

# Genetic Dissection of the Exit from Pluripotency in Mouse Embryonic Stem Cells by CRISPR-based Screening

Meng Li



Newnham College  
Wellcome Trust Sanger Institute  
University of Cambridge

This dissertation is submitted for the degree of

*Doctor of Philosophy*

September 2017

## **Declaration**

This dissertation describes work carried out from June 2014 to July 2017 under the supervision of Dr Kosuke Yusa at the Wellcome Trust Sanger Institute.

I hereby declare that this dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except where specifically indicated in the text.

I 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.

This thesis does not exceed the prescribed word limit of 60,000 words as set by the Degree Committee for the Faculty of Biology.

Meng Li

30th September 2017

## **Acknowledgements**

I would like to thank foremost my supervisor Kosuke Yusa, for his guidance and support during the past four years. I consider myself extremely lucky to have a supervisor who cared so much about my work and provided the best supervision he possibly can. His immense knowledge and great passion for science was and will always be a source of inspiration to me.

I would also like to thank the rest of my thesis committee: Prof Allan Bradley and Dr Pentao Liu, for the insightful discussions and encouragement. I am very grateful to my external supervisor Prof Austin Smith, who has been a great source of advice throughout my PhD. I also thank Dr Martin Leeb and Dr Meng Amy Li, for sharing their knowledge on genetic screening and ESC culture.

I am thankful to everybody in the the Yusa lab. I would particularly like to thank Hiroko, who is one of the warmest and kindest person I have ever met. She has given me enormous help and encouragement. This journey would not be as great without her. Jason also deserves special mention for helping me in the lab and always being there to give advice. I thank Kasia, Shannade, Yanzi, Leandro and Swee Hoe for all the help and support they have given me. Thanks also to the Vallier lab members especially Sapna for always being there to cheer me up. I would also like to thank the members of staff at the Flow Cytometry Facility, without whom all the large-scale cell sorting would not be possible.

I am also very glad for having had the luck of being surrounded by great friends. Special thanks to Ziyang, Giulia and Haruna for all the warm encouragement and support. Thanks also to Tom for proof-reading part of this thesis.

Finally, I would like to express my gratitude to my parents, for their constant encouragement and unconditional love. Without their support I would never have gotten to where I am today.

## Abstract

The ground state naive pluripotency is established in the epiblast of the blastocyst and can be captured by culturing mouse embryonic stem cells (mESCs) with MEK and GSK3 inhibitors (2i). The transcription network that maintains pluripotency has been extensively studied with the indispensable core factors being Oct4, Sox2 and Nanog, together with other ancillary factors reinforcing the network. However, how this network is dissolved at the onset of differentiation is still not fully understood. To identify genes required for differentiation in an unbiased fashion, I conducted a genome-wide CRISPR-Cas9-mediated screen in Rex1GFPd2 mESCs. This cell line expresses GFP specifically in the naive state and rapidly down-regulate upon differentiation. I differentiated mutagenised mESCs for two days and sorted mutants that kept higher GFP expression. gRNA representation was subsequently analysed by sequencing. I identified 563 and 8 genes whose mutants showed delayed and accelerated differentiation, respectively, at a false discovery rate (FDR) cut-off of 10%. The majority of the previously known genes were identified in my screen, suggesting faithful representation of genes regulating differentiation. Detailed screening result analysis revealed a comprehensive picture of pathways involved in the dissolution of naive pluripotency. Amongst the genes identified are 19 mTORC1 regulators and components of the mTORC2 complex. Deficiency in the TSC and GATOR complexes resulted in mTORC1 upregulation in consistent with previous studies. However, they showed opposite phenotype during ESC differentiation: TSC complex knockout cells showed delayed differentiation, whereas GATOR1 deficiency accelerated differentiation I found that the pattern of GSK3b phosphorylation is highly correlated with differentiation phenotype. I conclude that mTORC1 is involved in pluripotency maintenance and differentiation through cross-talk with the Wnt signalling pathway. My screen has demonstrated the power of CRISPR-Cas9-mediated screen and provided further insights in biological pathways involved in regulating differentiation. It would be interesting to explore the remaining unstudied genes for better understanding of the mechanisms underlying mESC differentiation.

## Abbreviation

2i	2 inhibitors
AP	Alkaline phosphate
<i>Blm</i>	Bloom's syndrome
BMP	Bone morphogenetic protein
cDNA	complementary DNA
ChIP	Chromatin immunoprecipitation
CRISPR	Clustered regularly interspaced short palindromic repeats
crRNA	CRISPR RNA
CRISPRi	CRISPR interference
CRISPRa	CRISPR-mediated gene activation
dCas9	nuclease-deficient Cas9
DE score	Depletion/Enrichment score
DSB	Double stranded break
dsRNA	double stranded RNA
DT-A	Diphtheria toxin fragment-A
EC	Embryonic carcinoma
EGC	Embryonic germ cells
ENU	N-ethyl-N-nitrosourea
EpiSC	Epiblast stem cell
ESC	Embryonic stem cell
ESCRT	Endosomal sorting complexes required for transport
FACS	Fluorescence-activated cell sorting
FBS	Fetal bovine serum
FDR	False discovery rate
GAP	GTPase-activating protein
GATOR	Gap activity toward rags
GeCKO	CRISPR-Cas9 Knockout
gRNA	guide RNA
GSEA	Gene set enrichment analysis
HDR	Homologous directed repair
HOPS	Homotypic fusion and protein sorting
HR	Homologous recombination
ICM	Inner cell mass
indels	Insertion or deletions
iPSC	Induced pluripotent stem cell

---

KSR	Knockout serum replacement
LIF	Leukemia inhibitory factor
LncRNA	Long non-coding RNA
LTR	Long terminal repeat
M15L	Knockout DMEM supplemented with 15%FBS and LIF
m <sup>6</sup> A	N <sup>6</sup> A-methyladenosine
MAGeCK	Model-based analysis of genome-wide CRISPR/Cas9 knockout
MMR	Mismatch repair
MOI	Multiplicity of infection
mTOR	Mammalian target of rapamycin
mTORC1/2	mTOR complex 1/2
NGS	Next-generation sequencing
NHEJ	Non-homologous end joining
NMD	Non-sense mediated decay
NuRd	Nucleosome remodeling deacetylase
OSN	Oct4/Sox2/Nanog
RA	Retinoid acid
RIGER	RNAi gene enrichment ranking
RNP	Ribonucleoproteins
RSA	Redundant siRNA activity
PAM	Protospacer adjacent motif
PB	piggyBac
PGC	Primordial germ cells
PRC	Polycomb repressive complex
RISC	RNA-inducing silencing complex
RNAi	RNA interference
RVDs	Repeat variable diresidues
SAM	synergistic activation mediator
SB	Sleeping beauty
siRNA	short interfering RNA
shRNA	short hairpin RNA
ssODNs	single-strand oligodeoxynucleotides
TALEN	Transcription activator-like effector nuclease
tracrRNA	trans-encoded smallRNA
TKO	Toronto knockOut
TSC	Tuberous sclerosis complex
XEN	Extra-embryonic endoderm
ZFN	Zinc finger nuclease

# Contents

<b>1</b>	<b>Introduction</b>	<b>14</b>
1.1	Reverse Genetics	15
1.1.1	Gene targeting	15
1.1.2	RNA interference (RNAi)	16
1.1.3	Genome engineering with programmable nucleases	17
1.1.3.1	The repair pathways and applications	18
1.1.3.2	Programmable nucleases before the CRISPR era	19
1.1.3.2.1	ZFN	19
1.1.3.2.2	TALEN	20
1.1.3.3	CRISPR-Cas systems	21
1.1.3.3.1	The discovery of CRISPR-Cas systems	21
1.1.3.3.2	The diversity of CRISPR-Cas systems	21
1.1.3.3.3	CRISPR-Cas9 system as a genome-editing tool	23
1.1.3.3.4	Structure and working mechanism of Cas9	25
1.1.3.3.5	Off-target effect	26
1.1.3.3.6	Applications of CRISPR-Cas9 system	26
1.2	Forward genetics	29
1.2.1	Mutagenesis using chemical and physical agents	29
1.2.1.1	N-ethyl-N-nitrosourea (ENU)	29
1.2.1.2	Irradiation	30
1.2.2	Insertional mutagenesis	30
1.2.2.1	Retroviral-mediated mutagenesis	31
1.2.2.2	DNA transposon-mediated mutagenesis	31
1.2.3	The use of <i>Blm</i> -deficient and haploid cell lines in genetic screens	32
1.2.3.1	<i>Blm</i> -deficient ESC systems	32
1.2.3.2	Haploid cell lines	32

1.2.4	RNAi-mediated screens . . . . .	33
1.2.5	CRISPR-Cas9-mediated screens . . . . .	34
1.2.5.1	The establishment of CRISPR-Cas9 screening technology . . . . .	34
1.2.5.2	Screening format . . . . .	35
1.2.5.2.1	Arrayed screening . . . . .	35
1.2.5.2.2	Pooled screening . . . . .	35
1.2.5.3	Applications of CRISPR-Cas9-mediated screens . . . . .	36
1.2.5.4	Experimental design of CRISPR-Cas9-mediated screens . . . . .	38
1.2.5.4.1	gRNA design . . . . .	38
1.2.5.4.2	Cas9 and gRNA delivery . . . . .	39
1.2.5.4.3	Gene identification and data analysis . . . . .	40
1.3	Embryonic stem cells (ESCs) . . . . .	42
1.3.1	Early development of mouse embryo . . . . .	42
1.3.2	Derivation of mouse ESCs . . . . .	43
1.3.3	Regulation of the pluripotency state . . . . .	45
1.3.3.1	Extrinsic signalling pathways . . . . .	46
1.3.3.1.1	LIF-mediated signalling pathway . . . . .	46
1.3.3.1.2	TGF- $\beta$ -mediated pathway . . . . .	48
1.3.3.1.3	FGF/MAPK pathway . . . . .	48
1.3.3.1.4	Wnt signalling pathway . . . . .	49
1.3.3.1.5	Serum/LIF culture and 2i culture . . . . .	51
1.3.3.2	Transcription factor network . . . . .	52
1.3.3.2.1	Core pluripotency factors . . . . .	52
1.3.3.2.2	Ancillary pluripotency regulators . . . . .	55
1.3.3.3	Epigenetic regulation . . . . .	56
1.3.3.3.1	DNA methylation . . . . .	56
1.3.3.3.2	Bivalent domains . . . . .	57
1.3.3.3.3	Heterochromatin organisation . . . . .	58
<b>2</b>	<b>Materials and Methods</b>	<b>59</b>
2.1	Cell Culture . . . . .	60
2.1.1	Materials . . . . .	60
2.1.1.1	Cell lines . . . . .	60
2.1.1.2	Media components, inhibitors and other reagents used in cell culture . . . . .	60



---

2.1.1.3	Dissociation agents . . . . .	61
2.1.1.4	Other chemicals and kits . . . . .	61
2.1.1.5	Media . . . . .	61
2.1.2	Methods . . . . .	62
2.1.2.1	Routine culture and maintenance mESCs . . . . .	62
2.1.2.2	Differentiation condition . . . . .	62
2.1.2.3	Transfection of mESCs . . . . .	62
2.1.2.4	Lentivirus production and transduction . . . . .	63
2.1.2.5	Genetic manipulation of cell lines . . . . .	63
2.1.2.6	Commitment assay . . . . .	64
2.1.2.7	Flow cytometry and cell sorting . . . . .	65
2.2	Molecular Biology . . . . .	66
2.2.1	Materials . . . . .	66
2.2.1.1	Molecular chemicals and Kits . . . . .	66
2.2.1.2	Immunoblotting . . . . .	67
2.2.2	Methods . . . . .	69
2.2.2.1	Isolation of nucleic acids . . . . .	69
2.2.2.2	gRNA cloning . . . . .	69
2.2.2.3	Lysate PCR . . . . .	69
2.2.2.4	Illumina library preparation . . . . .	70
2.2.2.5	Western blotting . . . . .	71
2.3	Bioinformatics analysis . . . . .	72
2.3.0.1	RNA-Seq analysis . . . . .	72
2.3.0.2	gRNA sequencing result analysis . . . . .	72
2.3.0.3	Gene Set Enrichment Analysis (GSEA) analysis . . . . .	72
<b>3</b>	<b>Screening Preparation and Condition Optimisation</b>	<b>73</b>
3.1	Introduction . . . . .	74
3.2	Results . . . . .	75
3.2.1	Analysis of self-renewal and differentiation conditions . . . . .	75
3.2.2	Establishment of Cas9 expression in Rex1:GFPd2 cell line . . . . .	78
3.2.3	Proof of principle studies . . . . .	80
3.2.4	A Preliminary screen . . . . .	82
3.3	Discussion and Conclusion . . . . .	85

<b>4</b>	<b>Screening Result and Pathway Analysis</b>	<b>88</b>
4.1	Introduction . . . . .	89
4.2	Results . . . . .	92
4.2.1	Screening Strategy and Result Analysis . . . . .	92
4.2.2	Screening Result Validation . . . . .	94
4.2.3	Pathways Analysis . . . . .	96
4.2.3.1	Signalling pathways . . . . .	96
4.2.3.2	mRNA related pathways . . . . .	98
4.2.3.3	Chromatin modifiers . . . . .	99
4.2.3.4	Transcription factors and other transcriptional regulatory proteins . . . . .	101
4.2.3.5	Mitochondria-related pathways . . . . .	102
4.2.3.6	Endocytosis and vesicle trafficking . . . . .	103
4.2.4	Discussion and Conclusion . . . . .	105
4.2.4.1	Functional genomics with CRISPR-Cas9 . . . . .	105
4.2.4.2	Integration of the regulatory network . . . . .	106
4.2.4.3	Metabolic regulation . . . . .	107
<b>5</b>	<b>The Role of mTOR-related Pathways in Pluripotency/Differentiation Regulation</b>	<b>108</b>
5.1	Introduction . . . . .	109
5.2	Results . . . . .	114
5.2.1	Generation of knockout cell lines and phenotype validation . . . . .	114
5.2.2	Disruption of GATOR1 complex induces mTORC1-activated nega- tive feedback loop . . . . .	117
5.2.3	<i>Tsc2</i> knockout demonstrated unconventional cross-talk between mTOR and PI(3)K-Akt pathways . . . . .	120
5.2.4	Transcriptional changes in <i>Nprl2</i> and <i>Tsc2</i> knockouts . . . . .	122
5.3	Discussion and conclusion . . . . .	126
<b>6</b>	<b>Conclusions and Future Perspectives</b>	<b>128</b>
6.1	The immeasurable potential of CRISPR-Cas9-mediated genetic screen in stem cell biology . . . . .	129
6.2	The complex regulatory network that drives dissolution of pluripotency . .	132

6.3 The prominent role of mTOR and its related proteins in pluripotency and differentiation regulation . . . . .	134
--	-----

# List of Figures

1.1	CRISPR-mediated DNA interference in microbial adaptive immunity. . . .	22
1.2	The mechanism of CRISPR-mediated genome engineering. . . . .	24
1.3	Overall structure of <i>Streptococcus pyogenes</i> Cas9 and gRNA. . . . .	25
1.4	Lineage segregation in mouse blastocyst. . . . .	44
1.5	Extrinsic signalling pathways that regulate pluripotency. . . . .	46
3.1	Analysis of Rex1GFP profile under maintenance conditions. . . . .	76
3.2	Analysis of the Rex1GFP profile under differentiation conditions. . . . .	77
3.3	Cas9 knockin . . . . .	78
3.4	Cas9 function assay. . . . .	79
3.5	Positive control study with stable knockout clones. . . . .	81
3.6	Positive control study with single-gRNA knockouts. . . . .	83
3.7	Preliminary screen result analysis. . . . .	86
4.1	Screening Strategy and Result Analysis. . . . .	93
4.2	Screening Result Validation. . . . .	95
4.3	Cellular pathways underpinning the initiation of differentiation. . . . .	97
4.4	Rex1GFP differentiation profile of Oct4 knockout mESCs. . . . .	101
4.5	Validation of genes in mitochondria-related pathways. . . . .	104
5.1	Structure of mTORC1 and mTORC2. . . . .	109
5.2	Structure of mTOR regulating complexes and mechanistic target of mTORC1 amino acid sensing pathway. . . . .	111
5.3	Generation of stable KO cell lines for mTORC1 regulating genes. . . . .	115
5.4	mTOR-related gene knockout phenotype validation. . . . .	116
5.5	GATOR1 KO phenotype resulted from activated GSK3. . . . .	119

5.6	The unconventional cross-talk between mTOR and PI(3)K-Akt pathways in <i>Tsc2</i> KO. . . . .	121
5.7	Transcriptional changes in <i>Nprl2</i> and <i>Tsc2</i> knockouts. . . . .	123
5.8	Pathway analysis of differentially expressed gene sets. . . . .	125
5.9	Summary of major relevant signalling components upstream and down- stream of mTORC1 and mTORC2. . . . .	126

# Chapter 1

## Introduction

## 1.1 Reverse Genetics

Reverse genetics refers to a gene-driven approach, which analyses the phenotypic consequences of directed mutations of a target gene. Emerged in the end of 1980s, reverse genetics set off since the development of transgenic organisms, where a genotype is designed, constructed *in vitro* and introduced to the mouse germline. The resulting transgenic animals display a phenotype dependent on the location and design of the mutation, which allows characterisation of a gene and eventual understanding of its underlying biology [213]. Since the 1990s, a large number of genes have been cloned and their knockout animals were generated. After the completion of mouse genome project, high-throughput reverse genetics became the major approach [151]. In this section, I will describe different strategies in reverse genetics, from homologous recombination (HR)-mediated modifications to RNA interference, and finally, the use of programmable nucleases, including zinc finger nucleases (ZFNs), transcription activator-like effectors nucleases (TALENs) and the revolutionary clustered regularly interspaced short palindromic repeats (CRISPR)-Cas9 technology.

### 1.1.1 Gene targeting

Gene targeting, defined as the introduction of site-specific modification into the genome by HR, has enabled genetic manipulations ranging from simple gene disruptions and point mutations to insertions and inversions, even conditional knockouts or knockins [32] [284]. The principle of gene targeting was developed in yeast, where DNA fragments with homology to yeast DNA sequence can integrate into its genome [306]. The first HR-based mutagenesis in mammalian cells was achieved by Smithies et al. showing successful integration by HR of a plasmid into the *beta*-globin locus of human erythroleukaemia cells [386]. In parallel with Smithies's work, Capecchi and colleagues independently achieved HR-mediated repair of a defective neomycin resistant gene in transformed mouse fibroblast cell lines [415].

Subsequent to the derivation and establishment of mouse embryonic stem cell (ESC) culture and the demonstration of its germ line transmission capacity [107] [34], the first HR in mouse ESCs was achieved at the selectable hypoxanthine phosphoribosyl transferase (Hprt) gene locus [94][414]. The targeting of non-selectable genes such as *Int-2* and *c-Abl* later also became possible [254][365]. These early experiments have provided a basis for the generation of genetically modified mice, which became invaluable tools for understanding

the functions of mammalian genes at the organism level and producing models for human diseases.

A targeting vector typically consists of three basic portions: a 5' homology arm, a positive selectable marker such as neomycin resistance gene (*neo*) and a 3' homology arm. Successfully targeted cells are positively selected by neomycin (G418) or other antibiotics such as puromycin or hygromycin depending on the selectable marker used. HR clones can be further enriched using the negative selection markers, such as the diphtheria toxin fragment A(DT-A) or thymidine kinase(TK), which are placed outside the homology arms. If the targeting vector is integrated randomly in the genome, the negative selection marker is also most likely to be retained and exerts toxicity, which will lead to cell death. Site-specific recombinases such as Cre or FLP are routinely used to remove the selection marker genes to leave minimal impact at the locus. The expression of these recombinases can be controlled in a time- and/or tissue- dependent manner using the Tamoxifen-ERT2 system and tissue specific promoter, respectively, which are particularly useful for a detailed study of genes whose inactivation would be otherwise lethal.

### 1.1.2 RNA interference (RNAi)

RNAi is defined as the process of suppression of the expression of a target gene via specific destruction of its mRNA by exogenous or endogenous double stranded RNA (dsRNA) [100]. Its ability of gene silencing in a sequence-specific manner has made it a powerful tool for investigating the function of a gene. The RNAi pathway was first discovered by Fire and Montgomery in 1998 when they injected long dsRNA into *Caenorhabditis elegans* and observed specific cytoplasmic degradation of the mRNA molecules containing the same sequence as the injected dsRNA [114] [280]. The use of long dsRNA was soon proven to be effective in flies and plants but not in mammalian cells due to a nonspecific interferon response [379] [389]. Subsequently, it was discovered that the use of 21-28 nucleotide short interfering RNAs (siRNAs) or short hairpin shRNAs (shRNAs) can prevent the interferon response and achieve sequence-specific silencing in mammalian cells [102] [43].

