developmental Biology*, vol. 48, no. 10, pp. 1095–1104, 2004.
- [63] M. Eiraku and Y. Sasai, "Mouse embryonic stem cell culture for generation of three-dimensional retinal and cortical tissues," *Nature Protocols*, vol. 7, no. 1, pp. 69–79, 2012.
- [64] B. Wdziekonski, P. Villageois, and C. Dani, "Differentiation of Mouse Embryonic Stem Cells and of Human Adult Stem Cells into Adipocytes.," *Current Protocols in Cell Biology*, vol. 34, no. 1, pp. 23.4.1–23.4.14, 2007.
- [65] J. Chal, M. Oginuma, Z. Al Tanoury, B. Gobert, O. Sumara, A. Hick, F. Bousson, Y. Zidouni, C. Mursch, P. Moncuquet, O. Tassy, S. Vincent, A. Miyanari, A. Bera, J. M. Garnier, G. Guevara, M. Hestin, L. Kennedy, S. Hayashi, B. Drayton, T. Cherrier, B. Gayraud-Morel, E. Gussoni, F. Relaix, S. Tajbakhsh, and O. Pourquie, "Differentiation of pluripotent stem cells to muscle fiber to model Duchenne muscular dystrophy," *Nature Biotechnology*, vol. 33, no. 9, pp. 962–969, 2015.
- [66] S. J. Bruce, R. W. Rea, A. L. Steptoe, M. Busslinger, J. F. Bertram, and A. C. Perkins, "In vitro differentiation of murine embryonic stem cells toward a renal lineage," *Differentiation*, vol. 75, no. 5, pp. 337–349, 2007.
- [67] K. Hayashi, H. Ohta, K. Kurimoto, S. Aramaki, and M. Saitou, "Reconstitution of the Mouse Germ Cell Specification Pathway in Culture by Pluripotent Stem Cells," *Cell*, vol. 146, no. 4, pp. 519–532, 2011.
- [68] J. Kim and S. H. Orkin, "Embryonic stem cell-specific signatures in cancer: insights into genomic regulatory networks and implications for medicine," *Genome Medicine*, vol. 3, no. 11, p. 75, 2011.
- [69] H. Wang, M. La Russa, and L. S. Qi, "CRISPR/Cas9 in Genome Editing and Beyond," *Annual Review of Biochemistry*, vol. 85, no. 1, pp. 227–264, 2016.
- [70] W. Wang, C. Lin, D. Lu, Z. Ning, T. Cox, D. Melvin, X. Wang, A. Bradley, and P. Liu, "Chromosomal transposition of PiggyBac in mouse embryonic stem cells," *Proceedings of the National Academy of Sciences*, vol. 105, no. 27, pp. 9290–9295, 2008.
- [71] S. Labialle, L. Yang, X. Ruan, A. Villemain, J. V. Schmidt, A. Hernandez, T. Wiltshire, N. Cermakian, and A. K. Naumova, "Coordinated diurnal regulation of genes from the Dlk1-Dio3 imprinted domain: implications for regulation of clusters of non-paralogous genes," *Human Molecular Genetics*, vol. 17, no. 1, pp. 15–26, 2008.
- [72] K. D. Tartof and C. A. Hobbs, "Improved media for growing plasmid and cosmid clones," *Bethesda Research Laboratories Focus*, vol. 9, p. 12, 1987.

- [73] D. Kim, G. Pertea, C. Trapnell, H. Pimentel, R. Kelley, and S. L. Salzberg, “TopHat2 : accurate alignment of transcriptomes in the presence of insertions , deletions and gene fusions,” *Genome Biology*, vol. 14, no. 4, p. R36, 2013.
- [74] S. Anders, P. T. Pyl, and W. Huber, “HTSeq - A Python framework to work with high-throughput sequencing data,” *Bioinformatics*, vol. 31, no. 2, pp. 166–169, 2015.
- [75] M. I. Love, W. Huber, and S. Anders, “Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2,” *Genome Biology*, vol. 15, no. 12, p. 550, 2014.
- [76] X. Xiao, J. A. Douthwaite, Y. Chen, B. Kemp, S. Kidd, J. Percival-Alwyn, A. Smith, K. Goode, B. Swerdlow, D. Lowe, H. Wu, W. F. Dall’Acqua, and P. S. Chowdhury, “A high-throughput platform for population reformatting and mammalian expression of phage display libraries to enable functional screening as full-length IgG,” *mAbs*, vol. 9, no. 6, pp. 996–1006, 2017.