To function as a silencing effector, siRNAs need to be processed from longer precursors by a member of the RNase III family called Dicer, in an ATP-dependent manner [18]. Processed siRNAs are typically 21-23 nucleotides long, which are subsequently loaded onto a group of Argonaute proteins called RNA-Inducing Silencing Complex (RISC) with some help from Dicer [149] [103]. The loaded RNA duplex then unwinds itself, leaving



one strand as a guide for target recognition, whereas the other passenger strand gets discarded [478]. When the siRNA-guided RISC reaches its complementary RNA target, an endonucleolytic cleavage is induced by the PIWI domain of RISC at the 10th nucleotide counting from the 5' end [478]. Following the initial cut, the mRNA target dissociates from the siRNA, and cellular exonucleases join in to complete the degradation process [149]. It has been observed that certain imperfect matches between siRNA and mRNAs can be tolerated, which causes the off-target effect [357].

siRNAs can be chemically synthesised and introduced into mammalian cells directly by transfection. The transfected siRNAs bypass the dicing step and directly incorporate into RISC for target mRNA degradation [357]. However, synthesised siRNAs have several drawbacks such as the expensive chemical synthesis process and low transfection efficiency in certain cell types. Furthermore, the siRNA molecules are unstable and become diluted as cell divide, which cannot constitutively sustain stable gene knockdown. To circumvent these problems, a plasmid-based system was developed, where siRNAs are expressed in the form of short hairpin RNAs (shRNAs) [43]. Once the expression cassette is introduced into the cell, shRNA is constitutively transcribed by RNA polymerase III promoter and forms a stem-loop structure, which is processed by Dicer and other RNAi-related machineries to into a 20-25 nucleotides double-stranded siRNA [43]. The processed siRNAs can then be loaded to RISC and carry out target mRNA degradation as described [43]. As an alternative method to plasmid transfection, retro- and lentiviral transduction are frequently used for shRNA delivery [459] [446] [42]. The transduction efficiency can be optimised to a high level, to nearly 100% in some cell types [100]. Furthermore, it was shown that the shRNA constructs can be designed to be embedded in the context of endogenous miRNA precursor sequence, improved the knockdown efficiency up to 12-fold higher [88] [376].

### 1.1.3 Genome engineering with programmable nucleases

Conventional gene targeting via HR is a powerful approach to achieve gene inactivation and enable gene function interrogation. However, it is usually a tedious process given that the efficiency of HR is extremely low in higher eukaryotic cells, which lead to the need for the labour-intensive and time-consuming selection/screening procedure. A study using a rare-cutting endonuclease, I-SceI, showed that the gene targeting efficiency increased by more than 2 orders of magnitude with the expression of endonuclease [337]. This observation provided the first evidence that HR is stimulated by the introduction of DNA double-

stranded break (DSB). From there onwards, targeted genome engineering became widely adopted, which allows precise and efficient genome editing via inducing a site-specific DSB followed by generation of desired modifications during subsequent DNA repair.

### **1.1.3.1 The repair pathways and applications**

DNA DSBs are potentially lethal to cells. Generally, they are repaired via one of the two major mechanisms: non-homologous end joining (NHEJ) or homology directed repair (HDR). NHEJ-mediated DSB repair involves direct ligation of the broken ends, which is error-prone, and often results in small insertions or deletions (indels). HDR is a template dependent pathway, which allows perfect restoration of the broken ends. Thus HDR has been exploited to achieved genetic modifications such as targeted gene insertion, correction and point mutation.

#### **Gene disruption**

Gene disruption can be achieved by the error-prone repair pathway NHEJ. Indels generated by NHEJ often give rise to frameshifts in the protein coding region, which result in premature termination followed by non-sense mediated decay, and the final consequence of gene knockout.

#### **Gene addition or tag ligation**

By co-transfecting a site-specific nuclease with a targeting vector bearing locus-specific homology arms, the transgene can be efficiently incorporated into the desired site. Alternatively, using specific nuclease that generates defined overhangs, large transgenes (up to 200kb) can be inserted into the targeted loci via NHEJ-mediated ligation.

#### **Point mutation or gene correction**

Targeting vectors or single-strand oligodeoxynucleotides (ssODNs) can be co-delivered with programmable nucleases to correct point mutations or introduce single-nucleotide variations via the HDR-mediated repair pathway. Compared to the use of targeting vector, ssODNs are much simpler in design and can be synthesised in a few days. Such advance in technology has greatly enhanced the efficiency of disease modelling, and could potentially be applied in cell and gene therapy.

#### **Chromosomal rearrangement**

Two simultaneous DSBs made on the same chromosome can result in chromosomal deletion, inversion, duplication or other rearrangements. If DSBs are introduced on two different chromosomes, chromosomal translocation can be achieved, which opens up oppor-

tunities for creating models for genetic defects caused by large chromosomal rearrangements.

### 1.1.3.2 Programmable nucleases before the CRISPR era

#### 1.1.3.2.1 ZFN

Although the study performed by Rouet et al. with I-SceI demonstrated improved targeting efficiency, type II restriction enzymes are not suitable for introducing unique DSBs in eukaryotic genomes due to their short recognition sites. A novel restriction endonuclease with longer recognition site, preferably 16-18bp in length, is required for the general use of gene targeting in eukaryotes. ZFN was the first programmable nuclease that demonstrated the potential to cleave any arbitrary DNA sequences [197]. It is composed of a DNA binding domain, which is adapted from the prevalent class of eukaryotic transcription factor – zinc finger proteins (ZFPs), and a nuclease domain derived from the restriction enzyme FokI.

The versatility of ZFN arises from its DNA binding module ZFPs, which typically contains a tandem array of Cys2-His2 fingers [273]. Each zinc finger (ZF) is composed of approximately 30 amino acid residues folded to a unique  $\beta\beta\alpha$  structure that is stabilised by a zinc ion. The crystal structure suggested that ZF binds DNA by inserting its  $\alpha$ -helix into the major groove of the double helix and recognises a 3bp sequence via making contact of the amino acids within the  $\alpha$ -helix and their target 3 nucleotides [312].

This modular structure has made ZFP a suitable component for the design of custom DNA binding protein. Facilitated by the discovery of a highly conserved linker sequence, researchers were able to generate ZFPs for a specific DNA sequence by identifying individual ZF modules for each triplet component and link them together. However, it soon became clear that the recognition of DNA by ZFs is not truly independent or modular, and that each ZF's activity is largely influenced by its neighbours [167][447][274]. To circumvent the constraints of simple modular assembly, strategies to generate context dependence of ZF modules, such as oligomerized pool engineering (OPEN) and context-dependent assembly (CoDA), were developed [244][349].

FokI was identified as a desirable subunit for generating programmable nuclease because its sequence recognition domain and endonuclease domain are structurally separated. This provides an opportunity for swapping the recognition domain with other DNA-binding proteins [59]. The FokI nuclease domain must form a dimer to cleave DNA [26], there-

fore two ZFNs are required to bind adjacent sites with appropriate spacing for efficient dimerisation. The dimerisation process increased DNA binding stringency, resulting in increased specificity. However, off-target cleavages can still arise from the homodimerisation of FokI monomers [275] [398]. To increase specificity further, FokI domain was engineered to cleave only as a heterodimerised pair [275] [398].

ZFN was first applied to *Drosophila melanogaster* [23][22]. Since then it has been used to modify endogenous genes in a wide range of organisms such as frog oocytes, mice, rats, plants, zebrafish and nematodes, as well as cultured cells such as human cancer cell line, mouse ES cells, human ES cells, and human iPS cells [21][230] [133] [51] [268] [283] [97] [139] [234] [493]. Furthermore, ZFN has also been applied in the development of novel therapy. The first clinical trial using ZFN to target the *CCR5* gene in T cells from HIV-infected patients has already been completed in 2014 [407].

#### 1.1.3.2.2 TALEN

After more than a decade of research and development on ZFNs, the discovery of a simpler modular DNA recognition protein, namely transcription activator-like effectors (TALEs), provided an alternative platform as a customisable endonuclease for genome editing. TALEs are naturally occurring proteins from the plant pathogenic bacteria *Xanthomonas* [27]. The ability of these proteins to bind DNA was first discovered in 2007 [336]. The binding process is mediated by an array of highly conserved 33-35 amino acid repeats, each of which recognises a single base pair in the major groove. The nucleotide specificity of each repeat is determined by two hypervariable amino acids positioned at 12 and 13, which are named as repeat variable diresidues (RVDs) [245] [82].

The discovery of TALEs attracted great interests in the field, and its DNA recognition code was deciphered shortly afterwards [28] [145]. Four different RVD residues NN, NI, HD and NG are the most widely used for the recognition of G, A, C and T, respectively [179]. Subsequently, chimeric TALE nucleases (TALENs) were generated by combining the TALE-based DNA recognition domain and the FokI nuclease domain [74]. Like ZFNs, TALENs work as pairs with the DNA binding sites designed to locate 12-25bp apart [74].

The one-to-one correspondence of TALE-DNA binding repeats provided greater design flexibility than triplet-confined ZFNs, which renders TALENs to be designed to target almost any given DNA sequences [195]. With a comparable targeting efficiency to

ZFNs, TALENs seem to be an easier option for non-specialist researchers [348] [333] [157]. However, due to the extensive identical repeats, expression vector construction could be challenging. To overcome this problem, several strategies have been developed such as the ‘Golden Gate’ cloning system, high-throughput solid-phase assembly and ligation-independent cloning techniques [54] [333] [37] [362].

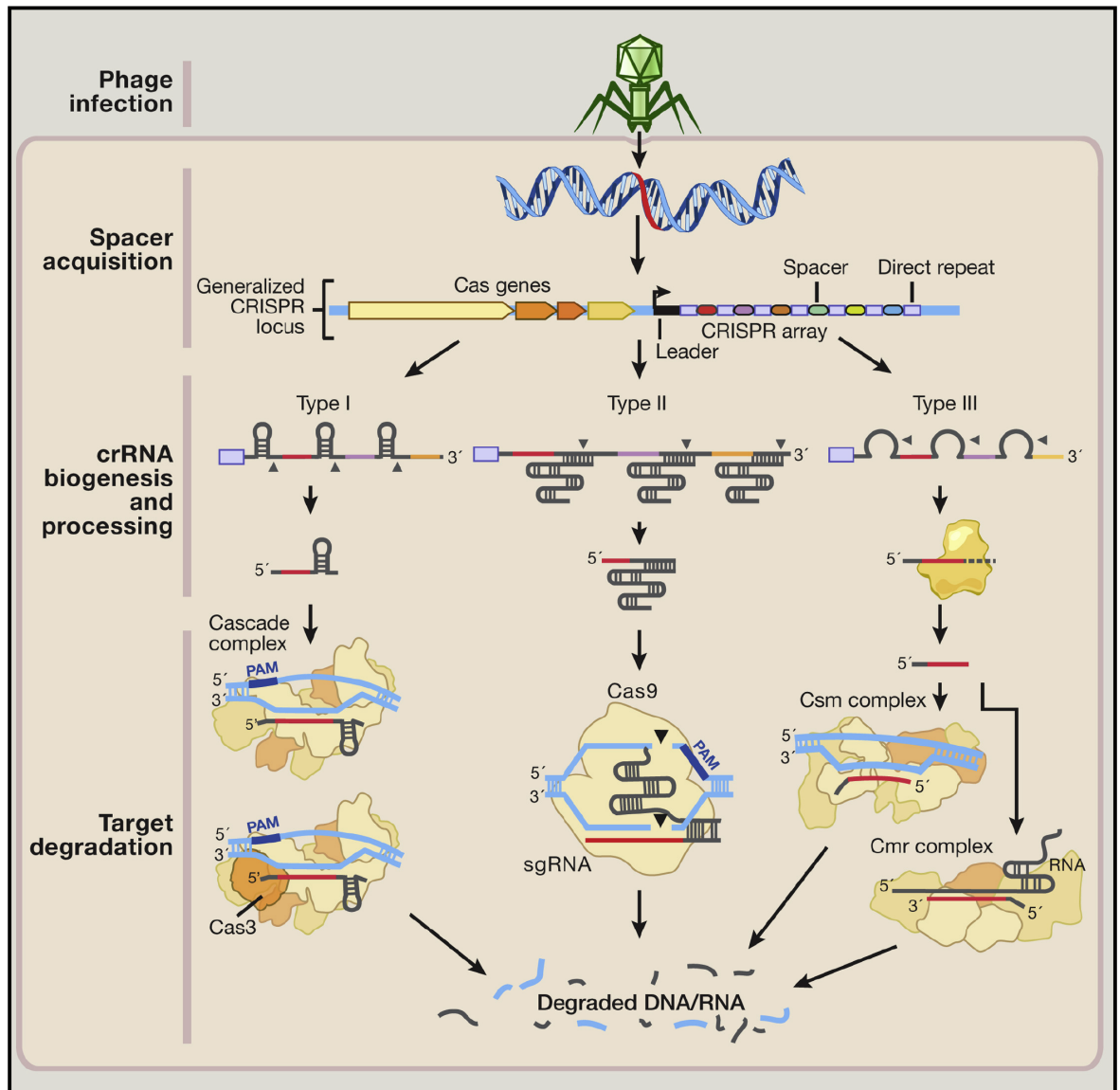
### 1.1.3.3 CRISPR-Cas systems

**1.1.3.3.1 The discovery of CRISPR-Cas systems** The CRISPR repeats were first identified in *Escherichia coli* in a study of *iap* enzyme in 1987. Ishino et al. observed an unusual structure of five highly homologous sequences of 29 nucleotides arranged as direct repeats with non-repetitive 32 nucleotides interspacing [168]. The biological significance of such structure remained elusive at the time. Over a decade later, Mojica et al. reported the wide spread of such short regularly spaced repeats among prokaryotic genomes [278]. Subsequently, Jansen et al. named these short repeats as clustered regularly interspaced short palindromic repeats (CRISPR) and identified the CRISPR-associated (Cas) genes [172].

In 2005, three groups independently published results showing the homology of CRISPR spacers with extrachromosomal elements such as phages and plasmids [279] [324] [29]. This observation, together with the evidence of the correlation between phage resistance and the CRISPR spacers, suggested that the acquisition of CRISPR elements may be related to foreign DNA invasion and CRISPR may function as a bacterial adaptive immune system [279] [29] [247]. This hypothesis was soon proved by Barrangou et al., who have demonstrated that the removal or addition of particular CRISPR spacers modified the phage resistance phenotype of the bacteria [12]. The natural mechanism of CRISPR system as part of the adaptive immunity is shown in Figure 1.1.

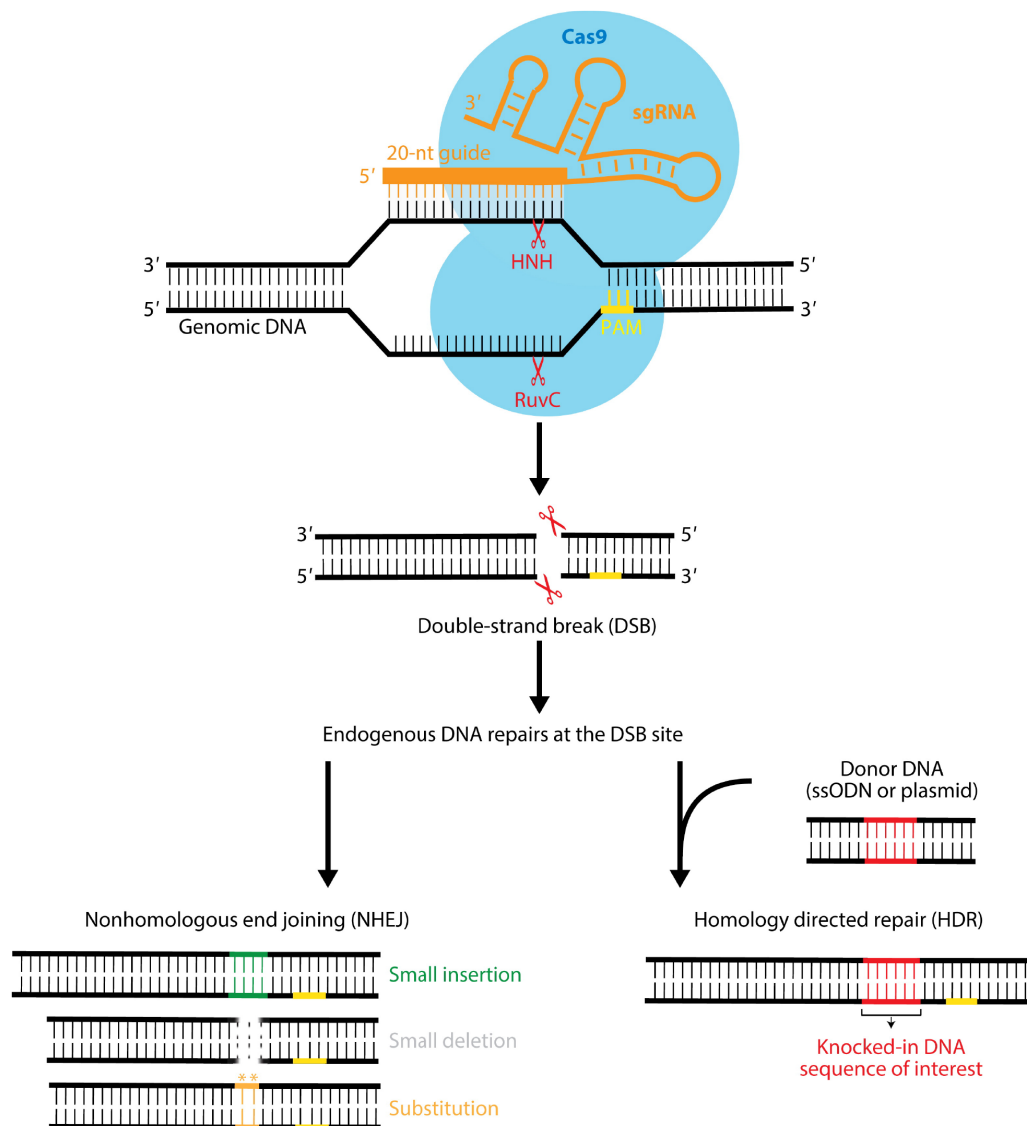
**1.1.3.3.2 The diversity of CRISPR-Cas systems** The CRISPR-Cas system has been classified into six types based on the configuration of their effector modules [248] [452] [373]. The six types can be further grouped into two major classes. Type I, III, and IV are considered as Class 1 system, where the targeted cleavage requires several Cas proteins and crRNAs to form an effector complex [248] [452]. Type II, V, and VI are grouped into the Class 2 system, where all functions of the effector complex are carried out by a single RNA-guided endonuclease, such as Cas9 [248] [452]. The utilisation of a single-component effector protein makes the Class 2 system a favourable choice to exploit

in genome editing applications. Among all the subtypes in the class 2 system, Cas9 from type II has been the most extensively characterised effector and widely utilised as a genome engineering tool.



**Figure 1.1:** CRISPR-mediated DNA interference in microbial adaptive immunity. A typical CRISPR locus comprises a set of Cas9 genes, a unique noncoding RNA called the trans-activating CRISPR RNA (tracrRNA), and an array of repetitive sequences interspaced by a range of non-repetitive sequences referred as spacers. Following the invasion of foreign genetic elements from bacteriophages or plasmids, the Cas enzymes acquire new spacers into the exogenous protospacer sequences and install them into the CRISPR locus within the prokaryotic genome. The crRNA biogenesis and processing follow distinct pathways in each type of CRISPR system. In type I and III CRISPR systems, the pre-crRNA transcript is cleaved within the repeats and further processed to produce matured crRNA before being loaded onto effector proteins complexes for target recognition and degradation. In type II system, tracrRNA hybridises with the direct repeats which then gets cleaved and processed by RNase III and other nucleases. The processed crRNA-tracrRNA hybrid forms a complex with Cas9 to degrade DNA matching its guide RNA sequence. Image is taken and adapted from Hsu et al., 2014 [160].