- [77] N. Gaztelumendi and C. Nogués, “Chromosome Instability in mouse Embryonic Stem Cells,” *Scientific Reports*, vol. 4, p. 5324, 2014.
- [78] P. Rebuzzini, T. Neri, M. Zuccotti, C. A. Redi, and S. Garagna, “Chromosome number variation in three mouse embryonic stem cell lines during culture,” *Cytotechnology*, vol. 58, no. 1, pp. 17–23, 2008.
- [79] A. Sugawara, K. Goto, Y. Sotomaru, T. Sofuni, and T. Ito, “Current Status of Chromosomal Abnormalities in Mouse Embryonic Stem Cell Lines Used in Japan,” *Comparative Medicine*, vol. 56, no. 1, pp. 31–34, 2006.
- [80] S.-H. Moon, J.-S. Kim, S.-J. Park, J.-J. Lim, H.-J. Lee, S. M. Lee, and H.-M. Chung, “Effect of chromosome instability on the maintenance and differentiation of human embryonic stem cells in vitro and in vivo,” *Stem Cell Research*, vol. 6, no. 1, pp. 50–59, 2011.
- [81] D. A. Lightfoot, A. Kouznetsova, E. Mahdy, J. Wilbertz, and C. Höög, “The fate of mosaic aneuploid embryos during mouse development,” *Developmental Biology*, vol. 289, no. 2, pp. 384–394, 2006.
- [82] P. Liu, H. Zhang, A. McLellan, H. Vogel, and A. Bradley, “Embryonic Lethality and Tumorigenesis Caused by Segmental Aneuploidy on Mouse Chromosome 11,” *Genetics*, vol. 150, no. 3, pp. 1155–1168, 1998.
- [83] M. Sugiura-Ogasawara, Y. Ozaki, K. Katano, N. Suzumori, T. Kitaori, and E. Mizutani, “Abnormal embryonic karyotype is the most frequent cause of recurrent miscarriage,” *Human Reproduction*, vol. 27, no. 8, pp. 2297–2303, 2012.
- [84] A. Nagy, J. Rossant, R. Nagy, W. Abramow-Newerly, and J. C. Roder, “Derivation of completely cell culture-derived mice from early-passage embryonic stem cells,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 90, no. 18, pp. 8424–8428, 1993.

- [85] X.-Y. Li, Q. Jia, K.-Q. Di, S.-M. Gao, X.-H. Wen, R.-Y. Zhou, W. Wei, and L.-Z. Wang, "Passage number affects the pluripotency of mouse embryonic stem cells as judged by tetraploid embryo aggregation," *Cell and Tissue Research*, vol. 327, no. 3, pp. 607–614, 2007.
- [86] W. Dean, L. Bowden, A. Aitchison, J. Klose, T. Moore, J. J. Meneses, W. Reik, and R. Feil, "Altered imprinted gene methylation and expression in completely ES cell-derived mouse fetuses: association with aberrant phenotypes," *Development*, vol. 125, no. 12, pp. 2273–2282, 1998.
- [87] T. T. Puck, "Studies of the Life Cycle of Mammalian Cells," *Cold Spring Harbor Symposia on Quantitative Biology*, vol. 29, pp. 167–176, 1964.
- [88] D. J. Rodda, J.-L. Chew, L.-H. Lim, Y.-H. Loh, B. Wang, H.-H. Ng, and P. Robson, "Transcriptional Regulation of Nanog by OCT4 and SOX2," *Journal of Biological Chemistry*, vol. 280, no. 26, pp. 24731–24737, 2005.
- [89] K. Takahashi and S. Yamanaka, "Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors," *Cell*, vol. 126, no. 4, pp. 663–676, 2006.
- [90] P. Zhang, R. Andrianakos, Y. Yang, C. Liu, and W. Lu, "Kruppel-like Factor 4 (Klf4) Prevents Embryonic Stem (ES) Cell Differentiation by Regulating Nanog Gene Expression," *Journal of Biological Chemistry*, vol. 285, no. 12, pp. 9180–9189, 2010.
- [91] N. Festuccia, R. Osorno, F. Halbritter, V. Karwacki-Neisius, P. Navarro, D. Colby, F. Wong, A. Yates, S. R. Tomlinson, and I. Chambers, "Esrrb Is a Direct Nanog Target Gene that Can Substitute for Nanog Function in Pluripotent Cells," *Cell Stem Cell*, vol. 11, no. 4, pp. 477–490, 2012.