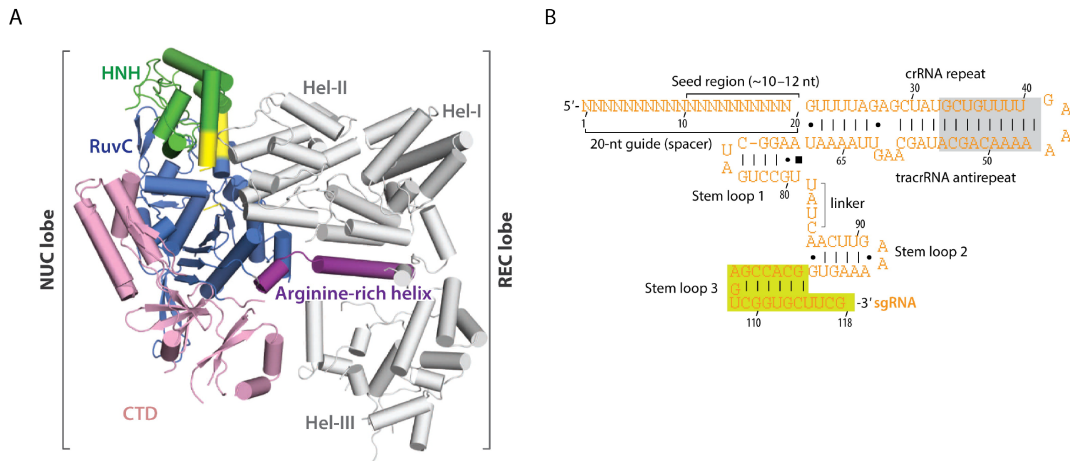
**1.1.3.3.3 CRISPR-Cas9 system as a genome-editing tool** As the CRISPR field started to attract more interests, researchers soon unraveled more details about the molecular mechanisms of the CRISPR-Cas system. Brouns et al. showed that the spacer sequences were transcribed, cleaved by the CRISPR RNA nuclease and act as guide RNAs [39]. Marraffini et al. demonstrated for the first time that the Cas protein was able to target DNA directly [255]. Moineau and colleagues showed that the *Streptococcus thermophilus* CRISPR1/Cas system cleaves plasmid and bacteriophage double-stranded DNA at specific sites within the proto-spacer sequence [129]. Subsequently, more studies were carried out and in particular, molecular mechanisms of the type II CRISPR system were extensively characterised. Charpentier and colleagues identified the trans-encoded small RNA (tracrRNA), which was required for the maturation of CRISPR RNA (crRNA) and Cas9 loading [81] (Figure 1.3 (B)). Soon afterwards, Siksnys and colleagues published detailed biochemical characterisation of Cas9-mediated cleavages. Most importantly, they demonstrated that Cas9 can be programmed to a specific target site by changing the sequence of the crRNA [130]. Like Siksnys, Doudna and Charpentier's groups also showed that the Cas9-crRNA-tracrRNA complex could cleave purified DNA *in vitro*. They also demonstrated that Cas9 could be programmed with a custom-designed crRNA to cut at a target site of their choosing. Furthermore, they showed that both crRNA and tracrRNA were required for Cas9 to function and the two RNAs could be fused as a single guide RNA (sgRNA), which has become a widely accepted concept used in genome-editing [176]. Siksnys, Doudna and Charpentier's work unleashed the potential of the universal programmable RNA-guided DNA endonuclease, and is considered profoundly important in the field of genome editing (Figure 1.2). The final highlight of the adaptation of CRISPR-Cas9 system as a genome-editing tool is the successful demonstration of targeted genome cleavage in mammalian cells by Zhang and colleagues [78]. Similar findings were published in the same issue of Science by the Church group and in Nature Biotechnology by the Joung group and the Kim group [250] [164] [71]. Since then, the CRISPR-Cas9 system has been widely adopted by the scientific community to edit genomes of a wide range of cells and organisms.



**Figure 1.2:** The mechanism of CRISPR-mediated genome engineering. To perform gene-editing, a guide RNA can be designed and constructed by fusing a crRNA containing the targeting sequence to a tracrRNA that facilitates DNA cleavage by Cas9. Binding of the PAM sequence and a matching target triggers Cas9 nuclease activity which allows it to produce a DSB 3bp upstream of the PAM site. DNA DSBs are typically repaired by NHEJ or HDR. In the error-prone NHEJ pathway, indels are introduced, frequently lead to the disruption of gene function. In the presence of a donor template, HDR pathway can be initiated to create desired mutations through homologous recombination, which allows precise gene modification such as gene knock-in and base correction. Image is taken and adapted from Jiang and Doudna, 2017 [174].



**1.1.3.3.4 Structure and working mechanism of Cas9** The Cas9 is a bilobed enzyme consists of the  $\alpha$ -helical recognition (REC) lobe and the nuclease (NUC) lobe [177] [293]. The REC lobe is essential for gRNA and DNA binding, whereas the NUC lobe is composed of the RuvC-like nuclease domain, HNH nuclease domain and a carboxyl-terminal domain (CTD) (Figure 1.3 (A)) [177] [293]. Jinek et al. and Gasiunas et al. have independently showed that the RuvC domain of the Cas9 enzyme cleaves the non-complementary strand while the HNH domain cleaves the complementary site [176] [130]. The carboxyl-terminal domain is responsible for the interaction with the protospacer adjacent motif (PAM) [293]. It was shown that the Cas9 enzyme adopts an auto-inhibited confirmation in the absence of gRNA. Upon RNA loading, the two lobes of Cas9 undergo a conformational rearrangement and form a central channel, which accommodates the RNA-DNA heteroduplex [177].



**Figure 1.3:** Overall structure of *Streptococcus pyogenes* Cas9 and gRNA. (A) Ribbon representation of the crystal structure of Cas9 protein. The nuclease (NUC) lobe contains the HNH and RuvC nuclease domains as well as the more variable C-terminal domain (CTD). The REC lobe consists of three alpha-helical domains (Hel-I, Hel-II, and Hel-III) and no structural similarity was found with other known proteins. (B) Schematic representation of the sgRNA secondary structure. Grey box denotes the extra repeat/antirepeat region in the full-length gRNA scaffold. Yellow box indicates the stem loop 3 of sgRNA, which is not essential for Cas9 function and was truncated in the sgRNA-bound structure. Image is taken and adapted from Jiang and Doudna, 2017 [174].

The PAM sequence plays an important role in target binding. While the exact PAM sequence is dependent on the bacteria or archaea species, 5' NGG 3' is the PAM site for the most widely used *Streptococcus pyogenes* Cas9 [78] [161]. Biochemical studies showed that Cas9 uses PAM recognition to identify target sites while scanning large DNA molecules [391]. Furthermore, competition assay suggested that DNA unwinding and RNA-DNA heteroduplex formation initiated at the PAM site and proceeded towards the distal 5' end

of the target sequence [391]. The crystal structure of the REC lobe also indicated that the eight PAM-proximal nucleotides in the Cas9-bound gRNA were exposed for base-pairing with target DNA, which supports the theory that the 8-12nt PAM proximal ‘seed’ region is critical for target DNA recognition [293] [176] [78] [126] [161] [311] [249].

**1.1.3.3.5 Off-target effect** Because genome-editing results in permanent alterations within the genome, the off-target effect of Cas9 nuclease is of particular concern, especially for clinical applications. A series of studies have investigated this issue using mismatched gRNA libraries, *in vitro* selection and reporter assays to monitor the ratio of off-target cleavage frequency [126] [161] [311] [80] [72] [249]. Beyond the previous knowledge that mismatches at the 3’ ‘seed’ region are less tolerated, these studies further demonstrated that the overall off-target effect depends on the number, position, and distribution of mismatches within the protospacer sequence [126] [161] [311] [249]. In addition, it was shown that the ‘NGG’ PAM sequence was not absolutely required, as the ‘NAG’ PAM can also be tolerated, albeit at a lower frequency. Therefore, it is worth considering both NGG and NAG PAM sequences in off-target analysis. Several groups have designed algorithms to select gRNAs with minimal off targets based on these findings [158] [161] [9] [61]. Furthermore, the amount of Cas9 enzyme expressed in the cell will also affect the off-target effect. It was reported that high Cas9 concentration increases the chance of off-targets [126]. To improve specificity, Cas9 was converted into a nickase by mutating either the HNH or RuvC domain. The double-nicking strategy can then be adopted using a pair of gRNAs and Cas9 nickase. Such method is based on the hypothesis that two simultaneous adjacent off-target binding and cleavage is much more unlikely than a single cleavage. It was shown that Cas9 nickase improved targeting specificity by up to 1500 times compare to wild-type Cas9 [329] [369] [249]. In addition to the usage of double nickase, shorter gRNA truncated by two or three nucleotides at the distal end could also reduce off-target activity, potentially due to greater mismatch sensitivity [127].

**1.1.3.3.6 Applications of CRISPR-Cas9 system** The CRISPR-Cas9-mediated genome editing system has been broadly used in gene function studies, disease modelling and potentially, gene therapy. In addition to DSB-based genome editing, Cas9 nucleases was engineered into RNA-guided DNA binding protein by mutating the RuvC and HNH nuclease domains [326]. This nuclease-deficient Cas9 (dCas9) can be fused with functional effector domains such as transcriptional activators, suppressors and chromatin modifiers.

The CRISPR-dCas9-based transcriptional and epigenetic regulators allow both loss-of-function and gain-of-function perturbations precisely and rapidly without major disruption of the local genomic architecture, which supplements the wild type CRISPR-Cas9 knockout function. Collectively, wild-type Cas9 and dCas9-mediated transcriptional and epigenetic regulators form a complete toolbox for comprehensive genomic study from all directions.

### **Genome editing**

The use of CRISPR-Cas9 platform has greatly accelerated the efficiency of generating cellular models as well as transgenic organisms. For generating cellular models, Cas9 and gRNAs can be easily introduced into the cell via transient plasmid transfection, viral transduction, or as ribonucleoproteins (RNPs). For generating transgenic animal models, it was shown that Cas9 protein and gRNA could be directed injected into fertilised zygotes, which bypassed the ESC targeting stage and shortened the generation time for mutant mice to only several weeks [462] [431]. In addition, the multiplexing capability of Cas9 provides a possibility for studying polygenic diseases, such as diabetes, schizophrenia and heart disease. Furthermore, many studies have reported using CRISPR-Cas9 system to correct disease-related mutations *in vitro* [465] [221] [364] [222] [457]. In 2016, the US National Institute of Health (NIH) approved the first clinical trial involving the use CRISPR-Cas9 to modify T cells from cancer patients. Although still a long way to reach the clinic, CRISPR-Cas9 technology holds prominent potential in treating inherit genetic disease, infectious diseases and cancer.

### **Transcriptional modulation**

It was shown that dCas9 alone can repress gene expression by blocking transcriptional elongation, RNA polymerase binding and transcription factor binding, which is a phenomenon referred as CRISPR interference (CRISPRi) [326]. However, the knockdown effect by dCas9 itself is not effective in mammalian cells ( 2-fold) [326]. Gilbert et al. demonstrated that dCas9 fused to KRAB domain can specifically and stably repress endogenous genes in human cells [136]. Similarly, Konermann et al. showed that dCas9-SID4X can mediate repression of Sox2 in 293 cells [204]. Although both studies demonstrated effective dCas9-mediated gene repression, the current CRISPRi needs further improvements as the knockdown effect is still partial.

Similar to CRISPRi, dCas9 can be fused to transcription activating domains, such as VP64 and p65AD to create the effect of CRISPR-mediated gene activation (CRISPRa)

[249] [204] [136] [317]. In addition to direct fusion of dCas9 with activating domains, gRNAs can be further engineered to include protein-interacting RNA aptamers for the recruitment of RNA binding proteins fused to functional effectors such as VP64 [477]. These scaffold RNAs can be used as a defined sets to generate synthetic multi-gene transcriptional programs to rewire cell fates or engineer metabolic pathways [477]. To enhance the efficiency of CRISPRa, Joung and colleagues expressed a dCas9-VP64 fusion protein and multiple gRNAs and showed dose dependent synergistic activation [243]. Similarly, Jaenisch and colleagues created the CRISPR-on system with dCas9 fused with VP48 and showed that the a cluster of 3-4 gRNAs could achieve more efficient gene induction [70]. Alternatively, other groups have developed strategies which exploit the synergy of multiple transcriptional activators. Chavez et al. generated a tripartite activator, with dCas9 fused to VP64-p65-Rta, and used to direct neuronal differentiation from iPS cells [62]. Another example is the synergistic activation mediator (SAM) system, which consists of dCas9-VP64 and gRNA-MS2, which selectively recruit MCP fused with activating domains of p65 and HSF1 [205]. Finally, Gilbert et al. and Tanenbaum et al. developed a protein scaffold called SunTag and applied to recruit multiple copies of VP64 activator modules to a single activating site [406] [135].

### **Epigenetic control**

Epigenetic modifications are crucial for controlling gene expression. Similar to transcriptional regulation, CRISPR-dCas9 can be used to recruit epigenetic modifiers and reshape the epigenome at a defined locus. Hilton et al. designed a programmable CRISPR-dCas9-based acetyltransferase by fusing dCas9 with the catalytic core of acetyltransferase p300. dCas9-p300 catalysed H3K27 acetylation at its target sites, resulting in specific transcriptional activation of genes from targeted enhancers [155]. Another study showed that the dCas9-LSD1 fusion can efficiently suppress the expression of genes controlled by the targeted enhancers [189]. Thakore et al. demonstrated that dCas9-KRAB is able to disrupt the activity of targeted enhancer via induction of H3K9me3 [411]. Furthermore, Liu et al. established a system where targeted DNA demethylation and methylation can be achieved by dCas9-Tet1 and dCas9-Dnmt3a, respectively. The dCas9 epigenetic effectors allow both loss-of-function and gain-of-function perturbations precisely and rapidly without major disruption of the local genomic architecture [229].

## 1.2 Forward genetics

In contrast to reverse genetics, forward genetics starts with isolation of mutants that show a particular phenotype and works to identify the causative genetic change. The most fundamental advantage of forward genetics is the unbiased nature of inquiry, which requires no hypothesis nor the molecular basis of the biological process to be studied, making it a powerful approach for new and unexpected discoveries [282].

Long before the post-genomic era, forward genetics was the approach in genetic research as most of the studies were based on the observations of a particular phenotype. Many genes were even named after their mutant phenotype. The phenotypes observed at the time were mostly caused by spontaneous mutations, and the subsequent identification of the causative genes usually involves positional cloning and mapping [282]. Examples include the wingless (*wg*) gene from *Drosophila* and obese (*Ob*) gene from mouse [368] [486]. Since spontaneous mutations arise at a rate insufficient to perform systematic genetic studies, accelerating the rate of mutagenesis became the pressing need.

Artificial mutagenesis can be achieved by exposing organisms to physical agents such as irradiation, chemical mutagens such as ENU, or biological mutagens such as the transposon system. These mutagens are able to introduce random mutations in a cell or organism at high efficiency, providing opportunities to perform loss-of-function genetic screens. A genetic screen is an experimental approach to identify individuals that exhibit the phenotype of interest in a mutant pool. Driven by genome sequence data, technologies such as RNA interference (RNAi) and most recently the CRISPR-Cas9 system allow scientists to generate multiplexed libraries and perform genome-wide screens in a targeted high-throughput manner, giving genetic screens the unprecedented power to discover novel genetic functions and biological pathways.

### 1.2.1 Mutagenesis using chemical and physical agents

#### 1.2.1.1 N-ethyl-N-nitrosourea (ENU)

N-ethyl-N-nitrosourea (ENU) is alkylating agent widely utilised for generating mutations in the mouse genome[340]. Given that ENU is able to induce nonsense and missense mutations, it is possible to generate null, hypomorphic and hypermorphic alleles, which diversifies the consequent phenotypes and widens its application [181]. ENU induces mutations at a relatively high efficiency. It has been shown that the mutation rate of ENU is approximately 1 in 1000 in mouse gametes and 1 in 200 in mouse ESCs[156].

Most of the ENU-induced mutations has been identified in the coding region or intron boundaries, which makes it great for studying gene functions, but not ideal for the regulatory regions [193].

### 1.2.1.2 Irradiation

Ionizing radiation (IR) including X-rays and  $\gamma$ -rays is a powerful mutagen which typically generates large deletions and complex chromosomal rearrangements such duplications, translocations and inversions. An X-ray-induced forward mutation study in Chinese hamster cells performed by Chu provided detailed analysis of the non-linear relationship between induced mutation rate and the dose of X-ray exposure [75]. It was found that X-irradiation could generate mutations at a rate of  $13 - 50 \times 10^{-5}$  per locus, which is a lower yield of mutations than chemical mutagens [339] [334].

Although ENU and irradiation are a powerful tools for genome-wide genetic screens, the difficulty in identifying the causative point mutation remains the biggest hurdle. The typical strategy for causative gene identification involves genetic mapping to narrow the region containing the mutation to a manageable size, followed by DNA sequencing. To perform genetic mapping, the affected mice need to be out-crossed to a different bred strain and the resulted F1 progeny are then backcrossed or intercrossed, which can take up to 40 weeks and involves up to hundreds of animals [193] [282]. With the development and popularisation of next-generation sequencing (NGS) technology, as well as the availability of ready-made genomic fragments in vectors such as bacterial artificial chromosomes (BACs) and the complementary DNA (cDNA) library, this process could be simplified and shortened to up to two weeks, albeit still highly labour-intensive and resource-dependent [282] [416].

### 1.2.2 Insertional mutagenesis

Insertional mutagenesis refers to the method in which an exogenous DNA integrates into the host genome and causes disruption or alteration of genes nearby the insertion site. There are mainly two categories of insertional mutagens, namely retroviral vectors and DNA transposons. Both of the vectors are flexible with the ‘cargo’ they accommodate, and can be constructed to carry various molecular elements based on the experimental design. Unlike ENU and irradiation, which have the biggest bottleneck as the identification of causative mutations, insertional mutagenesis involves integration of a piece of DNA whose identity is known, which can be used as a molecular tag to identify mutated genes.

### 1.2.2.1 Retroviral-mediated mutagenesis

Retroviral insertional mutagenesis is an experimental approach for gene discovery, taking advantage of the retrovirus life cycle: the proviral DNA integrates into the genome and results in gene expression alteration. Integration of proviruses can introduce both loss-of-function and gain-of-function mutations. The former are resulted when the provirus inserts into exogenic regions, whereas the later are produced when the enhancer element in the long terminal repeat (LTR) region of the retrovirus drives aberrant gene expression. There are several limitations to the use of retroviral-mediated mutagenesis. Most critically, retroviruses exhibit strong preferences for integration sites. It has been shown that retroviruses have both ‘hot’ and ‘cold’ insertion spots and preferentially target the 5’ end of expressed genes [276] [455]. Additionally, retroviruses carry strong enhancers in their LTRs that can deregulate genes located hundreds of kbs away, which can complicate the identification of causative gene [287].

### 1.2.2.2 DNA transposon-mediated mutagenesis

DNA transposons are mobilised in a non-replicative ‘cut and paste’ manner and has been developed as a widely used molecular tool for insertional mutagenesis [79]. Among various categories of DNA transposons, Sleeping beauty (SB) and piggyBac (PB) are more popular choices as methods of genome editing [200] [383] [171].

SB is a reconstructed DNA transposon from fossil fragments found in the salmon genome. It exclusively integrates into a TA dinucleotide, which is duplicated upon integration and flank the transposed element [171] [428]. The SB transposon can be mobilised from either an exogenous plasmid or a donor site on the chromosome, and every excision made by the SB transposase leaves a 3 bp ‘footprint’ [240]. PB is a moth-derived transposon system active in a wide range of organisms. Unlike SB, PB recognises a short TTAA sequence for insertion and excises the transposon without a footprint, which makes it a more precise and defining system [122]. Because PB can excise precisely from the donor site, it is especially useful when a transgene is transiently required, for example, in generating transgene-free iPSCs [449] [476]. One of the examples of PB-mediated genetic screen was a study performed by Rad and colleagues, in which PB was applied as a tool for genome-wide mutagenesis in mice, and many novel cancer genes were uncovered [327]. Both SB and PB demonstrate a strong tendency of local-hopping, meaning that the excised transposon preferentially integrates near its original location, which is unfavourable in conducting

genome-wide genetic screens [240] [99] [116].

### 1.2.3 The use of *Blm*-deficient and haploid cell lines in genetic screens

#### 1.2.3.1 *Blm*-deficient ESC systems

Recessive genetic screens in mammalian systems were hampered by their diploid nature. Although homozygous mutant organisms could be generated by breeding, there was no efficient approach to induce homozygous mutations in cultured cells, until the establishment of the Bloom's syndrome protein (*Blm*)-deficient ESCs [241]. Bloom syndrome is a recessive genetic disorder associated with genomic instability and cancer-prone phenotypes [140]. It has been demonstrated that *Blm*-deficient cells have increased incidence of homologous recombination which led to the increased loss of heterozygosity [241] [140]. Performing recessive screen on *Blm*-deficient genetic background increases the chance of recovering biallelic mutations. Two groups independently exploited this phenotype and conducted genetic screens in *Blm*-deficient ESCs. In one study performed by Yusa et al., a tetracycline-induced *Blm* ESC line was combined with ENU mutagenesis to successfully identify genes that are involved in the glycosylphosphatidylinositol (GPI)-anchor synthesis [475]. In the other approach, Guo et al. applied the retrovirus gene-trap system in *Blm*-deficient ESCs to select genes in the mismatch repairs pathway [144].

#### 1.2.3.2 Haploid cell lines

Another technical advance in overcoming the challenge of recessive screens in eukaryotic systems is the establishment of haploid cell lines. The first haploid human cell line (KBM7) was isolated from a chronic myeloid leukaemia patient in 1999, but it was not until a decade later its use as a tool for genetic screen was showcased [208] [52]. The mouse haploid ESC lines was generated in 2011 independently by two groups [104] [219]. Analysis revealed that these cells exhibited typical mouse ESC morphology and expressed pluripotency markers including Oct4, Klf4, Sox2, Nanog and Rex1 [104] [219]. Both of the studies demonstrated the utility of haploid ESCs for genetic screens. Leeb et al. conducted a pilot screen for mismatch repair genes in the presence of 6-TG using the gene trap PB transposon vector, and successfully identified *Msh2* and *Hprt* [219]. He and others then proceeded to a large-scale genome-wide screen with the haploid mutant library to study the exit of pluripotency and successfully recovered novel pluripotency regulators *Zfp706* and *Pum1* [217]. Elling et al. generated a haploid mutant library with retrovirus and challenged it with toxin ricin. As a result, they identified multiple genes involved in



ricin processing pathways, some of which had never been reported before [104]. These results illustrated the potential of haploid cells for large-scale forward genetic screens. However, haploid cells are unstable, and cells often undergo auto-diploidization, which requires regular selection such as cell sorting to maintain the haploid nature of the cell culture. Additionally, due to the limitation of the derivation process, genetic screens using haploid cells are only limited to a few cell types.