- [92] H. Niwa, T. Burdon, I. Chambers, and A. Smith, "Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3," *Genes and Development*, vol. 12, no. 13, pp. 2048–2060, 1998.
- [93] D. L. Ruzicka and R. J. Schwartz, "Sequential activation of alpha-actin genes during avian cardiogenesis: vascular smooth muscle alpha-actin gene transcripts mark the onset of cardiomyocyte differentiation," *Journal of Cell Biology*, vol. 107, no. 6, pp. 2575–2586, 1988.
- [94] Y. M. Cho, J. M. Lim, D. H. Yoo, J. H. Kim, S. S. Chung, S. G. Park, T. H. Kim, S. K. Oh, Y. M. Choi, S. Y. Moon, K. S. Park, and H. K. Lee, "Betacellulin and nicotinamide sustain PDX1 expression and induce pancreatic β -cell differentiation in human embryonic stem cells," *Biochemical and Biophysical Research Communications*, vol. 366, no. 1, pp. 129–134, 2008.
- [95] M. Kanai-Azuma, Y. Kanai, J. M. Gad, Y. Tajima, C. Taya, M. Kurohmaru, Y. Sanai, H. Yonekawa, K. Yazaki, P. P. L. Tam, and Y. Hayashi, "Depletion of definitive gut endoderm in Sox17-null mutant mice," *Development*, vol. 129, no. 10, pp. 2367–79, 2002.

- [96] T. T. Schwindt, F. L. Motta, G. F. Barnabé, C. G. Massant, A. O. Guimarães, M. E. Calcagnotto, J. B. Pesquero, and L. E. Mello, "Effects of FGF-2 and EGF removal on the differentiation of mouse neural precursor cells," *Anais da Academia Brasileira de Ciências*, vol. 81, no. 3, pp. 443–452, 2009.
- [97] S. M. Pollard, R. Wallbank, S. Tomlinson, L. Grotewold, and A. Smith, "Fibroblast growth factor induces a neural stem cell phenotype in foetal forebrain progenitors and during embryonic stem cell differentiation," *Molecular and Cellular Neuroscience*, vol. 38, no. 3, pp. 393–403, 2008.
- [98] P. Stankiewicz and J. R. Lupski, "Structural Variation in the Human Genome and its Role in Disease," *Annual Review of Medicine*, vol. 61, no. 1, pp. 437–455, 2010.
- [99] K. Klinger, G. Landes, D. Shook, R. Harvey, L. Lopez, P. Locke, T. Lerner, R. Osathanondh, B. Leverone, T. Houseal, K. Pavelka, and W. Dackowski, "Rapid Detection of Chromosome Aneuploidies in Uncultured Amniocytes by Using Fluorescence In Situ Hybridization (FISH)," *Am. J. Hum. Genet.*, vol. 51, pp. 55–65, 1992.
- [100] X.-Y. Lu, M. T. Ohung, C. A. Shaw, K. Pham, S. E. Neil, A. Patel, T. Sahoo, C. A. Bacino, P. Stankiewicz, S. Lee Kang, S. Lalani, C. Chinault, J. R. Lupski, S. W. Cheung, and L. Beaudet, Arthur, "Genomic Imbalances in Neonates With Birth Defects: High Detection Rates by Using Chromosomal Microarray Analysis," *Pediatrics*, vol. 122, no. 6, pp. 1310–1318, 2008.
- [101] G. Guo, L. Pinello, X. Han, S. Lai, L. Shen, T. W. Lin, K. Zou, G. C. Yuan, and S. H. Orkin, "Serum-Based Culture Conditions Provoke Gene Expression Variability in Mouse Embryonic Stem Cells as Revealed by Single-Cell Analysis," *Cell Reports*, vol. 14, no. 4, pp. 956–965, 2016.
- [102] J. Wray, T. Kalkan, S. Gomez-Lopez, D. Eckardt, A. Cook, R. Kemler, and A. Smith, "Inhibition of glycogen synthase kinase-3 alleviates Tcf3 repression of the pluripotency network and increases embryonic stem cell resistance to differentiation," *Nature Cell Biology*, vol. 13, no. 7, pp. 838–845, 2011.
- [103] I. Chambers, J. Silva, D. Colby, J. Nichols, B. Nijmeijer, M. Robertson, J. Vrana, K. Jones, L. Grotewold, and A. Smith, "Nanog safeguards pluripotency and mediates germline development," *Nature*, vol. 450, no. 7173, pp. 1230–1234, 2007.
- [104] A. M. Singh, T. Hamazaki, K. E. Hankowski, and N. Terada, "A Heterogeneous Expression Pattern for Nanog in Embryonic Stem Cells," *Stem Cells*, vol. 25, no. 10, pp. 2534–2542, 2007.
- [105] Y. Toyooka, D. Shimosato, K. Murakami, K. Takahashi, and H. Niwa, "Identification and characterization of subpopulations in undifferentiated ES cell culture," *Development*, vol. 135, no. 5, pp. 909–918, 2008.
- [106] H. Niwa, K. Ogawa, D. Shimosato, and K. Adachi, "A parallel circuit of LIF signalling pathways maintains pluripotency of mouse ES cells," *Nature*, vol. 460, no. 7251, pp. 118–122, 2009.

- [107] M. A. Canham, A. A. Sharov, M. S. H. Ko, and J. M. Brickman, "Functional Heterogeneity of Embryonic Stem Cells Revealed through Translational Amplification of an Early Endodermal Transcript," *PLoS Biology*, vol. 8, no. 5, p. e1000379, 2010.
- [108] D. L. C. van den Berg, W. Zhang, A. Yates, E. Engelen, K. Takacs, K. Bezstarosti, J. Demmers, I. Chambers, and R. A. Poot, "Estrogen-Related Receptor Beta Interacts with Oct4 To Positively Regulate Nanog Gene Expression," *Molecular and Cellular Biology*, vol. 28, no. 19, pp. 5986–5995, 2008.
- [109] B. McClintock, "The origin and behavior of mutable loci in maize," *Proceedings of the National Academy of Sciences*, vol. 36, no. 6, pp. 344–355, 1950.
- [110] U. Lakshmipathy, B. Pelacho, K. Sudo, J. L. Linehan, E. Coucouvanis, D. S. Kaufman, and C. M. Verfaillie, "Efficient Transfection of Embryonic and Adult Stem Cells," *Stem Cells*, vol. 22, no. 4, pp. 531–543, 2004.
- [111] Y. Yang and G. Oliver, "Review series mammalian lymphatic vasculature," *The Journal of Clinical Investigation*, vol. 124, no. 3, pp. 888–897, 2014.
- [112] H. Sumida, K. Noguchi, Y. Kihara, M. Abe, K. Yanagida, F. Hamano, S. Sato, K. Tamaki, Y. Morishita, M. R. Kano, C. Iwata, K. Miyazono, K. Sakimura, T. Shimizu, and S. Ishii, "LPA4 regulates blood and lymphatic vessel formation during mouse embryogenesis," *Blood*, vol. 116, no. 23, pp. 5060–5070, 2010.
- [113] E. Torsney, R. Charlton, A. G. Diamond, J. Burn, J. V. Soames, and H. M. Arthur, "Mouse Model for Hereditary Hemorrhagic Telangiectasia Has a Generalized Vascular Abnormality," *Circulation*, vol. 107, no. 12, pp. 1653–1657, 2003.
- [114] T. N. Sato, Y. Tozawa, U. Deutsch, K. Wolburg-Buchholz, Y. Fujiwara, M. Gendron-Maguire, T. Gridley, H. Wolburg, W. Risau, and Y. Qin, "Distinct roles of the receptor tyrosine kinases Tie-1 and Tie-2 in blood vessel formation," *Nature*, vol. 376, no. 6535, pp. 70–74, 1995.
- [115] D. V. Parums, J. L. Cordell, K. Micklem, A. R. Heryet, K. C. Gatter, and D. Y. Mason, "JC70: a new monoclonal antibody That detects vascular endothelium associated antigen on routinely processed tissue sections," *Journal of Clinical Pathology*, vol. 43, no. 9, pp. 752–757, 1990.
- [116] J. E. Park and A. Barbul, "Understanding the role of immune regulation in wound healing," *American Journal of Surgery*, vol. 187, no. 5A, pp. 11S–16S, 2004.
- [117] J. Palis, J. Malik, K. E. McGrath, and P. D. Kingsley, "Primitive erythropoiesis in the mammalian embryo," *International Journal of Developmental Biology*, vol. 54, pp. 1011–1018, 2010.
- [118] K. D. Hansen, S. E. Brenner, and S. Dudoit, "Biases in Illumina transcriptome sequencing caused by random hexamer priming," *Nucleic Acids Research*, vol. 38, no. 12, p. e131, 2010.