#### 1.2.4 RNAi-mediated screens

The first large-scale genetic screen using siRNAs in mammalian cells was performed in 2003 to study the mechanism of TRAIL-induced apoptosis. Aza-Blanc et al. transfected 510 siRNAs targeting 510 genes in HeLa cells and used AlamarBlue as a cell viability read out [7]. The screen has successfully uncovered several modulators of TRAIL-induced apoptosis including DOBI and MIRSA [7]. Zerial and colleagues performed the first genome-wide transfected siRNA screen in combination with automated imaging analysis. The authors discovered a number of kinases involved in endocytosis, suggesting that signalling functions are built into the machinery of endocytosis [313]. The Bernards group developed the first large-scale virus-based shRNA library, which contained approximately 23,000 shRNAs targeting around 8000 human genes. It was used in a pooled infection screen, from which the authors could identify one known and five new modulators contributing to the p53-dependent proliferation arrest [16]. One complication about this study was that each of the individual clone contained several shRNA inserts, which required further analysis to identify the shRNA responsible for the observed phenotype.

Compared to the methods described previously, RNAi was the first technology that supports a fully-controlled targeted screen. However, with the classical genetic approaches, one can plan for gain-of-function screens or identify mutations that are not in the coding regions, which is limited in RNAi [96]. In addition, large discrepancy were often observed between the results of a similar RNAi-based screen performed by several groups [35] [206] [488]. This is probably due to the false-positive hits resulted from the off-target effect of siRNAs. Therefore, secondary screens are often necessary to identify the true hits, and the phenotype needs to be verified by a second independent siRNA targeting the same transcript. Furthermore, siRNAs almost never completely deplete the target mRNA, which often results in false negatives [96].

## 1.2.5 CRISPR-Cas9-mediated screens

### 1.2.5.1 The establishment of CRISPR-Cas9 screening technology

Soon after the successful adaptation of CRISPR-Cas9 as a genome editing tool in mammalian cells, three groups independently generated genome-wide gRNA libraries and performed functional genetic screens with Cas9 nuclease. The Yusa lab constructed a genome-wide library with 87,897 gRNAs targeting 19,150 mouse genes and applied the resulted library in a recessive screens to identify genes that modulate susceptibility to *Clostridium speticum* alpha-toxin and 6-TG [202]. As a result, all known essential components of the GPI-anchor biosynthesis pathway has been identified, together with 13 genes whose function in alpha-toxin resistance had not been reported. Analysis of the 6-TG resistance screen revealed all known factors including four MMR genes and *Hprt*. Similarly, the Zhang group designed a genome-scale CRISPR-Cas9 knockout (GeCKO) library with 64,751 gRNAs targeting 18,080 human genes and successfully identified essential genes in cancer and pluripotent stem cell lines [366]. The authors also demonstrated the use of GeCKO library for positive selections, which uncovered both known and novel genes whose loss conferred resistance to vemurafenib in melanoma cell lines [366]. The Sabatini/Lander group built a library with 73,151 gRNAs targeting 7114 genes and 100 non-targeting gRNAs as control. With this library, they screened for genes that function in the DNA MMR pathway in the presence of 6TG using the haploid cell line KBM7, and identified genes encoding four components of the MMR pathway [434]. They also screened for resistance to etoposide in diploid cell line HL60 and revealed hits including TOP2A as well as CDK6 whose role in this pathway was previously unknown [434]. Those proof-of-principle studies demonstrated the power of CRISPR-Cas9-mediated genome-wide screens and uncovered its potential to address a wide-range of biological questions.

Compared to the other mutagenesis methods described earlier in this chapter, the CRISPR-Cas9 system allows high-throughput gene knockout in a targeted manner with pre-defined cutting sites. Importantly, the Cas9-mediated mutagenesis exhibited high bi-allelic mutation efficiency, which is essential for its application in the mammalian systems. In addition, it is straightforward to identify the causative mutations as gRNA itself can serve as a molecular barcode. Direct comparison suggested that the CRISPR-Cas9 system outperformed RNAi, which is also a reprogrammable, easy-to-perform genome-editing technology [202]. This is probably due to the fact that RNAi can almost never achieve complete suppression of gene silencing and its off-target effect often complicate the analysis and results in poor

consistency. Using a set of ‘gold standard’ essential and non-essential genes as targeting controls, Evers et al. showed that CRISPR outperformed shRNA-based system with lower noise, better consistency and lower off-target effect [108]. Similarly, Munoz et al showed that CRISPR-based screens consistently identify more lethal genes than RNAi in cancer cell lines, indicating lower false-negative rate [285].

### 1.2.5.2 Screening format

#### 1.2.5.2.1 Arrayed screening

CRISPR-Cas9-mediated screen can be carried out in two formats: arrayed or pooled format. The choice depends on the experimental aim. Arrayed screening is usually carried out in multi-well plates with each well containing one or a few gRNAs targeting a single gene. Two major advantages are associated with arrayed screens. First, the causative mutation can be easily identified as the constituents of each well are known. Second, given that each well has a single known genetic perturbation, it allows the investigation of a much wider range of phenotypes such as high-content imaging [367]. Recently, Metzakopian et al. generated the first two individually-cloned CRISPR-Cas9 genome-wide arrayed gRNA libraries with a complexity of 34,332 gRNAs for human and 40,860 gRNAs for mouse genome, covering 17,166 human and 20,430 mouse genes [272]. These libraries expanded the toolbox for comprehensive gene editing and offered an opportunity to perform screens at a single-gene level. However, arrayed screens are labour-intensive and time-consuming, as reagents have to be prepared individually. Accessibility to automated robotic equipment is often necessary for plate handling and can be its limitation on wider usage.

#### 1.2.5.2.2 Pooled screening

In contrast to arrayed screens, pooled screening is usually less expensive and do not require highly automated equipment. Although direct phenotypic assessment is limited for each gRNA and a more careful experimental design is needed, pooled screening is a powerful and fast approach for systematically investigating plenty of biological questions. The simplicity and high efficiency of the CRISPR-Cas9 technology make it an ideal system for pooled functional genomic screen. The *in-silico* designed gRNA libraries can be chemically synthesised as a pool of oligonucleotides, which is subsequently cloned into plasmid vectors, usually lentivirus expressing vectors. A mutant cell library can be generated from transduction of lentivirus library, followed by the application of selection stress. Mutations causing the phenotype of interest can be identified by next generation sequencing based

on the representation of gRNAs.

A pooled genetic screen can be designed for either positive or negative selection of gRNAs. Positive selection screening identifies genes that are enriched after applying the selection pressure. It is most commonly used to identify perturbations that confer resistance to a toxin, inhibitor, drug, or pathogen [367]. For example, in the screen aiming to identify genes involved in the GPI-anchor biosynthesis pathway, most mutants transduced with irrelevant gRNAs were depleted from the population due to the susceptibility to alpha-toxin, whereas cells transduced with gRNAs targeting genes involved in the GPI-anchor biosynthesis pathway lacked the cellular receptor, and became resistant to alpha-toxin, thereby getting enriched after selection [202]. Positive selections usually produce clearer results as the expected hits are few. In contrast, negative selection screening is to identify genes that are depleted during the selection process. The most typical negative selection screen is the one to identify essential or fitness genes, that are required for cell survival and/or proliferation. After a certain period of cell culture, mutants transduced with gRNAs targeting essential genes will deplete or ‘drop out’ from the population. Compared to positive selection screening, negative selection screening is more technically challenging and require higher screening sensitivity due to the fact that level of depletion can be limited.

### **1.2.5.3 Applications of CRISPR-Cas9-mediated screens**

The CRISPR-based screening technology has provided a great opportunity for systematic identification of essential genes. Wang et al. constructed a genome-wide gRNA library to screen for genes required for proliferation and survival in the near-haploid KBM7 cell line [433]. The screen results were validated by an orthogonal retroviral gene-trap screen and benchmarked with functional profiling in yeast. As a result, the authors were able to uncover a group of uncharacterised essential genes in various cellular pathways [433]. Hart et al. designed and generated a second-generation CRISPR knockout library referred to as The Toronto KnockOut (TKO) library and applied it to screen for essential genes in a range of cell lines derived from different parts of the human body [153]. With this approach, they could identify five-times more fitness genes than previously described in shRNA screens, and were able to classify the ‘core’ fitness genes and context-specific fitness genes, which provided insights to the biological differences between cell types [153].

CRISPR-Cas9 loss-of-function screening has also been applied to the non-coding region

of the genome. Canver et al. developed a screen using tiling gRNA library for saturated mutagenesis of non-coding elements *in situ*, which provided insights into the function and organisation of the BCL11A enhancer [50]. Similar studies have been performed by others to analyse the regulatory regions of *NF1*, *NF2* and *CUL3* loci, *POU5F1* locus and TP53 and ESCR1 transcription binding sites [350] [207] [86]. In addition, the Gersbach group designed the CRISPR-Cas9-based epigenomic regulatory element screen (CERES), which utilises the dCas9-KRAB repressor and dCas9-p300 activator to target the DNase I hypersensitive sites around genes of interest. Both loss- and gain-of-function screens were conducted for the  *$\beta$ -globin* and *HER2* loci, which revealed known and previously unidentified regulatory elements [198]. Similarly, another study performed by Fulco et al. exploited the CRISPRi library to screen for regulatory elements of *MYC* and *GATA1*, and identified nine distal enhancers that control gene expression and cellular proliferation. Not limited to the discovery of novel enhancer elements, the CRISPR-based screens were also utilised to study the function of lncRNA using cell growth as a readout [491] [228]. Zhu et al. constructed a paired-guide RNA library and uncovered 51 lncRNA that positively or negatively regulated cell growth [491]. Using a CRISPRi-based gRNA library, Liu et al. identified approximately 500 functional lncRNAs out of 17,000 screened [228]. These studies demonstrated the potential of CRISPR-Cas9 screens to unravel the functions of non-coding genome.

Although the most of the CRISPR-Cas9 screens were performed in *in vitro* systems, it has also been applied *ex vivo* in primary dendritic cells to study regulators of the bacterial lipopolysaccharide response [310]. Similarly, Chen et al. performed CRISPR-Cas9 screen in mice to study tumour growth and metastasis [65]. Recently, Manguso et al. performed an *in vivo* screen using a library encoding 9872 gRNAs targeting 2368 genes to identify genes that synergise with or cause resistance to PD-L1 checkpoint blockade [251].

The development of dCas9 opened up alternative options to conventional knockout screens. The Weissman lab constructed a genome-wide library targeting the transcription regulatory regions of approximately 16,000 genes and applied it with a dCas9-KRAB fusion protein to achieve CRISPR interference (CRISPRi) screens [135]. CRISPRi is not as efficient as CRISPR-Cas9 mediated gene knockout, but it could be a suitable option to screen for genes that are essential for cell viability. Also, because the CRISPRi-mediated knock-down effect is reversible if an inducible dCas9 is used, it allows extra on/off control to be incorporated in screening design [135]. The Weissman library can also be used for CRISPR activation (CRISPRa) screen when utilised with dCas9-VP64 with the SunTag system for

signal amplification. Another available version of CRISPRa library was designed by the Zhang lab, which contains 70,290 gRNAs targeting every coding isoform from the RefSeq database [205]. In this system, gene activation is mediated by dCas-VP64 combined with modified gRNA that recruits MS2-p64-HSF1, which is also referred as the Synergistic Activation Mediator (SAM) [205]. Both CRISPRa systems have exhibited ability to increase gene expression, however, the degree of increment depends of the targeted gene. Nonetheless, the CRISPRa libraries give researchers the opportunity to perform gain-of-function screens and study the biological pathways from a different angle.

#### 1.2.5.4 Experimental design of CRISPR-Cas9-mediated screens

##### 1.2.5.4.1 gRNA design

The outcome of a CRISPR-Cas9-based screen is directly determined by the design of the gRNA library. The key to an effective gRNA library is to maximise the overall on-target efficacy and minimise any off-target activity. Generally, the basic design process follows a set of conventional rules. First, gRNAs should be designed against the constitutive coding exons [202] [366] [434]. Second, all available gRNA candidates should be screened based on the potential off-target matches in the genome. Any gRNAs having a perfect match of its seed region elsewhere in the genome should ideally be removed from the library [202] [366] [434] [176]. Third, gRNAs with very low or high GC content, as well as homopolymer stretches should be avoided [434].

Wang et al. described some of these early rules and performed additional tests on gRNA efficacy [434]. By analysing the performance of gRNAs targeting the essential ribosomal gene sets, they found that gRNAs having pyrimidines at the last four nucleotides were disfavoured [434]. Consistently, Tzelepis et al. and Hart et al. also observed a strong bias against uridine at the last few positions of the gRNA, which is due to the premature termination of transcription by RNA polymerase III [421] [153] [456]. The Root group took an approach of a tiling library covering all possible gRNAs for a selection of cell surface markers and used flow cytometry to measure the performance of each gRNA [93] [92]. In contrast to what has been reported by Vakoc and colleagues that higher knockout efficiency can be achieved by targeting the functional domains of a protein [370], the Root lab showed that gRNAs targeting the 90% of the N-terminal protein coding sequences exhibited similar efficiency, which allowed more gRNA selection flexibility due to expanded target site window [92]. In addition to target site selection, the Yusa group demonstrated that a modification of the gRNA scaffold improved its targeting efficiency significantly,

and incorporated this change in their version 2 gRNA libraries [64][421]. Finally, the false-negative error due to lack of efficacy can be improved by increasing the number of gRNAs per gene [305]. However, a library with a larger number of gRNAs require a larger screening scale, and a balance needs to be considered by taking cost, space and number of screening conditions into consideration.

The off-target effect of gRNAs are predicted based on the position, number and nucleotide identity of potential mismatches. Although analysis in the early proof-of-principle screening studies demonstrated low off-target activity at predicted sites [202] [366] [434], the recent unbiased DBS prediction revealed unexpected potential off-target activity [125] [418]. Furthermore, a series of ChIP-seq of dCas9 coupled with various gRNAs showed a large number of off-targets binding events [211] [456]. Though such analysis is not feasible for a large-scale gRNA library, it indicates that there is a room for further improvement of gRNA specificity in library design. Potential measures include paying attention to alternative PAM sites, gRNA modification and utilisation of the double-nicking approaches. Although off-target activity is deeply concerned in clinical applications, it is unlikely to affect the performance of a genetic screen, due to the fact that any false-positive hits can be identified by comparing to the phenotype of other gRNAs targeting the same gene.

#### 1.2.5.4.2 Cas9 and gRNA delivery

Almost all the published CRISPR-Cas9-based screens to date have unanimously used lentivirus to deliver gRNA libraries. The idea has been adapted from the delivery of shRNAs libraries in RNAi-based screens. The reason of its popularity is mainly because they can stably integrate into the host genome. The virus titre at transduction needs to be carefully titrated to achieve a reasonably low multiplicity of infection (MOI) is achieved at transduction, so that the majority of transduced cells have been infected with one virus particle.

There are mainly two strategies to generate Cas9-expressing cell line: the first is to deliver Cas9 and gRNA in a single virus as demonstrated by Shalem et al. [366], and the second is to establish a stable Cas9-expressing cell line by knockin or viral transduction. Several evidence suggested that the prior-generation of Cas9-expressing cell line is advantageous because a clonal cell line with high Cas9 activity can be selected. Tzelepis et al. showed that subcloned Cas9-expressing HT29 cell line exhibited higher efficiency of mutagenesis [421]. Similar phenomena has been observed in Huh7.5, HeLa, 293T and HT1080 cell lines [490]. Such improvement in efficiency is especially important for negative selection

screens where higher screening sensitivity is required. However, it might not be applicable in primary cells, where the one-vector approach or Cas9-transgenic mouse can be used [322].

#### 1.2.5.4.3 Gene identification and data analysis

At the end of a pooled CRISPR-Cas9 screen, cell pellets are collected from both treated and control samples, and genomic DNA is extracted from them. The lentiviral integrant, which contains gRNA sequence, will be amplified by PCR and analysed by Illumina sequencing. The necessary sequencing depth is largely dependent on the design of the screen and resulted final cell number to be sequenced. In the case of a positive selection screen, often only a small number of cells are collected at the end of the selection process, hence only a few million reads will be enough. In contrast, negative selection screens, where the change of gRNA representation can be subtle and cell population at the end of the screen is usually large, require much deeper sequencing depth. Typically the Illumina HiSeq platform is used with 30-40 million reads for a population up to 100 million cells.

Following sequencing, statistical analysis needs to be performed to determine the significance of any changes between control and experimental samples at the gRNA-level, as well as gene-level. A range of algorithms designed for differential RNA-Seq expression analysis or shRNA screens were employed to analyse CRISPR-Cas9 screening data. Shalem et al. used the RNAi Gene Enrichment Ranking (RIGER) algorithm, which examined the positions of the gRNAs targeting the same gene and assessed whether the set was skewed towards the top of the list based on Kolmogorov-Smirnov statistic. An enrichment score was calculated based on this algorithm followed by a permutation test [239] [366]. The Wei group adopted an R package called DESeq2 to perform the statistical analysis of gRNA abundance, where gRNAs were ranked by the average fold changes [490]. Although algorithms for differential expression analysis such as DESeq2, edgeR and baySeq can be used to evaluate the statistical significance of hits in the CRISPR/Cas9-based screens, they can only perform the analysis at the gRNA level. Algorithms for shRNA screens such as RIGER and Redundant siRNA Activity (RSA) are also not ideal [225]. Followed the need of a computational algorithm suitable for CRISPR/Cas9-mediated screens, the Liu group developed an algorithm called Model-based Analysis of Genome-wide CRISPR/Cas9 Knockout (MAGeCK) [225].

In MAGeCK, read counts from all the conditions were normalised by the median ratio method, followed by the use of a negative binomial (NB) model to estimate the statistical



significance between the treatment and control samples [225]. The gRNAs are ranked based on  $P$ -value calculated from the NB model and a modified robust ranking aggregation ( $\alpha$ -RRA) algorithm is used to evaluate the statistical significance at the gene level [225]. Specifically, the  $\alpha$ -RRA ranks a gene by comparing the skew of its gRNA sets to the uniform null model and calculates the false discovery rate (FDR) from an empirical permutation test using the Benjamini-Hochberg procedure [225]. The authors demonstrated more robust and sensitive CRISPR-Cas9 screening data analysis using MAGeCK compared to the other existing algorithms [225]. The reason is probably because MAGeCK has considered the fact that not all gRNAs targeting the same gene are working equally well and the  $\alpha$ -RRA algorithm can remove the gRNAs with low targeting efficiency, which is more likely to produce more accurate gene level analysis.

## 1.3 Embryonic stem cells (ESCs)

### 1.3.1 Early development of mouse embryo

The mouse embryonic development begins in a fertilised egg packed within the protective glycoprotein layer called zona pellucida. The fertilised egg, also known as zygote, is capable of generating all embryonic and extraembryonic lineages. Such ability is defined as totipotency. The first five cell cycles of the embryonic development, which is referred to as the cleavage divisions, occurs without increase in total cell mass. Cells generated from cleavage divisions are called blastomeres. Blastomeres retain both embryonic and extraembryonic potential until the late eight-cell stage, when cell polarity is established and compaction takes place to form morula. During compaction, the spacial location of each individual cell is instructive for their subsequent lineage differentiation. Cells located on the outside develop into the first extraembryonic lineage called trophoblast, which is essential for implantation and will subsequently differentiate into placenta. In contrast, the inner cells of a morula are biased toward forming the inner cell mass (ICM).