- [119] S. T. da Rocha, M. Tevendale, E. Knowles, S. Takada, M. Watkins, and A. C. Ferguson-Smith, "Restricted co-expression of Dlk1 and the reciprocally imprinted non-coding RNA, Gtl2: Implications for cis-acting control," *Developmental Biology*, vol. 306, no. 2, pp. 810–823, 2007.
- [120] P. Georgiades, M. Watkins, M. A. Surani, and A. C. Ferguson-Smith, "Parental origin-specific developmental defects in mice with uniparental disomy for chromosome 12," *Development*, vol. 127, no. 21, pp. 4719–4728, 2000.
- [121] M. Tevendale, M. Watkins, C. Rasberry, B. Cattanach, and A. C. Ferguson-Smith, "Analysis of mouse conceptuses with uniparental duplication/deficiency for distal chromosome 12: Comparison with chromosome 12 uniparental disomy and implications for genomic imprinting," *Cytogenetic and Genome Research*, vol. 113, no. 1-4, pp. 215–222, 2006.
- [122] S.-P. Lin, P. Coan, S. T. da Rocha, H. Seitz, J. Cavaille, P.-W. Teng, S. Takada, and A. C. Ferguson-Smith, "Differential regulation of imprinting in the murine embryo and placenta by the Dlk1-Dio3 imprinting control region," *Development*, vol. 134, no. 2, pp. 417–426, 2007.
- [123] M. Stadtfeld, E. Apostolou, H. Akutsu, A. Fukuda, P. Follett, S. Natesan, T. Kono, T. Shioda, and K. Hochedlinger, "Aberrant silencing of imprinted genes on chromosome 12qF1 in mouse induced pluripotent stem cells," *Nature*, vol. 465, no. 7295, pp. 175–181, 2010.
- [124] Y. Stelzer, C. S. Shivalila, F. Soldner, S. Markoulaki, and R. Jaenisch, "Tracing Dynamic Changes of DNA Methylation at Single-Cell Resolution," *Cell*, vol. 163, no. 1, pp. 218–229, 2015.
- [125] Y. Stelzer, H. Wu, Y. Song, C. S. Shivalila, S. Markoulaki, and R. Jaenisch, "Parent-of-Origin DNA Methylation Dynamics during Mouse Development," *Cell Reports*, vol. 16, no. 12, pp. 3167–3180, 2016.
- [126] H. G. Leitch, K. R. Mcewen, A. Turp, V. Encheva, T. Carroll, N. Grabole, W. Mansfield, B. Nashun, J. G. Knezovich, A. Smith, M. A. Surani, and P. Hajkova, "Naive pluripotency is associated with global DNA hypomethylation," *Nature Structural and Molecular Biology*, vol. 20, no. 3, pp. 311–316, 2013.
- [127] J. Yang, D. J. Ryan, W. Wang, J. C.-H. Tsang, G. Lan, H. Masaki, X. Gao, L. Antunes, Y. Yu, Z. Zhu, J. Wang, A. A. Kolodziejczyk, L. S. Campos, C. Wnag, F. Yang, Z. Zhong, B. Fu, M. A. Eckersley-Maslin, M. Wood, Y. Tanaka, X. Chen, A. C. Wilkinson, J. Bussell, J. White, R. Ramirez-Solis, W. Reik, B. Göttgens, S. A. Teichmann, P. P. L. Tam, N. Hiromitsu, X. Zou, L. Lu, and P. Liu, "Establishment of mouse expanded potential stem cells," *Nature*, vol. 550, no. 7676, pp. 393–397, 2017.
- [128] D. Humpherys, K. Eggan, H. Akutsu, K. Hochedlinger, W. M. Rideout, D. Biniszkiewicz, R. Yanagimachi, and R. Jaenisch, "Epigenetic instability in ES cells and cloned mice," *Science*, vol. 293, no. 5527, pp. 95–97, 2001.
- [129] J. Wray, T. Kalkan, and A. G. Smith, "The ground state of pluripotency," *Biochemical Society Transactions*, vol. 38, no. 4, pp. 1027–1032, 2010.

- [130] J. Choi, A. J. Huebner, K. Clement, R. M. Walsh, A. Savol, K. Lin, H. Gu, B. Di Stefano, J. Brumbaugh, S.-Y. Kim, J. Sharif, C. M. Rose, A. Mohammad, J. Odajima, J. Charron, T. Shioda, A. Gnirke, S. Gygi, H. Koseki, R. I. Sadreyev, A. Xiao, A. Meissner, and K. Hochedlinger, "Prolonged Mek1/2 suppression impairs the developmental potential of embryonic stem cells," *Nature*, vol. 548, no. 7666, pp. 219–223, 2017.
- [131] X. S. Liu, H. Wu, X. Ji, Y. Stelzer, X. Wu, S. Czauderna, J. Shu, D. Dadon, R. A. Young, and R. Jaenisch, "Editing DNA Methylation in the Mammalian Genome," *Cell*, vol. 167, no. 1, pp. 233–247, 2016.
- [132] X. Xu, Y. Tao, X. Gao, L. Zhang, X. Li, W. Zou, K. Ruan, F. Wang, G.-L. Xu, and R. Hu, "A CRISPR-based approach for targeted DNA demethylation," *Cell Discovery*, vol. 2, p. 16009, 2016.
- [133] K. Yoshimi, T. Kaneko, B. Voigt, and T. Mashimo, "Allele-specific genome editing and correction of disease-associated phenotypes in rats using the CRISPR-Cas platform," *Nature Communications*, vol. 5, p. 4240, 2014.
- [134] G. P. Smith, "Filamentous Fusion Phage: Novel Expression Vectors That Display Cloned Antigens on the Virion Surface," *Science*, vol. 228, no. 4705, pp. 1315–1317, 1985.
- [135] R. R. Porter, "The Hydrolysis of Rabbit γ -Globulin and Antibodies with Crystalline Papain," *Biochemical Journal*, vol. 73, no. 1, pp. 119–127, 1959.
- [136] G. M. Edelman, "Dissociation of γ -globulin," *Journal of the American Chemical Society*, vol. 81, no. 12, pp. 3155–3156, 1959.
- [137] G. M. Edelman and M. D. Poulik, "Studies on structural units of the γ -globulins," *Journal of Experimental Medicine*, vol. 113, no. 5, pp. 861–884, 1961.
- [138] T. Clackson, H. R. Hoogenboom, A. D. Griffiths, and G. Winter, "Making antibody fragments using phage display libraries," *Nature*, vol. 352, no. 6336, pp. 624–8, 1991.
- [139] G. Winter, A. D. Griffiths, R. E. Hawkins, and H. R. Hoogenboom, "Making Antibodies by Phage Display Technology," *Annual Review of Immunology*, vol. 12, no. 1, pp. 433–455, 1994.
- [140] H. R. Hoogenboom and G. Winter, "By-passing Immunisation. Human Antibodies from Synthetic Repertoires of Germline VH Gene Segments Rearranged in Vitro," *Journal of Molecular Biology*, vol. 227, no. 2, pp. 381–388, 1992.
- [141] A. M. Basson, "Signaling in Cell Differentiation and Morphogenesis," *Cold Spring Harbor Perspectives in Biology*, vol. 4, no. 6, p. a008151, 2012.
- [142] C. E. Chan, A. P. Lim, P. A. MacAry, and B. J. Hanson, "The role of phage display in therapeutic antibody discovery," *International immunology*, vol. 26, no. 12, pp. 649–657, 2014.
- [143] C. Lloyd, D. Lowe, B. Edwards, F. Welsh, T. Dilks, C. Hardman, and T. Vaughan, "Modelling the human immune response: performance of a 1011 human antibody repertoire against a broad panel of therapeutically relevant antigens," *Protein Engineering Design and Selection*, vol. 22, no. 3, pp. 159–168, 2009.

- [144] T. J. Vaughan, A. J. Williams, K. Pritchard, J. K. Osbourn, A. R. Pope, J. C. Earnshaw, J. McCafferty, R. A. Hodits, J. Wilton, and K. S. Johnson, "Human Antibodies With Subnanomolar Affinities Isolated From A Large Non-immunized Phage Display Library," *Nature Biotechnology*, vol. 14, no. 3, pp. 309–314, 1996.
- [145] X. Zhang, Y. Zhou, K. R. Mehta, D. C. Danila, S. Scolavino, S. R. Johnson, and A. Klibanski, "A Pituitary-Derived MEG3 Isoform Functions as a Growth Suppressor in Tumor Cells," *Journal of Clinical Endocrinology and Metabolism*, vol. 88, no. 11, pp. 5119–5126, 2003.