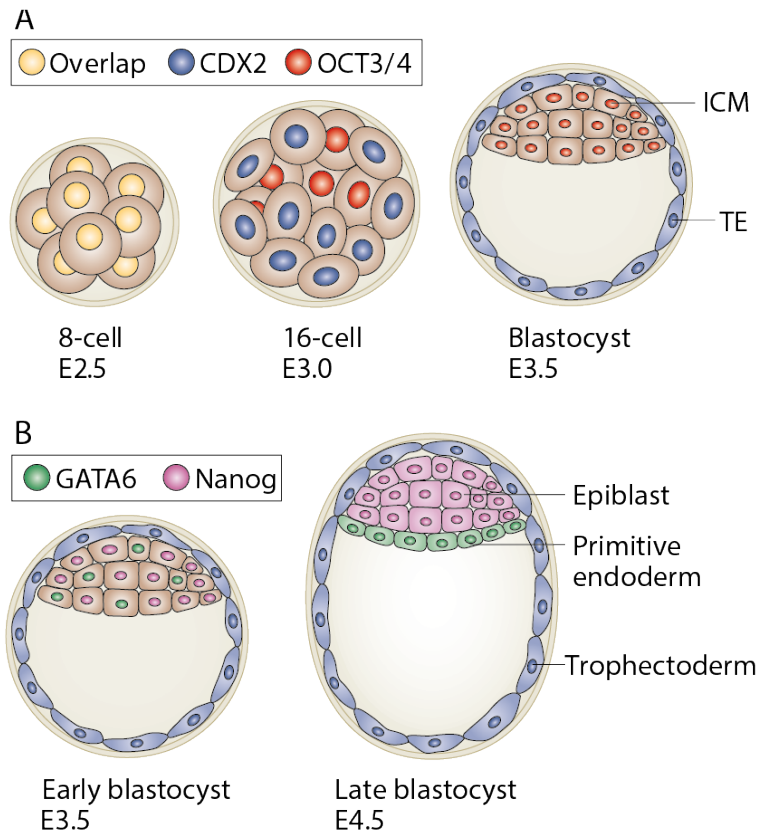
Specification of the trophoblast lineage appears to be mediated primarily by the Hippo signalling pathway, which conveys positional information into lineage-specific gene expression. In an early embryo, the Hippo pathway is active in the inner cells, where Yap1 is phosphorylated by Lats1/2 and degraded. As a consequence, Yap1 is unavailable to act as a co-activator for the key trophoblast transcription factor Tead4, resulting in failure to activate the trophoblast programme. On the contrary, in the outer cells, where Lats1/2 is inactive, Yap1 pairs with Tead4 and upregulates Cdx2, Gata3 and eomesodermin, which collectively drive commitment to the trophoblast lineage. In line with this model, Lorthongpanich et al. showed that knockdown of LATS kinases by injecting siRNA into mouse zygotes caused lineage misspecification and resulted in the generation of a TE-like lineage in the morula [236]. Once upregulated, the expression of Cdx2 and Eomes is maintained by Elf5 through a positive feedback loop to reinforce the commitment to the trophoblast lineage [288]. Despite both being important for trophoblast (TE) specification, Cdx2 and Eomes seem to play different roles. It was shown that Cdx2-deficient blastocysts failed to repress Oct4 and Nanog in the outer cells, which led to the failure of the segregation of ICM and TE, whereas Eomes mutant blastocysts could implant and showed normal Cdx2 and Oct4 expression [394] [430]. These observations suggested that Cdx2 is the earlier TE inducer in morula and Eomes is required for further TE differentiation at the blastocyst stage.

Coincident with the specification of the TE, the establishment of ICM is under the influence of the upregulated Oct4. In the absence of Oct4, inner cells fail to develop into mature ICM but rather divert into TE [289]. Oct4 acts cooperatively with Sox2 to regulate the expression of several pluripotent genes, including *Fgf4* and *Nanog* [13] [33]. It was shown that *Sox2*-null embryo was able to develop normal ICM, but fail to maintain an epiblast or further differentiation [6].

After the segregation of trophoblast and ICM, the trophoblast pumps fluid into the blastocyst to form a cavity known as the blastocoel. At this stage, the ICM start being partitioned into the epiblast and primitive endoderm as a consequence of differential gene expression. This specification is first observed when *Nanog* and *Gata6* begin to express in a mutually exclusive manner. Cells expressing primitive endoderm markers such as *Gata6*, *Gata4*, *Sox17* and *Pdgfr* gradually move away from *Nanog*-positive cells, and eventually form a morphologically distinguishable epithelium layer adjacent to the cavity. This process is regulated by the FGF signalling pathway as embryos deficient of *Grb2*, *Fgf4* or *Fgfr2* fail to form the primitive endoderm [63] [69][291]. At the same time, *Nanog*-expressing cells remain restricted to the inner space between the trophectoderm and primitive endoderm and develop into the pluripotent epiblast. These pluripotent cells are thought to be in the ‘ground state’, which is characterised by their unrestricted differentiation capacity and flexibility to the formation of all embryonic lineages [258].

### 1.3.2 Derivation of mouse ESCs

ESCs are characterised by its ability to sustain self-renewal and remain as undifferentiated for an extended period of time in culture. When injected into adult mice, ESCs give rise to multi-differentiated teratocarcinoma. Their full differentiation potential was revealed by blastocysts injection, which yields chimeric mice with high contribution from the injected ESCs to all tissues, including functional colonisation of the germ line [34]. Competence of germline transmission suggests that ESCs can be exploited as a vehicle for introducing genetic modifications into mice [335]. The fact that ESCs are permissive to multiple rounds of sophisticated genetic manipulation and their ability of clonal expansion enables isolation of mutants with desired genetic modification. These groundbreaking discoveries led to the creation of transgenic mice, which became an immensely powerful technology for basic research and the development of new therapies.



**Figure 1.4:** Lineage segregation in mouse blastocyst. (A) At E2.5, the eight blastomeres retain both embryonic and extraembryonic potential, which is reflected by the overlapping expression of *Cdx2* and *Oct4*. At the blastocyst stage, *Oct4* is expressed exclusively in the inner blastomeres, which leads to the formation of inner cell mass (ICM). As the early trophoctoderm (TE) inducer, *Cdx2* is exclusively expressed in the outer blastomeres. (B) At E3.5, the ICM shows mosaic 'salt and pepper' expression of *Nanog* and *GATA6*. *GATA6*-positive cells are subsequently sorted to the distal surface of the ICM, where they give rise to the primitive endoderm. *Nanog*-positive cells exclusively give rise to the pluripotent epiblast. Image taken and adapted from Arnold and Robertson, 2009 [4].

It was a challenge to derive ESCs directly from the embryo until Martin and Evans independently succeeded in isolation and maintenance of pluripotent cell lines [107] [260]. The original derivation of mouse ESCs involved explanting blastocysts or isolated ICMs on a layer of mitotically inactivated fibroblasts called 'feeders', in medium containing fetal calf serum. This method was developed empirically from the early research on embryonal carcinoma (EC) cells. Smith et al. later demonstrated the ability of leukaemia inhibitory factor (LIF) in replacing feeder cells both in derivation and long-term culture. The derivation process in serum-containing culture is inefficient as the emergence of ESCs only happens after dissociation and replating of the primary outgrowth. With the advent of 2 inhibitors (2i), namely GSK3 inhibitor CHIR99021 and MEK inhibitor PD032590, the derivation process became more efficient [290].

It has been well-established that mouse ESCs are originated from the mid-blastocyst-stage at embryonic day E4.5. However, it has been shown that ESCs can also be derived from early-blastocyst-stage at E3.5 or even from eight-cell-stage blastomeres, suggesting that ESCs may represent a very early developmental stage. Attempts to derive pluripotent cell lines from implanted mouse embryos had not been successful for a long time until a different pluripotent cell type was established from the postimplantation-epiblast using a different culture condition [38] [410]. These cells are referred to as epiblast stem cells (EpiSCs). Unlike ESCs, EpiSCs do not rely on LIF or 2i to maintain pluripotency, instead require FGF and activin. EpiSCs exhibit some pluripotency features such as expression of Oct4, the ability to differentiate *in vitro* and form teratocarcinomas, but they cannot contribute effectively to blastocyst chimeras. Gene expression analysis revealed that EpiSCs exhibit relative low expression of ICM-specific genes such as *Rex1*, *Sox2* and *Nanog*, but upregulation of late epiblast markers such as *Fgf5*, *Brachyury* and *Sox17* [410] [38]. These evidence indicated that EpiSCs represent a more advanced state of pluripotency, which is called primed pluripotency

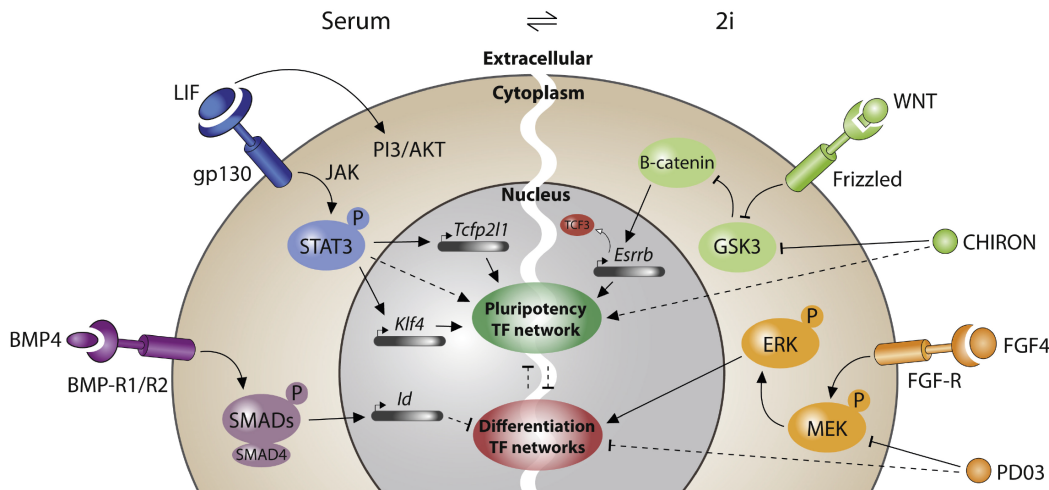
In addition to ESCs and EpiSCs, which have been derived from the epiblasts of the blastocyst, other stem cell lines have been established from other lineages of the early embryo. Examples include embryonic germ (EG) cells which can be derived from primordial germ cells (PGCs) in embryos between E8.5 and E11.5, permanent trophoblast stem cell lines from early post-implantation trophoblasts and extra-embryonic endoderm (XEN) cell lines from the primitive endoderm lineage [265] [331] [405] [209]. The establishment of these cell lines has provided powerful models for the dissection of the molecular mechanism underlying lineage specification in early embryonic development.

### 1.3.3 Regulation of the pluripotency state

Pluripotency is defined as a capacity of a cell to give rise to all the specialised cell types of an adult organism. The derivation of ESCs made it possible to capture pluripotency indefinitely *in vitro*, and provided an extraordinary tool to investigate the molecular mechanisms that govern pluripotency. Accumulating evidence suggested that the ESC identity is sustained through integrated actions of multiple extrinsic signalling pathways with intrinsic transcription regulatory network, reinforced by epigenetic modifiers.

### 1.3.3.1 Extrinsic signalling pathways

The ability of ESCs to retain pluripotency is stabilised by the continuous input of extrinsic cues. Such requirement is owing to the auto-inductive stimulus, in particular FGF, which promotes the exit from pluripotency. Multiple extrinsic factors need to be fed into the system to counterbalance the self-inductive differentiation signals and reinforce the pluripotent network (Figure 5.1).



**Figure 1.5:** Extrinsic signalling pathways that regulate pluripotency. Filled arrows represent activation of target activity and bars indicate inhibition. Solid lines indicate a direct or known downstream target, whereas a dashed line shows an indirect or inferred effect. BMP4 signalling functions via phosphorylating Smads to activate *Id* genes. LIF signalling affects many pathways but its positive effect on pluripotency is primarily via JAK-mediated phosphorylation of STAT3, which activates *Tcfp2l1* and *Klf4*. Canonical WNT signalling inhibits GSK3 activity leading to stabilisation of  $\beta$ -catenin, which in turn abrogates TCF3-mediated repression of pluripotency genes including *Esrrb*. CHIRON closely mimics WNT signaling by inhibiting GSK3. FGF signalling activates the MAPK pathway which promotes the transition to a primed state. Two inhibitors (2i), CHIRON and PD03, stabilise naive pluripotency through inhibiting GSK3 and MAPK pathway respectively. Image taken and adapted from Hackett and Surani, 2014 [147]

#### 1.3.3.1.1 LIF-mediated signalling pathway

The discovery of LIF was driven by the urge to demystify the function of feeder cells. Through screening of a range of feeder cell types, medium collected from Buffalo rat liver cells was found to effectively suppress differentiation and sustain propagation of ESCs in the absence of feeders [382]. Further analysis of this medium led to the identification of the effective factor, LIF [381] [445]. LIF is able to replace feeders in both derivation and long-term culture of mouse ESCs. It is likely that feeder cells provide extra support other than LIF as ESCs cultured in serum and LIF in the absence of feeders demonstrate slightly larger heterogeneity and more differentiated cells. However, LIF is the major

functional component of feeders given that *LIF* knockout fibroblasts were unable to sustain undifferentiated ESCs growth [392].

LIF signals through binding to a heterodimeric receptor consists of gp130 and Lifr [131]. Both gp130 and Lifr are constitutively associated with tyrosine kinases from the JAK family [444]. These kinases become activated upon binding of LIF. There are four members of JAK family, among them, Jak1 was demonstrated to be the primary downstream effector of Lifr [388] [105]. Activated Jak1 initiates a cascade of tyrosine phosphorylation that stimulates three distinct signalling pathways: the JAK/STAT, PI(3)-kinase, MAPK pathways [388] [413] [302]. These pathways contribute to self-renewal, survival as well as differentiation [292].

The Stat proteins are a family of transcription factors, among which Stat3 is the crucial functional effector following stimulation by LIF. Stat3 is activated via phosphorylation by Jak1, which allows it to form a signalling competent dimer and translocates into the nucleus. It was shown that Stat3 is required for ESC maintenance. Over-expression of a dominant-negative Stat3 construct in ESCs led to abrogated self-renewal and differentiation, whereas over-expression of Stat3 is sufficient to sustain LIF-independent self-renewal [295] [264]. Molecular studies revealed that Stat3 promotes pluripotency via upregulating pluripotency genes such as *Klf4*, *c-Myc*, *Gbx2* and *Tfcp2l1* [148] [53] [297] [399] [466] [257]. Additionally, ChIP-Seq analysis revealed that STAT3 shares many common target site with Nanog, Oct4 and Sox2, indicating that JAK/STAT3 pathway directly feeds into the core pluripotency transcription network [68].

Activation of PI(3)-kinase pathway following LIF stimulation is driven by the association between Jak1 and the p85 subunit of PI(3)-kinase. It has been demonstrated that inhibition of PI(3)-kinase pathway in ESCs resulted in spontaneous differentiation, even in the presence of LIF. Furthermore, ESCs expressing an active form of Akt could be maintained in an undifferentiated state without LIF, indicating that constitutively active PI(3)-kinase signalling is sufficient for ESC maintenance [439].

Interestingly, in parallel to JAK/STAT and PI(3)-kinase pathway, LIF activates MAPK pathways via recruiting Shp2, which induces stimulation towards differentiation [47] [361]. It was demonstrated that *Socs3* null ESCs, which showed hyperactive MAPK pathway as a result of non-competition of binding to the receptors, were inclined to differentiate into primitive endoderm in the presence of LIF [401]. And this phenotype can be reversed by inhibition of Mek [118]. These observations suggests LIF induces competing pathways

and the self-renewal signal downstream of LIF is under a fine balance between positive and negative pluripotency regulatory pathways.

#### 1.3.3.1.2 TGF- $\beta$ -mediated pathway

The TGF- $\beta$  family comprises a broad range of proteins including TGF- $\beta$ , nodal, activins, Growth Differentiation Factors (GDFs), and Bone Morphogenetic Proteins (BMPs) [261]. Signalling mediated by TGF- $\beta$  ligands is transduced through two types of transmembrane kinases, the type I and type II receptors. In the canonical TGF- $\beta$  pathway, upon ligand binding, type II receptors phosphorylate type I receptors, which in turn phosphorylate and activate the downstream effector Smads. Smad1, Smad5 and Smad8 are activated by BMP receptors, whereas Smad2 and Smad3 are activated by TGF- $\beta$ /nodal/activins receptors. These receptor-activated Smads form trimers with the common Smad, Smad4, and translocate into the nucleus, where they interact with other transcription factors, co-activators or co-repressors to regulate the expression of target genes [343].

Bmp4 signalling plays an important role in mouse ESC maintenance through upregulation of the Id proteins, which are transcription factors inhibit neuronal differentiation [469]. It was shown that Bmp4 can substitute serum in ESC maintenance in the presence of LIF, indicating that Bmp4 might be the functional component of serum [469]. However, Bmp4 is known to promote mesoderm, endoderm and trophoblast differentiation [286] [315] [342] [424] [460], and LIF can block mesoderm and endoderm differentiation but not neural differentiation [470]. These evidence led to the proposal of a mechanism that LIF and Bmp4 act cooperatively in supporting pluripotency status by each suppressing differentiation towards specific fates. Because of this counterbalancing effect, ESCs in serum/LIF condition exist in an unstable environment with competing signals, which is reflected by the pluripotency heterogeneity.

The TGF- $\beta$ /activin signalling comprises another branch of canonical TGF- $\beta$  pathway, which was shown to be essential for pluripotency maintenance in human ESCs and mouse EpiSCs via Smad2/3-mediated Nanog activation [423]. However, activation of TGF- $\beta$ /activin signalling induces differentiation of mouse ESCs in the absence of LIF.

#### 1.3.3.1.3 FGF/MAPK pathway

Fgf signalling is activated via binding of Fgf to the receptor tyrosine kinase Fgfr, which leads to the formation of a complex between Fgfr, Frs2 $\alpha$ , Shp2, and Grb2. The complex formation facilitates the activation of the phosphorylation cascade through Ras-Raf-Mek-



Erk. Originally considered as an autocrine self-renewal signal, it was later confirmed that Fgf4 acts to stimulate ESCs towards lineage specification. Evidence includes that *Fgf*-deficient ESCs were severely compromised in neural and mesodermal differentiation [443]. Although formation of the Fgfr-Frs2 $\alpha$ -Grb2 can also activate the PI(3)-kinase pathway via Gab1, Mek1/2 was identified as the downstream effector of the Fgf4 signal based on the fact that the phenotype of *Fgf4*-null ESCs can be reproduced using Mek1/2 inhibitor [214] [47] [390]. Furthermore, Kunath et al. demonstrated that *Erk2*-null ESCs fail to commit differentiation and retain expression of pluripotency markers Oct4, Nanog and Rex1 [210]. Consistent with the observations *in vitro*, Mek inhibitor treated embryos failed to form blastocysts and generated enlarged epiblast [291]. Similar phenotype can be observed in *Grb2*-deficient embryos [63].

These discoveries led to the hypothesis that blocking Fgf/Mapk pathway facilitates maintenance of pluripotency. As predicted, it was shown that either of the Fgfr inhibitor SU5402 or MEK1/2 inhibitor PD184352 could replace the requirement for serum/BMP and support long-term ESC maintenance [471]. However, inhibition of FGF/ERK pathway is not sufficient to maintain ESC self-renewal without LIF.

Despite the widely used Mek inhibitor in ESC maintenance, its molecular mechanism remained unclear, until a study performed by Yeo et al. suggested that ERK2 drives differentiation through phosphorylation and destabilisation of Klf2. It was demonstrated that over-expression of Klf2 can replace Mek inhibition which allows stable culture under Gsk3 inhibition alone [467]. In addition, Tee et al. showed that Erk2 directly modulates chromatin features required for developmental gene expression via regulating PRC2 and RNAPII [409]. Notably, ESCs express Fgf4 under the regulation of Oct4 and Sox2. This indicates that the transcription factors essential for the establishment and maintenance of pluripotency also function as differentiation promoters [473].

#### 1.3.3.1.4 Wnt signalling pathway

In the absence of Wnt, Apc, Axin and GSK3 form a complex that phosphorylates  $\beta$ -catenin in coordination with Ck1 $\alpha$ , which marks it for ubiquitination and proteolysis [76]. In the presence of Wnt, Frizzled receptor forms a complex with Lrp5/6, which triggers the displacement of GSK3 from the destruction complex, allowing  $\beta$ -catenin to accumulate and translocate into the nucleus where it interacts with co-activators to drive transcription of target genes [451] [454].

A positive effect of Wnt signalling in promoting self-renewal was demonstrated by two studies focusing on the knockout phenotype of *Apc* and *Gsk3* in ESCs. Kielman et al. showed that constitutive activation of Wnt signalling via *Apc* mutation affected the differentiation potential of ESCs both *in vitro* and in teratomas [192]. Doble et al. generated the *Gsk3* DKO cell line, in which both Gsk3 $\alpha$  and Gsk3 $\beta$  were inactivated. The DKO cell line, which has elevated  $\beta$ -catenin, demonstrated severe defects in differentiation [91]. Notably, in both studies, the severity of the phenotype exhibited a dose-dependent manner which correlated to the *Apc* mutation or *Gsk3* functional alleles. Furthermore, Sato et al. showed that the addition of a GSK3 inhibitor BIO could facilitate maintenance of ESC and resulted in sustained expression of Oct4, Rex1 and Nanog [352]. Ogawa et al. demonstrated that supplementation of Wnt3a helped to maintain ESC self-renewal in the presence of LIF [301]. These evidence converged to the deduction that elevated  $\beta$ -catenin promotes ESC self-renewal and results in differentiation defects, which was confirmed by Wray et al. by showing that the absence of  $\beta$ -catenin eliminated the self-renewal response to Gsk3 inhibition [450]. They also showed that the responsiveness could be restored by truncated  $\beta$ -catenin lacking a transactivation domain [450]. This indicates that the transcriptional activation function is not required for  $\beta$ -catenin to confer differentiation resistance. Instead, it was found that the role of  $\beta$ -catenin in pluripotency arises through direct interaction with the transcription repressor Tcf3 (gene name *Tcf7l1*) [450]. Chip-Seq data revealed that Tcf3 shares binding site with Oct4 and Sox2 [109], and acts a repressor to antagonise their function [269]. Other key pluripotency factors repressed by Tcf3 include *Esrrb*, *Klf2* and *Nanog* [77] [316] [142] [450]. Interaction between  $\beta$ -catenin and Tcf3 abrogate its repressive effect on pluripotent genes and stabilises pluripotency programme [374] [453]. It has been shown that *Tcf7l1*-null ESCs exhibit enhanced self-renewal and differentiation defects [468] [142]. *Tcf7l1*-null embryos develop normally until profound defects was observed in axial patterning during implantation, which highlighted the prominent role of Tcf3 as a regulator for differentiation [270].