- [146] D. Astuti, F. Latif, K. Wagner, D. Gentle, W. N. Cooper, D. Catchpoole, R. Grundy, A. C. Ferguson-Smith, and E. R. Maher, "Epigenetic alteration at the DLK1-GTL2 imprinted domain in human neoplasia: analysis of neuroblastoma, pheochromocytoma and Wilms' tumour," *British Journal of Cancer*, vol. 92, no. 8, pp. 1574–1580, 2005.
- [147] C.-Y. Zhang, M.-S. Yu, X. Li, Z. Zhang, C.-R. Han, and B. Yan, "Overexpression of long non-coding RNA MEG3 suppresses breast cancer cell proliferation, invasion, and angiogenesis through AKT pathway," *Tumor Biology*, vol. 39, no. 6, p. 101042831770131, 2017.
- [148] C. Braconi, T. Kogure, N. Valeri, N. Huang, G. Nuovo, S. Costinean, M. Negrini, E. Miotto, C. M. Croce, and T. Patel, "microRNA-29 can regulate expression of the long non-coding RNA gene MEG3 in hepatocellular cancer," *Oncogene*, vol. 30, no. 47, pp. 4750–4756, 2011.
- [149] P. Wang, D. Chen, H. Ma, and Y. Li, "LncRNA MEG3 enhances cisplatin sensitivity in non-small cell lung cancer by regulating miR-21-5p/SOX7 axis," *Onco Targets and Therapy*, vol. 10, pp. 5137–5149, 2017.
- [150] Z.-Z. Tian, X.-J. Guo, Y.-M. Zhao, and Y. Fang, "Decreased expression of long non-coding RNA MEG3 acts as a potential predictor biomarker in progression and poor prognosis of osteosarcoma," *International Journal of Clinical and Experimental Pathology*, vol. 8, no. 11, pp. 15138–15142, 2015.
- [151] Y. Zhou, Y. Zhong, Y. Wang, X. Zhang, D. L. Batista, R. Gejman, P. J. Ansell, J. Zhao, C. Weng, and A. Klibanski, "Activation of p53 by MEG3 Non-coding RNA," *Journal of Biological Chemistry*, vol. 282, no. 34, pp. 24731–24742, 2007.
- [152] D. Hanahan and R. A. Weinberg, "The Hallmarks of Cancer," *Cell*, vol. 100, no. 1, pp. 57–70, 2000.
- [153] G.-Z. Qiu, W. Tian, H.-T. Fu, C.-P. Li, and B. Liu, "Long noncoding RNA-MEG3 is involved in diabetes mellitus-related microvascular dysfunction," *Biochemical and Biophysical Research Communications*, vol. 471, no. 1, pp. 135–141, 2016.
- [154] W. Su, W. Xie, Q. Shang, and B. Su, "The Long Noncoding RNA MEG3 is d Downregulated and i Inversely Associated with VEGF Levels in Osteoarthritis," *BioMed Research International*, vol. 2015, p. 356893, 2015.

- [155] F. E. Gordon, C. L. Nutt, P. Cheunsuchon, Y. Nakayama, K. A. Provencher, K. A. Rice, Y. Zhou, X. Zhang, and A. Klibanski, "Increased Expression of Angiogenic Genes in the Brains of Mouse Meg3-Null Embryos," *Endocrinology*, vol. 151, no. 6, pp. 2443–2452, 2010.
- [156] J. Liu, Q. Li, K.-S. Zhang, B. Hu, X. Niu, S.-M. Zhou, S.-G. Li, Y.-P. Luo, Y. Wang, and Z.-F. Deng, "Downregulation of the Long Non-Coding RNA Meg3 Promotes Angiogenesis After Ischemic Brain Injury by Activating Notch Signaling," *Molecular Neurobiology*, vol. 54, no. 10, pp. 8179–8190, 2017.
- [157] R. Ravi, B. Mookerjee, Z. M. Bhujwalla, C. Hayes Sutter, D. Artemov, Q. Zeng, L. E. Dillehay, A. Madan, G. L. Semenza, and A. Bedi, "Regulation of tumor angiogenesis by p53-induced degradation of hypoxia-inducible factor 1 α ," *Genes and Development*, vol. 14, no. 1, pp. 34–44, 2000.
- [158] T. D. Le Cras, R. Spitzmiller, K. Albertine, J. Greenberg, J. Whitsett, and A. Akeson, "VEGF causes pulmonary hemorrhage, hemosiderosis, and air space enlargement in neonatal mice," *American Journal of Physiology - Lung Cellular and Molecular Physiology*, vol. 287, no. 1, pp. L134–L142, 2004.
- [159] J. Rubenstein, N. Fischbein, K. Aldape, E. Burton, and M. Shuman, "Hemorrhage and VEGF expression in a case of primary CNS lymphoma.," *Journal of Neuro-Oncology*, vol. 58, no. 1, pp. 53–56, 2002.
- [160] D. Yang, J. M. Baumann, Y. Y. Sun, M. Tang, R. S. Dunn, A. L. Akeson, S. G. Kernie, S. Kallapur, D. M. Lindquist, E. J. Huang, S. S. Potter, H. C. Liang, and C.-Y. Kuan, "Overexpression of Vascular Endothelial Growth Factor in the Germinal Matrix Induces Neurovascular Proteases and Intraventricular Hemorrhage," *Science Translational Medicine*, vol. 5, no. 193, p. 193ra90, 2013.
- [161] X. Qu, K. Tompkins, L. E. Batts, M. Puri, and H. S. Baldwin, "Abnormal embryonic lymphatic vessel development in Tie1 hypomorphic mice," *Development*, vol. 137, no. 8, pp. 1285–1295, 2010.
- [162] H. G. Augustin, G. Y. Koh, G. Thurston, and K. Alitalo, "Control of vascular morphogenesis and homeostasis through the angiopoietin - Tie system," *Nature Reviews Molecular Cell Biology*, vol. 10, no. 3, pp. 165–177, 2009.
- [163] M. C. Puri, J. Rossant, K. Alitalo, A. Bernstein, and J. Partanen, "The receptor tyrosine kinase TIE is required for integrity and survival of vascular endothelial cells," *The EMBO journal*, vol. 14, no. 23, pp. 5884–5891, 1995.
- [164] S. Kuijper, C. J. Turner, and R. H. Adams, "Regulation of angiogenesis by Eph-Ephrin interaction," *Trends in Cardiovascular Medicine*, vol. 17, no. 5, pp. 145–151, 2007.
- [165] N. Takahashi, A. Okamoto, R. Kobayashi, M. Shirai, Y. Obata, H. Ogawa, Y. Sotomaru, and T. Kono, "Deletion of Gtl2, imprinted non-coding RNA, with its differentially methylated region induces lethal parent-origin-dependent defects in mice," *Human Molecular Genetics*, vol. 18, no. 10, pp. 1879–1888, 2009.

- [166] S.-P. Lin, N. Youngson, S. Takada, H. Seitz, W. Reik, M. Paulsen, J. Cavaille, and A. C. Ferguson-Smith, "Asymmetric regulation of imprinting on the maternal and paternal chromosomes at the Dlk1-Gtl2 imprinted cluster on mouse chromosome 12," *Nature Genetics*, vol. 35, no. 1, pp. 97–102, 2003.
- [167] L. T. Krebs, Y. Xue, C. R. Norton, J. R. Shutter, M. Maguire, J. P. Sundberg, D. Gallahan, V. Closson, J. Kitajewski, R. Callahan, G. H. Smith, K. L. Stark, and T. Gridley, "Notch signaling is essential for vascular morphogenesis in mice," *Genes and Development*, vol. 14, no. 11, pp. 1343–1352, 2000.
- [168] Y. Xue, X. Gao, C. E. Lindsell, C. R. Norton, B. Chang, C. Hicks, M. Gendron-Maguire, E. B. Rand, G. Weinmaster, and T. Gridley, "Embryonic lethality and vascular defects in mice lacking the Notch ligand Jagged1," *Human Molecular Genetics*, vol. 8, no. 5, pp. 723–730, 1999.
- [169] A. Fischer, N. Schumacher, M. Maier, M. Sendtner, and M. Gessler, "The Notch target genes *Hey1* and *Hey2* are required for embryonic vascular development," *Genes and Development*, vol. 18, no. 8, pp. 901–911, 2004.
- [170] M. Gasperowicz and F. Otto, "The Notch Signalling Pathway in the Development of the Mouse Placenta," *Placenta*, vol. 29, no. 8, pp. 651–659, 2008.
- [171] Y. Zhou, P. Cheunschon, Y. Nakayama, M. W. Lawlor, Y. Zhong, K. A. Rice, L. Zhang, X. Zhang, F. E. Gordon, H. G. W. Lidov, R. T. Bronson, and A. Klibanski, "Activation of paternally expressed genes and perinatal death caused by deletion of the *Gtl2* gene," *Development*, vol. 137, no. 16, pp. 2643–2652, 2010.