It was shown that the stimulation Wnt signalling facilitates ESC maintenance; however, activation of Wnt signalling alone is insufficient to maintain long-term ESC self-renewal [471] [301]. Remarkably, the combination of GSK3 inhibitor (CHIR9902) with the Fgfr inhibitor (SU5402) and Mek1/2 inhibitor (PD184352) could effectively maintain ESCs for an extended period of time even in the absence of LIF [471]. This system was referred to as ‘3i’, which evolved to ‘2i’ with the substitution of the SU5402 and PD184352 to a more potent and specific Mek inhibitor PD0325901.

### 1.3.3.1.5 Serum/LIF culture and 2i culture

Conventional condition (serum/LIF) is chemically undefined and often activate multiple conflicting pathways. As a result, it promotes a considerable degree of morphological, transcriptional and functional heterogeneity among cells [147]. This heterogeneity is reflected on the expression of a range of pluripotency-associated transcription factors, such as *Nanog*, *Rex1*, *Esrrb*, *Stella*, *Klf4*, *Tbx3* and *Hex* [57] [417] [154] [426] [49] [297]. Functional distinction was observed between cells with different expression levels of some of these factors. For example, cells with low *Nanog* expression exhibit moderate expression of primitive endoderm markers such as *Gata4* and *Hex1*, and epiblast marker *Fgf5* [49] [377] [186]. Similarly, *Rex1*-low cells were shown to have poor ability to form chimeras following blastocyst injection [417]. These observations indicate the existence of two distinct sub-populations in ESCs cultured in serum/LIF: naive pluripotent cells, and primed cells. The later is associated with expression of lineage markers and poor performance in pluripotency assays. Notably, purification of the primed cells by cell sorting and replating assay showed that these two subpopulations are interchangeable, and the transcriptional and functional differences exist in a dynamic equilibrium [451] [57]. Overall it suggested that ESCs in serum/LIF condition is maintained in a metastable naive pluripotent state, with a small proportion of cells cycling in and out of the ‘primed’, pluripotent state [147].

The development of 2i condition allowed maintenance of ESCs in stabilised naive state, characterised by its relatively spherical colony morphology with defined borders and lack of differentiating cells. 2i-cultured ESCs exhibit a homogeneous transcriptional and epigenetic state with uniform expression of *Nanog* and *Rex1* [471]. Transcriptome and epigenomic analysis showed that the 2i-cultured ES cells exhibit a profile comparable to that of E4.5 epiblast, which probably explains its higher chimera contribution [30][203]. This more robust naive pluripotency status in 2i condition is probably owing to the complete insulation of differentiation signals. It was thus proposed that ESCs in 2i represent the *in vitro* ‘ground state’, meaning a homogenous population with the potential to form all embryonic lineages unbiasedly [471]. The *in vitro* ground state is the most optimised state of naive pluripotency to date and the closest model to the pre-implantation epiblast [147]. However, a recent study reported that prolonged MEK1/2 suppression resulted irreversible epigenetic changes that compromise the developmental potential of ESCs [73].

### 1.3.3.2 Transcription factor network

#### 1.3.3.2.1 Core pluripotency factors

##### Oct4

Oct4, also known as Oct3, is a member of the POU transcription factor family encoded by the *Pou5f1* gene. Oct4 regulates gene expression by binding to the octamer motif ATGCAAAT within the promoter or enhancer region and was the first factor identified as a master transcription factor in pluripotency and lineage specification regulation [363] [220] [303].

Oct4 is absolutely essential for embryogenesis as *Oct4*-deficient embryos failed to develop ICM and die at the time of implantation [289]. The detection of Oct4 was made as early as the zygote stage, which is believed to be inherited from the oocyte [479]. Zygotic Oct4 expression can be detected at 4- or 8- cell stage in blastomeres until blastocyst formation. After the first lineage specification takes place, cells in the ICM retain Oct4 expression whereas cells in the trophoblast have little or no Oct4 expression [307] [318]. Upon implantation, transient up-regulation of Oct4 induces the formation of primitive endoderm, while in the epiblast, Oct4 expression remains uniformly and continuously high [307] [318]. During gastrulation, Oct4 expression is down-regulated and eventually confined to primordial germ cells. In cell culture systems, Oct4 is highly expressed in ESCs, ECs, and embryonic germ cells. Its expression is down-regulated upon induction of differentiation [307] [190] [318].

The critical role of Oct4 in pluripotency maintenance was uncovered by Niwa et al. For this, an inducible Oct4-expression system was established wherein Oct4 level can be modulated by the addition of tetracycline. Using this system, Niwa et al. demonstrated that a two-fold increase in Oct4 expression led to primitive endoderm differentiation, whereas repression of Oct4 caused dedifferentiation to trophectoderm [296]. Therefore, Oct4 expression needs to be tightly regulated in ESCs. It was shown that the Oct4 positive regulators include *Esrrb* and *Sall4*, whereas *Tcf3*, *Gcnf* and *Cdk2* mediate its negative regulation [485] [483] [353] [84] [404] [281]. In addition to its crucial function in the maintenance of pluripotency, Oct4 plays a role in regulating early cell fate. As mentioned previously in this Chapter that expression of the autocrine differentiation signal *Fgf4* was under the regulation of Oct4. It was shown that Oct4 formed a complex with *Cdx2*, which resulted in a reciprocal inhibition mechanism with mutually exclusive expression and facilitated the segregation of pluripotent stem cells and trophectoderm [298]. Similarly, others have

shown that sustained Oct4 expression induced specific lineage commitment in dependent on the condition. For instance, Shimosaki et al. reported that Oct4 upregulation in ESCs accelerated neurogenesis under serum-free culture condition [372]. Additionally, transient increase in Oct4 expression led to cardiac commitment [480]. Finally, Oct4 was found to play an important role in the reprogramming of somatic cells into induced pluripotent stem cells (iPSCs). As one of the groundbreaking works in the stem cell field, Takahashi and Yamanaka screened 24 factors and found that four transcription factors Oct4, Sox2, Klf4 and c-Myc were sufficient to reprogram fibroblast cells to pluripotent cells [400].

Among all pluripotency regulators, Oct4 was found to be central to the machinery. Most importantly, Oct4 acts as a fundamental coordinator that recruits factors with various functions to establish gene regulation programmes. Several mass spectrometry studies were performed and identified a large number of Oct4 interaction partners from families such as transcription factors, epigenetic modifiers, transcriptional coactivators and components of signalling pathways [425] [106] [89] [309]. In particular several chromatin remodelling complexes such as NuRD, SWI/SNF and LSD1 were found to interact with Oct4. The correlation of LSD1 and Oct4 was confirmed by Whyte et al., showing that LSD1 mediated pluripotency-related gene silencing during differentiation and this function was through recruitment by Oct4 [442]. Although many of the other proposed correlations found in these mass spectrometry analysis need to be validated, they demonstrated the prominent role of Oct4 in pluripotency regulatory network.

## **Sox2**

The most well-known partner of Oct4 is Sox2, which is also considered as one of the core pluripotency factors. Sox2 belongs to the Sry high mobility group (HMG) box (Sox) superfamily, which interact with DNA via the HMG domain with a consensus sequence. Like Oct4, Sox2 is also required for early embryogenesis. Homozygous *Sox2* mutant embryos die shortly after implantation, due to failure of the epiblast formation [6]. Sox2 is highly expressed in mouse ESCs. *Sox2*-null ES cells differentiated primarily to trophoectoderm-like cells, similar to Oct4-null ES cells [262]. Notably, forced expression of Oct4 in *Sox2*-deficient ESCs could rescue their the phenotypes, suggesting that the role of Sox2 in pluripotency maintenance is to sustain Oct4 expression [262]. Masui et al. also showed that Sox2 positively regulated the expression of Oct4 by promoting the expression of Oct4 positive regulators such as Nr5a2 and repressing Oct4 negative regulators such as Nr2f2 [262]. Oct4 and Sox2 bind DNA cooperatively and act synergistically on many pluripotency-related genes [1] [2]. Compared to Oct4, Sox2 is more widely expressed in

the developing embryo, from epiblast to trophoctoderm as well as later in the neuroectoderm [6] [191]. Sox2 has been reported to be in charge of neural differentiation by repressing other lineage regulators such as brachyury [436] [487].

### **Nanog**

Another key regulator that contributes to the core pluripotency circuit is Nanog. Nanog is a homeodomain-containing transcription factor whose role in pluripotency was first described by two groups independently in 2003. Chambers et al. discovered Nanog from a functional screen using an ESC cDNA library and found that forced expression of Nanog from transgene is sufficient to maintain ESC pluripotency with elevated Oct4 level independent of LIF [56]. In the same issue, Mitsui et al. reported the identification of Nanog by digital differential display comparing the expressed sequence tag libraries from ESCs and somatic tissues [277].

Deletion of Nanog results in preimplantation lethality, indicating its indispensability in early mouse embryo development. *Nanog*-null embryos failed to develop epiblast, instead cells either committed to trophoblast differentiation or progress to apoptosis [375] [277] [289]. The expression of Nanog in ICM follows a ‘salt-and-pepper’ pattern mutually exclusive with Gata6-expressing cells, which was shown to be essential for the formation of primitive endoderm and epiblast via potentiating Gata6 expression and providing support from a functional epiblast [375]. By the late blastocyst stage, Nanog expression become restricted to epiblast compartment, where it is uniformly expressed [375]. It was observed that Nanog expression in conventional ESC culture is heterogeneous. To investigate this phenomena, Chambers et al. generated heterozygous and null Nanog cell lines [57]. A reduction of self-renewal ability was observed in relation to the dosage of Nanog [57]. Surprisingly, it was observed that *Nanog*-null cells maintained the ability for self-renewal, albeit prone to differentiation [57]. Subsequent studies showed that Nanog-null cells were able to contribute to three germ layers. These findings suggested that Nanog mainly function in stabilising pluripotency by counteracting alternative gene expression states [57].

Like Oct4 and Sox2, Nanog has been shown to interact with a large number of protein partners ranging from transcription factors, chromatin modifiers and signalling pathway components, indicating its critical role as a core pluripotency regulator. Unlike Sox2 which is closely associated with Oct4, global mapping of Oct4 and Nanog binding sites showed only partial co-occupancy of Nanog and Oct4. In addition, transcriptome analysis after

shRNA knockdown of either Oct4 and Nanog showed distinct gene expression profile [232] [294]. It was thus proposed that Nanog has a complementary and partially overlapping gene regulation activities to Oct4/Sox2 [258].

### **The Oct4/Sox2/Nanog triumvrate**

Oct4, Sox2 and Nanog crossly regulate each other, forming an interconnected auto-regulatory and feedforward circuitry known as the Oct4/Sox2/Nanog (OSN) triumvrate [33]. They function cooperatively to activate the expression of genes required to maintain pluripotency, and at the same time repress genes involved in lineage specification [472] [33]. The ability of OSN to positively or negatively regulate gene expression is based on the interaction with other transcription factors and epigenetic machineries. Chen et al. demonstrated that the OSN collaboratively activate gene expression via binding to the enhancer site [68]. Most of these binding sites are occupied with coactivator p300 and mediator [68] [24] [182]. Furthermore, Yuan et al. showed that Oct4 recruits Setdb1, which catalyses the repressive histone modification H3K9me3 at genes associated with trophectoderm differentiation, such as *Cdx2* and *Tcfap2a* [474]. Similar findings have been reported by the Young lab [24]. Additionally, Liang et al. showed that Nanog and Oct4 associate with specific repressor protein from the NuRD complex, namely *Hdac1/2* and *Mta1/2*, to form a complex and co-occupy Nanog target genes for developmental gene repression [226]. Importantly, pluripotency signalling pathways are wired into the OSN circuitry to deliver exogenous information to the genome in the form of activate transcription factors [472]. The OSN binding sites are correlated with the binding of Stat3, Tcf3 and Smad1, which are the effectors of LIF/STAT3, Wnt and BMP4 signalling pathways. Loss of Oct4 leads to a loss of co-binding of these transcription factors, indicating that these pathways regulate pluripotency by directly deliver signals to the core regulatory circuitry [68] [232] [77] [472].

#### **1.3.3.2.2 Ancillary pluripotency regulators**

Further to the core pluripotency circuitry established by the OSN triumvrate, ESCs also express a repertoire of ‘ancillary’ pluripotency regulators that are individually dispensable but collectively reinforce naive pluripotency. A large-scale RNAi screen performed by Ivanova et al. identified several ancillary pluripotency factors including *Esrrb*, *Tbx3*, *Tcl1* and *Dppa4* [170]. Among them *Esrrb* appears to play an especially crucial effect, probably because it interacts with Oct4 and is directly up-regulated by Nanog [111] [485]. *Esrrb* over-expression showed an enhanced self-renewal phenotype and ESC pluripotency

can be sustained without LIF [111]. During development, *Esrrb* is required in placenta formation but not the embryo [242], indicating that its role in pluripotency can be substituted by alternative pathways. Consistently, it was shown that *Esrrb* knockout ESCs can be isolated and propagated in serum/LIF with sustained Oct4 expression [111]. It was also reported that *Esrrb* is the principal target of Tcf3 and forced expression of *Esrrb* render ESCs propagation without GSK3 inhibitor [259]. Furthermore, *Esrrb* is dispensable in the presence of LIF, confirming the functional compensation by LIF/STAT3 [259]. *Tbx3* is shown to be regulated by the PI3K/Akt pathway downstream of LIF [297], while *Klf4* is found to be a direct target of Stat3 [148]. Forced expression of *Tbx3* and *Klf4* is sufficient to maintain self-renewal of ESCs in LIF-free condition [297]. A study performed by Martello et al. identified another ancillary factor *Tfcp2l1*, which is another non-compensable downstream target of STAT3 [257]. Other ancillary factors include but not limited to Tcf3, *Klf2*, *Sall4*, *Prdm14*, *Pum1* and *Zfp706*. These factors are expressed uniformly in 2i but heregenously in serum.

### 1.3.3.3 Epigenetic regulation

#### 1.3.3.3.1 DNA methylation

Methylation at CpG dinucleotide is a repressive epigenetic modification at the level of DNA [385]. Once established, DNA methylation (5mC) is stably maintained by DNA methyltransferase 1 (DNMT1) and propagated through cell division. It was found that during early development, 5mC is dynamically erased, resulting in a globally hypomethylated state in the ICM [384][380]. It was proposed that the global hypomethylation is to remove epigenetic barriers and facilitate pluripotency acquisition [147].

There are mainly two possible mechanisms of DNA demethylation: the replication-independent active DNA demethylation and the replication-dependent passive DNA demethylation [438]. As an effector involved in the active DNA demethylation, the activation-induced cytidine deaminase (AID) has been shown to demethylate the promoters of *Oct4* and *Nanog* during human fibroblast reprogramming [20] [323]. However, controversial findings have been reported in certain mouse ESC lines [120]. The passive DNA demethylation pathway involves oxidation of 5mC to 5-hydroxy-methylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) by the ten-eleven translocation (TET) family of enzymes [223]. *Tet1* and *Tet2* are highly expressed in mouse ESCs and are down-regulated upon differentiation [169] [201]. It was demonstrated that the rate and extent of DNA demethylation was compromised in *Tet1* and *Tet2* deficient ESCs [112]. Furthermore, si-



lencing of Tet1 resulted in downregulation of pluripotency genes such as *Nanog*, *Esrrb* and *Klf4* [169] [123] [201].

As the mouse embryonic development proceeds, DNA methylation is reestablished by Dnmt3a and Dnmt3b. Embryos with mutant Dnmt3b was normal in early developmental stages but was defective in later stages [304]. Furthermore, it was shown that Dnmt-deficient ESCs exhibited normal self-renewal ability but lost differentiation potential [67] [419] [266]. Collectively, it suggests that DNA methylation is important for lineage specification but not ESC maintenance.

Recent studies demonstrated that ESCs cultured in 2i/LIF exhibited global hypomethylation which is comparable with the ICM, whereas ESCs in serum/LIF accumulated a much higher level (approximately 3-fold) of DNA methylation, which resembles the hypermethylated state of the postimplantation epiblast [113] [146]. Interestingly, the Nanog/Rex1-positive cells in serum/LIF also retain high global 5mC, indicating they may not be at the ground state, which is consistent with the heterogeneity and primed feature of serum/LIF cultured cells [113]. The difference in DNA methylation between serum/LIF and 2i/LIF indicates profound effect of exogeneic signalling pathways on the epigenetic landscape in ESC maintenance.

#### 1.3.3.3.2 Bivalent domains

The developmental promoters in pluripotent stem cells are featured by the co-presence of activating modification H3K4me3 and the repressive modification H3K27me3, which is a phenomenon known as bivalency [458] [8] [308]. These conflicting marks are commonly observed in pluripotent stem cells but very rarely in somatic cells [17] [267]. The bivalent signature is thought to keep lineage-specific genes silenced yet maintaining a poised state so that they can be rapidly reactivated in response to differentiation cues [438]. The H3K27 methylation is catalysed by Polycomb repressive complex 2 (PRC2), which is composed of Ezh2, Eed and Suz12. It has been shown that the polycomb complexes are dispensable for ESC self renewal, but simultaneous knockout of PRC1 and PRC2 ESCs failed to differentiate into three germ layers, suggesting that the repressive epigenetic modifications are primarily function in the initiation of differentiation rather than pluripotency maintenance [55] [218]. Notably, cells cultured in 2i/LIF exhibited decreased H3K27 modifications on bivalent domains compared to cells in serum/LIF. It was thought that this could be a an effect of Erk inhibition since Erk is required for the activity of Eed [409]. The methylation of H3K4 is mediated by the Trxthorax group (TrxG) complex

such as Wdr5. It was shown that Wdr5 physically interact with Oct4 and genome-wide protein localisation analysis revealed overlapping gene regulatory functions between Oct4 and Wdr5 [3]. Additionally, depletion of Wdr5 resulted down regulation of Oct4 target genes and resulted in loss of self renewal [3]. It has been widely accepted that the bivalent histone modifications are the unique feature of pluripotent stem cells that keeps genes in an inducible state and increases robustness at the same time. However, some evidence suggested that bivalency may be functionally dispensable [83] [429] [147]. Thus more work needs to be done to directly probe the function of bivalent domains in development.

### 1.3.3.3.3 Heterochromatin organisation

Higher order of chromatin remodelling has been shown as an important machinery that facilitates coordinated action on gene expression. H3K9 methylation marks constitutive heterochromatin in pericentric and telomeric regions. Immunostaining of heterochromatin protein1 (HP1) and H3K9me3 revealed a hyperdynamic and less compartmentalised structure in ESCs, indicating of chromatin reorganisation during differentiation [271]. This relatively diffused heterochromatin structure is a functionally important hallmark of pluripotent stem cells, which helps to maintain plasticity and establish higher-order chromatin structure upon differentiation. It was shown that Oct4 regulates H3K9 methylation via up-regulating *Jmjd1a* and *Jmjd2c*, which encode H3K9me2/3 demethylases [233]. Depletion of *Jmjd1a* and *Jmjd2c* resulted in down regulation of pluripotency genes and differentiation of ESCs [233].

The maintenance of pluripotency has been extensively studied over the past decades. However, up until now, there is still a lack of knowledge on the exact mechanism of the initial transition towards differentiation. With the advent of the CRISPR-Cas9 technology, I sought to perform a genome-wide knockout screen to study the exit of pluripotency in a comprehensive in-depth manner. In order to do that, specific investigation steps were undertaken. First, careful preparation and optimisation was conducted to establish the screening conditions. Second, the genome-wide CRISPR-Cas9-mediated screen was performed and result was analysed. Finally, the role of mTORC1-related pathways in the regulation of pluripotency and differentiation was investigated.

## Chapter 2

# Materials and Methods

## 2.1 Cell Culture

### 2.1.1 Materials

#### 2.1.1.1 Cell lines

Cells	Source
Rex1:GFPd2 reporter cell line	Mouse E14 with expression of a destabilised GFP (GFPd2) driven by the endogenous <i>Rex1</i> promoter [450]
Cas9-expressing JM8 cell line	Generated by Koike-Yusa et al. [202]
HEK293FT	Human embryonic kidney cells obtained from Invitrogen

#### 2.1.1.2 Media components, inhibitors and other reagents used in cell culture

Components	Supplier	Catalog No.
Knockout DMEM	Life Technologies	10829018
DMEM	Life Technologies	21969
Fetal Bovine Serum (FBS)	Life Technologies	10500064
GlutaMax (100X)	Life Technologies	35050061
Non-essential Amino Acids (100X)	Life Technologies	11140035
$\beta$ -Mercaptoethanol	Sigma-Aldrich	M6250
NDiff 227	Clontech	Y40002
Knockout Serum Replacement (KSR)	Life Technologies	10828
20% Bovine Serum Albumin (BSA)	Life Technologies	15260037
Penicillian-Streptomycin (100X)	Life Technologies	5140122
PD0325901	Selleck Chemicals	S1036
CHIR99021	Selleck Chemicals	S1263
Leukemia inhibitory factor (LIF)	Millipore	ESG1107
G418, Geneticin	Invitrogen	10131027
Rapamycin	Sigma Aldrich	R8781
Dimethyl Sulfoxide (DMSO)	Sigma Aldrich	276855
Phosphate Buffered Saline (PBS)	Sigma-Aldrich	806544

**2.1.1.3 Dissociation agents**

Components	Supplier	Catalog No.
Trypsin- 0.25% EDTA	Life Technologies	25200056
Accutase	Millipore	SCR005

**2.1.1.4 Other chemicals and kits**

Components	Supplier	Catalog No.
Leukocyte Alkaline Phosphatase kit	Sigma-Aldrich	86R-1KT
Paraformaldehyde (PFA)	Sigma-Aldrich	158127
Mytomycin C	Sigma-Aldrich	50-07-7
Polybrene	Sigma-Aldrich	86R-1KT
Lipofectamin LTX	Invitrogen	15338100

**2.1.1.5 Media**

Media	Components
Serum/LIF condition	KnockOut DMEM supplemented with: 15%FBS 1 x GlutaMax 1 x NEAA 0.1mM $\beta$ -mercaptoethanol 1000 U/mL LIF
2i/LIF condition	NDiff 227 supplemented with: 5%BSA 1%KSR 1 x NEAA 0.1mM $\beta$ -mercaptoethanol 1 $\mu$ M PD0325901 3 $\mu$ M CHIR99021 1000 U/mL LIF
Medium for feeder cells	KnockOut DMEM supplemented with: 10%FBS 1 x GlutaMax 1 x NEAA 0.1mM $\beta$ -mercaptoethanol
Medium for HEK293FT (D10)	DMEM supplemented with: 10%FBS 1 x GlutaMax

## 2.1.2 Methods

### 2.1.2.1 Routine culture and maintenance mESCs

ESCs cultured under serum/LIF condition were routinely passaged on mitotic-inactivated feeder layer everyday at a split ratio of 1:2-1:3. Trypsin-0.25% EDTA was used followed by mechanistic dissociation to achieve single cell suspension before replating. Feeder cells were prepared by treating MEF with 15  $\mu\text{g}/\text{mL}$  mitomycin C for 3 hours. Feeder cells were plated one day before passaging.

ESCs cultured in 2i/LIF were passaged every two or three days on 0.1% gelatin coated plastics at a split ratio of 1:8-1:10 depending on confluency. Accutase was used as a dissociation agent. Medium was replenished everyday. For freezing, mESCs were harvested and resuspended in freezing medium (90% culture medium, 10%DMSO) and stored overnight at  $-80^\circ\text{C}$  before transfer into liquid nitrogen for long-term storage.

### 2.1.2.2 Differentiation condition

ESCs were detached with Accutase and plated at a density of 10,000/cm<sup>2</sup> in N2B27 medium on 0.1% gelatin coated plates in the absence of LIF or 2i to allow differentiation. Medium was changed every day.

### 2.1.2.3 Transfection of mESCs

#### Lipofection

Lipofection was conducted using the Lipofectamine LTX Reagent as per manufacturer's protocol. For  $4 \times 10^5$  ESCs, the DNA transfection complex was prepared as follows: 2  $\mu\text{g}$  kit-purified DNA and 2  $\mu\text{PLUS}$  reagent were mixed with 500  $\mu\text{L}$  OptiMEM and incubated at room temperature for 5 minutes. LipofectamineLTX reagent (6  $\mu\text{L}$ ) was added and further incubated for 30 minutes. The DNA transfection complex was mixed with cell suspension and plated on MEF in serum/LIF. Medium was replenished after 3 hours.

#### Electroporation

A suspension of  $10 \times 10^7$  ESCs were mixed with 25  $\mu\text{g}$  DNA in 800  $\mu\text{L}$  PBS. The mixture was transferred to a 0.4 cm gap cuvette (Biorad) and electroporated at 240 V, 500  $\mu\text{F}$  using the Gene Pulser Xcell Electroporation systems (Biorad). The electroporated cells were plated on feeders in serum/LIF at a density of approximately 50,000 cells per cm<sup>2</sup>.

Selective drugs were added about 16 to 18 hours later. Medium was replenished every day. Colonies were picked and genotyped about 6 to 8 days later.

#### 2.1.2.4 Lentivirus production and transduction

##### Lentivirus production

HEK293FT cells were plated the day before transfection at a density to achieve 70-80% confluency at the time of transfection. To produce virus from one 10-cm cell culture dish, 5.4  $\mu\text{g}$  lentiviral transfer vector, 5.4  $\mu\text{g}$  psPax2, 1.2  $\mu\text{g}$  pMG2.G and 12  $\mu\text{L}$  PLUS reagent were mixed in 3 mL Opti-MEM and incubated for 5 minutes at room temperature. 36  $\mu\text{L}$  Lipofectamine LTX was added to the mixture and further incubated for 30 minutes at room temperature. At the end of the incubation, D10 was replaced with 5 mL OptiMEM from the dish culturing HEK293FT cells before DNA/Lipofectamine complex was added to the cells. The cells were incubated with the DNA/Lipofectamine complex for 6 - 8 hours, after which medium was replaced with 10 mL D10. The amount of reagents were scaled up or down accordingly depending on the amount of virus to be produced. Two days after transfection, viral supernatant was collected with a syringe and filtered through a 0.45  $\mu\text{m}$  filter cartridge. The filtered viral supernatant was aliquoted and stored at  $-80^\circ\text{C}$ .

##### Lentivirus transduction

$0.5 \times 10^6$  ESCs were resuspended in 500  $\mu\text{L}$  diluted viral supernatant supplemented with polybrene at a final concentration of 4  $\mu\text{g}/\text{mL}$ . The ESC-lentivirus suspension was incubated at  $37^\circ\text{C}$  for 30 minutes before being plated in serum/LIF on feeders. Medium was changed the next day. The amount of reagents was scaled up or down depending on the amount of cells to be transduced.

#### 2.1.2.5 Genetic manipulation of cell lines

##### Cas9 knockin

The *Rosa26* Cas9 targeting vector (Figure 3.2) was linearised with PacI and introduced into the Rex:GFPd2 cells by electroporation. Transfected cells were selected in 180 mM G418 for 8 days, after which single colonies were picked and expanded. Genomic PCR was performed using LongAmp Taq DNA polymerase (NEB) with the following primers:

Forward primer: TCGCATTGTCTGAGTAGGTGTCATTCTA

Reverse primer: CTAACAAAACGTCTCAACTTCAAGGTGA

### Generation of stable knockout cell lines

Several stable knockout mESC lines were generated using paired gRNAs to delete one 'critical' exon, which is defined as an exon that is common to all transcripts and creates a frame-shift mutation when deleted.. Cas9-expressing Rex1GFPd2 cells were transfected with 2 plasmids encoding gRNA#1 (1  $\mu$ g) , gRNA#2 (1  $\mu$ g). gRNA sequence as shown in Table:2.1. Three days after transfection, BFP-positive cells were sorted and plated in serum/LIF on MEF at a density of 1000 cells per 10cm dish. Single colonies were picked and genotyped after 6 to 8 days. PCR genotyping primer sequence as shown in Table:2.2.

**Table 2.1:** gRNA pairs used to generate stable knockout cell lines

Gene	gRNA1	gRNA2
<i>Tcf7l1</i>	GCTCCCAAAGAGCGGTGGTG	TGAAAGGAGCCACCGGTGAG
<i>Apc</i>	ACAAGCTAATACATATTGCC	GACAGTGCAGCTTTTAGATT
<i>Tsc1</i>	GGCGACATCAGGCTCAGCAC	GCAGCCATGTGTATGCGGGA
<i>Tsc2</i>	AAAAGGTGCTGCAGTTCACG	TCCAAGCTTAATGCATTAGG
<i>Nprl2</i>	CCGGACCCAACGTCCACTGA	GTGTGAGGCTTTAGTTGGGT
<i>Depdc5</i>	CACAGGGCACCCCATCATGT	ACAAAACATGCTCGTCTCTA
<i>Rraga</i>	GTCGCCACGGACTGGGCCTG	CTGATAGACGATGCTGGACC
<i>Rictor</i>	TGCTTGTCTATGCACAATTT	TAACAATTTAAGTCCGAGCT

**Table 2.2:** PCR genotyping primers

Gene	Forward primer	Reverse primer
<i>Tcf7l1</i>	AGCCATTTTGACGTCTGTCC	CCGAGAGCTCCTGTGTCAGAAC
<i>Apc</i>	CGTCAGTGCAGTGTTTCCTTC	AGTCTGAAGTCAGCCCAGGA
<i>Tsc1</i>	GGGGATAGGGATAGGGGTCT	ATGAACTGCAGGGTTTCTGG
<i>Tsc2</i>	GACAGGAGGCAAGCAGAAAC	GCTAGAGAAGGGCAGGGAGT
<i>Nprl2</i>	CAAAGTAGACCACTGGGTGGA	AGAAGAAGCTGATTGGCTGC
<i>Depdc5</i>	ACTCTCAGGGAAAAGGCAGA	TGCTTTTGCAAGTCAAGTCG
<i>Rraga</i>	TGGTCTGCTCTCGCTAGCTC	TTTGCCAGATTATTGAGGC
<i>Rictor</i>	ACGGTGGGACAGAAACTCAG	TCAAGCAGTTTCAGTGCCAC

#### 2.1.2.6 Commitment assay

ESCs were plated under differentiation condition for 26-28 hours before being detached and replated in N2B27 supplemented 2i/LIF at a density of approximately 20,000 cells per well of a 12-well plate. Uncommitted ESCs were allowed to expand for 6-8 days before AP staining was performed. Medium was replenished every day.



### 2.1.2.7 Flow cytometry and cell sorting

Flow cytometric analysis was performed to analyse Rex1-GFP expression. ESCs were dissociated into single cells by trypsin and resuspended in 1% (V/V) BSA in PBS followed by filtration through a 35  $\mu$ m cell strainer to remove any cell clumps. Filtered cells were immediately brought for analysis using the LSR Fortessa instrument (Becton Dickinson). Data was analysed using FlowJo.

For transduction efficiency measurements, harvested ESCs were fixed with 4% (V/V) PFA for 20 minutes. The fixed cells were washed twice with PBS before being resuspended in 1% BSA and analysed on the LSR Fortessa instrument.

To perform fluorescent activated cell sorting (FACS), harvested ESCs were filtered and resuspended in culture medium to a final density of approximately  $5 \times 10^6$  per ml and left on ice until sorting. The sorted cells were either pelleted down and stored at  $-80^\circ$  C for DNA/RNA extraction, or plated in culture medium supplement with Penicillin-Streptomycin for further analysis.

## 2.2 Molecular Biology

### 2.2.1 Materials

#### 2.2.1.1 Molecular chemicals and Kits

Reagents	Supplier	Catalog No.
1 kb DNA ladder	NEB	N3232S
100 bp DNA ladder	NEB	N3231S
Ampicillin	Sigma Aldrich	59349
DH5 $\alpha$ competent cells	Invitrogen	18263012
LongAmp Taq DNA polymerase	NEB	M0323S
Q5 Hot Start High-Fidelity 2XMaster Mix	NEB	M0494S
10x Ligation buffer	NEB	M0202S
T4 Ligase	NEB	M0202S
T4 PNK	NEB	M0201
Proteinase K	Roche	03115879001
DNeasy Blood and Tissue Kit	Qiagen	69504
Blood and Cell Culture DNA Midi Kit	Qiagen	13343
Blood and Cell Culture DNA Maxi Kit	Qiagen	13362
QIAprep Spin Miniprep Kit	Qiagen	27104
QIAquick PCR Purification Kit	Qiagen	28104
RNeasy Mini Kit	Qiagen	74104
Agencourt AMPure XP beads	Beckman	A63881
DynaMag-96 Side Magnet	Invitrogen	12331D

**2.2.1.2 Immunoblotting**

<b>Reagents</b>	<b>Supplier</b>	<b>Catalog No.</b>
RIPA buffer	Sigma-Aldrich	R0268
Protease Inhibitor Cocktail	Sigma-Aldrich	P8340
Phosphatase Inhibitor Cocktail	Sigma-Aldrich	P5726
Bradford protein assay	Bio-Rad	5000006
Colour Prestained Protein Standard (11-245 kDa)	NEB	P7712
NuPAGE MOPS SDS Running Buffer (20X)	Invitrogen	NP0001
NuPAGE Transfer Buffer (20X)	Invitrogen	NP00061
BSA lyophilized powder	Sigma-Aldrich	A9647
Nonfat dry milk blotting-grade blocker	Bio-Rad	1706404
NuPAGE Sample Reducing Agent (10X)	Invitrogen	NP0009
NuPAGE LDS Sample Buffer (4X)	Invitrogen	NP0007
NuPAGE Novex 4-12% Bis-Tris Protein Gels, 10-well	Invitrogen	NP0321
NuPAGE Novex 4-12% Bis-Tris Protein Gels, 12-well	Invitrogen	NP0322
Pierce TBS Buffer (20X)	Thermo Scientific	28358
Tween 20	Sigma-Aldrich	90050-64-5
Thick Blot Paper	Bio-Rad	1703932
Amersham ECL start Wester Blotting Detection Reagent	GE Healthcare	RPN3244
Amersham Hypercassette	GE Healthcare	RPN11642
Amersham Hyperfilm ECL	GE Healthcare	28906837
Amersham Hybond P 0.45 PVDF	GE Healthcare	10600023
Cell scraper	Corning	734-1537
Mini Trans-Blot Electrophoretic Transfer Cell	Bio-Rad	1703930
XCekk SureLock Mini-Cell Electrophoresis System	Thermo Fisher	EI0002

<b>Antigen</b>	<b>Supplier</b>	<b>Source</b>	<b>Dilution</b>
Akt-pThr308	Cell Signaling-9275	Rabbit/Polyclonal	1:1000
Akt-pSer473	Cell Signaling-9271	Rabbit/Polyclonal	1:1000
Akt	Cell Signaling-9272	Rabbit/Polyclonal	1:1000
S6K-pThr389	Cell Signaling-9205	Rabbit/Polyclonal	1:1000
S6K	Cell Signaling-9202	Rabbit/Polyclonal	1:1000
S6-pSer235/236	Cell Signaling-2211	Rabbit/Polyclonal	1:1000
S6	Cell Signaling-2217	Rabbit/Monoclonal	1:1000
GSK3 $\beta$ -pSer9	Cell Signaling-5558	Rabbit/Monoclonal	1:1000
GSK3 $\beta$	Santa Cruz-9166	Rabbit/Polyclonal	1:1000
PRAS40-pThr246	Cell Signaling-13175	Rabbit/Monoclonal	1:1000
PRAS40	Cell Signaling-2691	Rabbit/Monoclonal	1:1000
$\beta$ -Actin	Sigma-Aldrich-A2228	Mouse/Monoclonal	1:1000
Rabbit IgG-HRP	GE Healthcare-NA934	Donkey	1:1000
Mouse IgG-HRP	GE Healthcare-NA931	Donkey	1:1000

## 2.2.2 Methods

### 2.2.2.1 Isolation of nucleic acids

Plasmid DNA was obtained using the QIAprep Spin Miniprep Kit. Genomic DNA was obtained from mESCs collected from cell sorting or direct trypsinisation. DNeasy Blood and Tissue kit was used when cell number was less than 5 million. Blood and Cell Culture DNA Midi kit was used when cell number was in between 5 million and 20 million. Blood and Cell Culture DNA Maxi kit was used when cell number was above 20 million. RNA was extracted using the RNeasy Mini Kit.

### 2.2.2.2 gRNA cloning

gRNA expression vector (Figure 4.1 (A)) was linearised with BbsI and purified from gel. 1 pmol each of top and bottom strand oligo were mixed with 1  $\mu$ L of 10X T4 ligation buffer, 0.5  $\mu$ L of T4 PNK and dH<sub>2</sub>O to achieve a total volume of 10  $\mu$ L. Reaction was performed in a thermal cycler with the following conditions: 37 ° C for 30 minutes followed by 95 ° C for 5 minutes, after which temperature was set to ramp down to 25 ° C at 0.1 ° C/second. 14.2 fmol annealed ds-oligo was ligated with 3.7 fmol linearised vector by mixing with 1  $\mu$ L T4 ligase and 1  $\mu$ L 10X ligase buffer in a total volume of 10  $\mu$ L. The ligation mixture was incubated at 16 ° C for 4-16 hours. Ligated plasmid was introduced into DH5 chemical competent cells following standard bacteria transformation protocol. Bulk plasmid was prepared by inoculating 2 mL of 2X YT media (50  $\mu$ g/mL Amp) and shaking overnight at 37 ° C. Plasmid DNA was purified with Miniprep kit. To prepare single clone, transformed bacteria was plated on LB Amp plates followed by expanding in liquid culture. Plasmid sequence was checked by capillary sequencing.

### 2.2.2.3 Lysate PCR

Approximately 5,000 cells were collected per PCR tube and washed twice with PBS. 25  $\mu$ L water was added to the cell pellet and heated for 10 minutes at 90 ° C for cell lysis. 5  $\mu$ L proteinase K was added to each well and reaction was carried out at 55 ° C for 60 minutes. The enzyme was heat inactivated at 95 ° C for 10 minutes. PCR was carried out using 10  $\mu$ L cell lysate in a 50  $\mu$ L PCR reaction with LongAmp Taq DNA polymerase following the manufacturer's protocol.

### 2.2.2.4 Illumina library preparation

At the end of the screen, gRNAs were amplified from lentivirally transduced cells using the following primers:

Forward primer: ACACTCTTTCCCTACACGACGCTCTTCCGATCTCTTGTGGAAAG-GACGAAACA; Reverse primer: TCGGCATTCCTGCTGAACCGCTCTTCCGATCTC-TAAAGCGCATGCTCCAGAC. The following reaction mix was set up:

Reagent	Volume per reaction
2x Q5 HS HF	25 $\mu$ L
Primer mix (10 $\mu$ M each)	1 $\mu$ L
Genomic DNA	2 $\mu$ g
dH <sub>2</sub> O	up to 50 $\mu$ L

To cover the complexity of the gRNA library, approximately 72  $\mu$ g genomic DNA was analysed for the control samples. All the genomic DNA was analysed for sorted Rex1-GFP positive cells. The PCR reaction was set to run with the following programme:

Cycle number	Denature	Annealing	Extension
1	98 ° C, 30 seconds		
2-26	98 ° C, 30 seconds	61 ° C, 15 seconds	72 ° C, 20 seconds
27			72 ° C, 2 minutes

The PCR end product was purified using QIAGEN's PCR purification kit and used for second round PCR.

Reagent	Volume per reaction
2x KAPA HiFi HotStart ReadyMix	25 $\mu$ L
Primer mix (5 $\mu$ M each)	2 $\mu$ L
First round PCR product	1 ng
dH <sub>2</sub> O	18 $\mu$ L

Cycle number	Denature	Annealing	Extension
1	98 ° C, 30 seconds		
2-9	98 ° C, 10 seconds	66 ° C, 15 seconds	72 ° C, 20 seconds
10			72 ° C, 5 minutes

The end PCR product was purified using SPRI beads and transferred for Illumina sequencing.

### 2.2.2.5 Western blotting

ESCs were lysed on tissue culture plate directly by adding RIPA buffer supplemented with phosphatase inhibitor (1:100) and protease inhibitor (1:1000). Protein was purified from cell debris via centrifugation at 15000 rpm for 15 minutes at 4 ° C. Total protein content was quantified using Bradford Protein Assay (BioRad). Protein standard was prepared by diluting BSA to a range of 0.05 to 0.25 mg/mL. 10 µL protein standard and sample solution was added into duplicate wells of a 96-well clear flat-bottom plate. 200 µL Bradford protein assay dye reagent was added into each well. The mixture was incubated for at least 5 minutes before measuring the absorbance at 595 nm. The protein concentration was calculated based on the standard curve obtained from the readings of protein standards. The quantified protein samples were diluted with 4X sample buffer supplemented with reducing agent before being heat denaturation. The denatured protein samples were resolved at 180 V for 1 hour through a 4-12% Bis-Tris gel in SDS running buffer, then subsequently being transferred onto a PVDF membrane at 90 V for 1 hour at 4 ° C using the wet transfer method. Blots were blocked for 30 minutes in either 5% BSA or milk in TBS buffer plus 0.1% Tween 20 (TBS-T). Blots were probed in diluted primary antibody overnight at 4 ° C followed by three times washing in with TBS-T. Membranes were then probed with appropriate horse radish peroxidase (HRP)-conjugated secondary antibody for 1 hour at room temperature. Membranes were washed three times in TBS-T before the addition of ECL substrate and exposure onto the X-ray film.

## 2.3 Bioinformatics analysis

### 2.3.0.1 RNA-Seq analysis

The extracted RNA was sequenced on Illumina HiSeq2500 by 75-bp paired-ended sequencing. Sequencing data was analysed using Kallisto with the mouse RefSeq transcriptome as a reference [36]. Differential gene expression analysis was performed using DESeq2.

### 2.3.0.2 gRNA sequencing result analysis

The number of reads for each gRNA was counted with an in house scripted written by Yilong Li. The enrichment and depletion of gRNAs and genes were analysed by MAGeCK [225].

### 2.3.0.3 Gene Set Enrichment Analysis (GSEA) analysis

The GSEA analysis was performed using the online algorithm ‘GSEAPreranked’ developed and maintained by the Broad Institute

(<http://software.broadinstitute.org/cancer/software/genepattern/modules/docs/GSEAPreranked/1>)

[395]. Screening hits were pre-ranked according to the Depletion/Enrichment (DE) score, which was computed as:  $\log_{10}(\text{Depletion } P \text{ value}) + [-\log_{10}(\text{Enrichment } P \text{ value})]$ . The ‘Mitochondrion morphogenesis’ and ‘Oxidative phosphorylation’ gene sets were downloaded from the Molecular Signatures Database (MSigDB). The ‘Betschinger’ gene set contains all the validated genes from a published siRNA-based screen [19].



## Chapter 3

# Screening Preparation and Condition Optimisation

### 3.1 Introduction

Care must be taken with the design of a genome-wide screen to ensure the ultimate success of the endeavour. A successful screen must be robust and reproducible, with maximal sensitivity and minimal false-positives. Some of the major considerations are whether an assay needs to be used to measure the desired parameters, which marker to choose for that assay and whether the selected marker is truly indicative of that process. To monitor the very early stages of differentiation, I adopted the *Rex1*:GFPd2 reporter cell line kindly provided by the Smith lab. Two characteristics of the *Rex1* gene render it a desirable pluripotency marker. Firstly, as demonstrated by Masui et al., the *Rex1* function is dispensable for both the development of a mouse embryo and the maintenance of ESCs [263]. Secondly, *Rex1* expression is tightly restricted to the naive pluripotency compartment and is rapidly downregulated at the onset of differentiation [417], therefore providing an accurate biological focus and faithful pluripotency readout. Loss of *Rex1* expression leads to loss of clonogenicity under 2i/LIF conditions, indicating the irreversible exit of pluripotency [30] [185]. Notably, the downregulation of other naive pluripotency markers such as *Nanog*, *Klf2* and *Tfcp2l1* upon differentiation is earlier than that of *Rex1*, but self-renewal capacity is retained as long as *Rex1* is expressed [185]. The Smith group then generated the *Rex1*:GFPd2 reporter cell line, in which destabilised GFP with a half-life of 2 hours is expressed under the control *Rex1* promoter [450]. This reporter cell line enables almost real time monitoring of differentiation, hence providing great convenience in the fractionation of ESCs based on its naive pluripotency state by flow cytometry and further downstream analysis.

Screening parameters need to be designed to maximise the difference in gRNA abundance between treated and control samples. One of the crucial parameters is the time of examining, at which it is ideal to achieve a balance between assessing mutants with modest phenotypes and discarding a large number of outliers. In the context of a differentiation screen, conditions such as differentiation duration and FACS cut-off are crucial to success. Given the substantial complexity of the genome-wide gRNA library, it is challenging to maintain a sufficient coverage of gRNA representation during the screen. Therefore, a relatively simple differentiation protocol is favourable, especially when handling a large number of cells. A monolayer neuroectodermal differentiation method has been used in previous genetic screens to identify pluripotency regulators [469] [142] [217] [19], in which ESCs were induced to differentiate following LIF and inhibitors withdrawal. The same

can be applied in the CRISPR-Cas9 based screen and mutant candidates can be easily identified based on the persistence of Rex1GFP expression. Last but not least, it is noteworthy that there is natural variation between each cell as well as between each screen. Such variation is unavoidable and difficult to eliminate, but measures can be taken to keep it in an acceptable range. For example, it is worth checking the Cas9 activity to make sure it is uniformly active in the chosen cell line, so that the phenotype linked to a gRNA will not be masked by inactive Cas9. Including biological replicates will also help to reduce stochastic noise. Feasibility and cost needs to be taken into account when choosing the numbers of replicates.

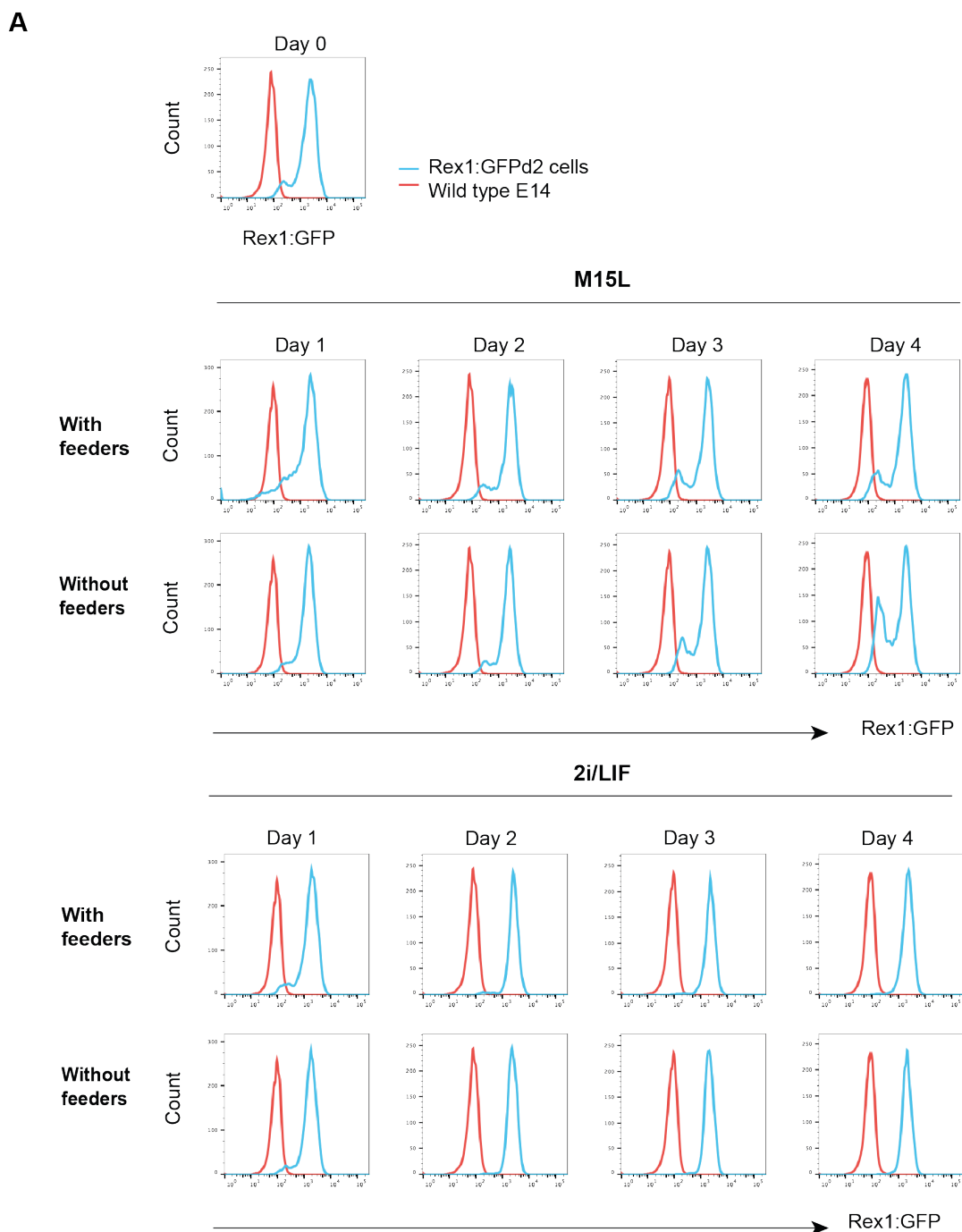
In this chapter, I describe the preparation work for the genome-wide CRISPR-Cas9 knockout screen, including analysis of culture and differentiation conditions, generation of constitutive Cas9-expressing Rex1:GFPd2 cell line, as well as proof-of-principle studies which target well-studied pluripotency-regulating genes, namely *Tcf7l1* and *Apc*. My aim was to optimise the screening conditions to achieve the highest possible sensitivity and robustness.

## 3.2 Results

### 3.2.1 Analysis of self-renewal and differentiation conditions

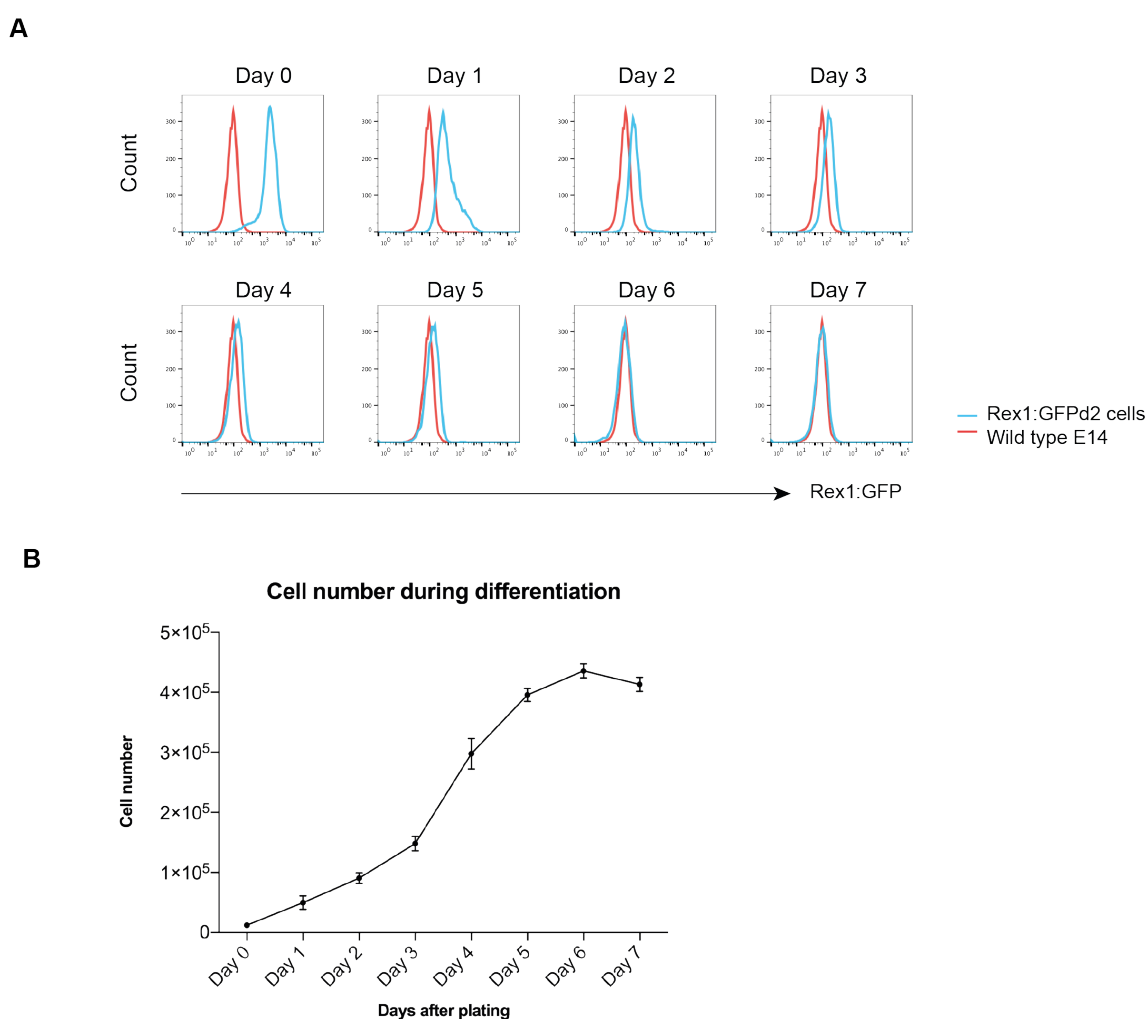
Establishing the pre-screening culture condition is important to achieve a successful differentiation screen. It is preferable to maintain ESCs as a relatively homogeneous population to minimise any biases prior differentiation. I first examined the Rex1GFP expression in serum/LIF and 2i/LIF culture, with or without feeder layer (Figure 3.1). Consistent with previous findings, *Rex1* expression was homogeneous in 2i/LIF but rather heterogeneous in serum/LIF. Notably, ESCs cultured in serum/LIF without feeders were inclined to lose *Rex1* expression, indicating a less stable pluripotency status. It is thus likely that feeders play additional roles in promoting self-renewal beyond contributing LIF. As expected, the presence of feeders did not make a difference in *Rex1* expression in 2i/LIF. As 2i effectively insulates any differentiation signals, so that the pluripotency status is rather stable without any further support. The heterogeneous ESC population became homogeneous after two days of culture in 2i/LIF, suggesting that the *Rex1*-negative primed cells can be either reversed to naive pluripotency or eliminated from the population, and that the two culture conditions are convertible. In observing the above, I decided to maintain ESCs in 2i/LIF before screening to achieve a feeder-free homogeneous population prior induction

of differentiation.



**Figure 3.1:** Analysis of Rex1GFP profile under maintenance conditions. (A) Comparison of Rex1GFP profile under different maintenance conditions. ESCs were cultured under serum/LIF condition on feeders (Day 0), and were subsequently split into four maintenance conditions: serum/LIF with or without feeders, and 2i/LIF with or without feeders. The Rex1GFP profiles of cells in each condition were measured everyday for four days. Blue - Rex1:GFPd2 cells; Red - GFP negative control, wild type ESC line E14.

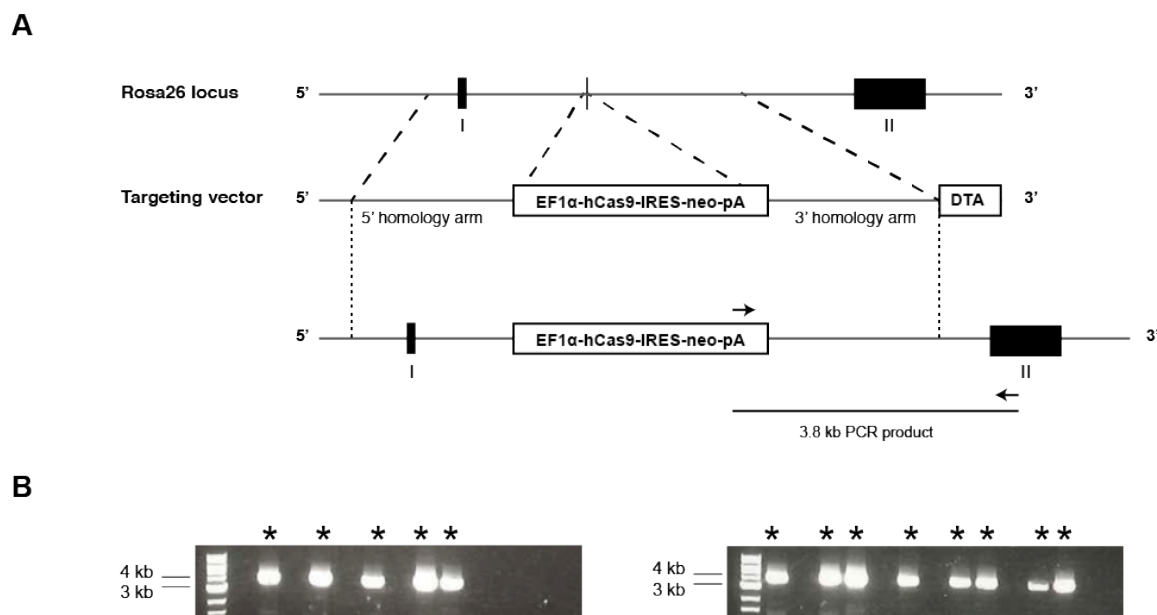
Following 2i withdrawal, the pluripotency network collapsed quickly, which was reflected by the rapid down regulation of *Rex1* expression (Figure 3.2 (A)). The most drastic transition occurred in the first three days, during which the majority of the cells lost *Rex1* expression. Taking the large complexity of the library into consideration, I then checked cell growth during differentiation to plan the scale of the screen and to ensure there was sufficient coverage of the gRNA representation and enough cells at the end of the screen for the extraction of genomic DNA. Cell number increased throughout the differentiation period until a plateau was reached on day five, indicating a healthy differentiation occurred in these conditions (Figure 3.2 (B)).



**Figure 3.2:** Analysis of Rex1GFP profile under differentiation conditions. (A) Rex1GFP expression after 2i withdrawal. The removal of 2i relieved ESCs from the shield of differentiation cues, which initiated spontaneous differentiation. ESCs were maintained in N2B27 with supplements but without 2i (described in Chapter 2). Rex1GFP expression profile was measured everyday for seven days. Blue - Rex1:GFPd2 cells; Red - GFP negative control, wild type ESC line E14. (B) ESC growth curve after 2i withdrawal. ESCs were plated on gelatin in N2B27 without LIF at a density of 10,000 cells per cm<sup>2</sup>. Cells were detached everyday for enumeration. Cell number per cm<sup>2</sup> was calculated.

### 3.2.2 Establishment of Cas9 expression in Rex1:GFPd2 cell line

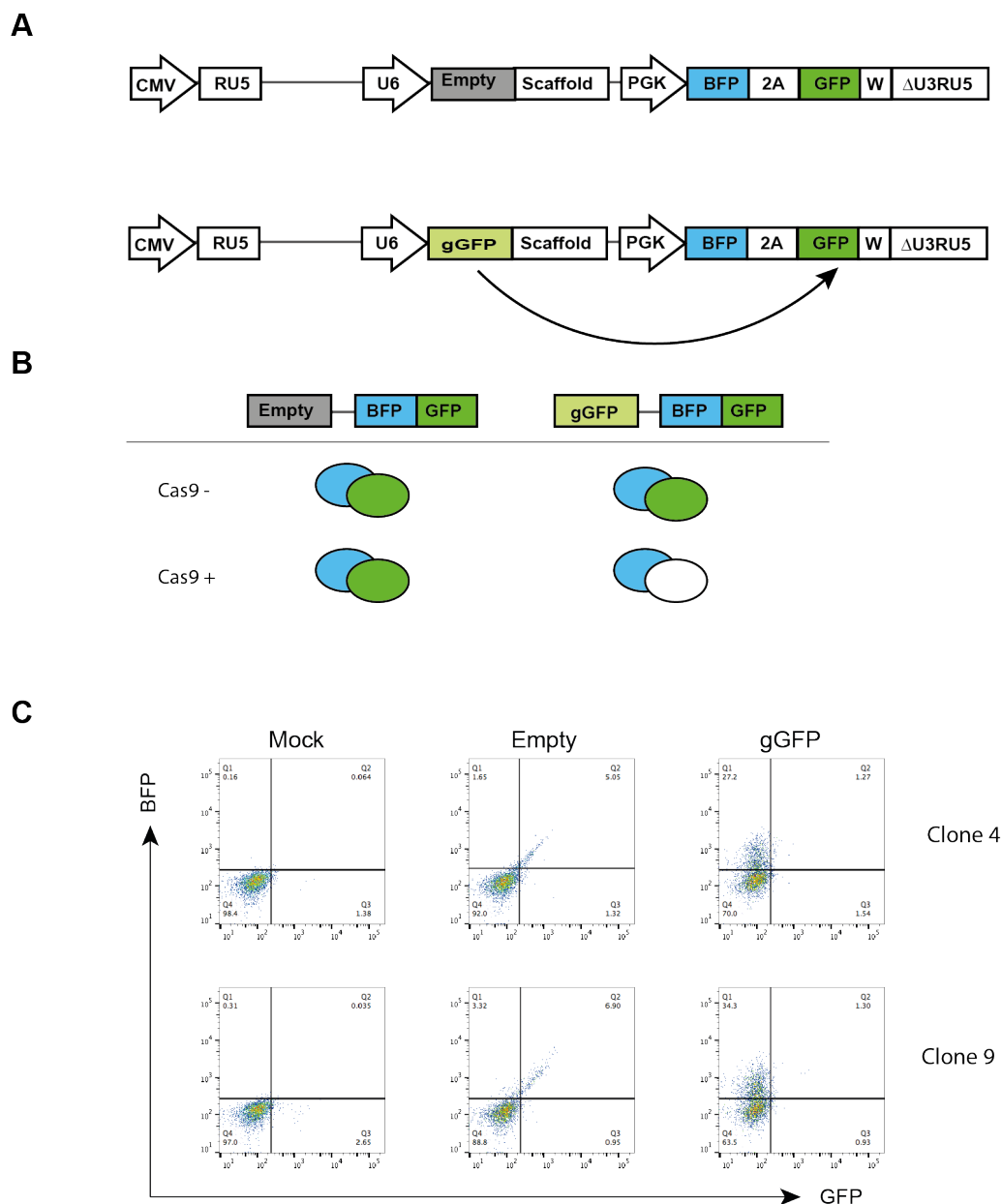
In order to perform the CRISPR-Cas9-based mutagenesis in the Rex1:GFPd2 cell line, I performed homologous recombination-mediated Cas9 knockin at the *Rosa26* locus using the previously published targeting vector [202] (Figure 3.3 (A)). G418-resistant colonies were picked and further expanded for genotyping. Out of 24 colonies analysed, 13 showed correct PCR bands (Figure 3.3 (B)).



**Figure 3.3:** Cas9 knockin. (A) Schematic diagram showing the targeting scheme at the *Rosa26* locus. EF1 $\alpha$ : elongation factor-1 $\alpha$  promoter; hCas9: human codon-optimised SpCas9; IRES: internal ribosomal entry site; neo: neomycin resistant gene; DTA: diphtheria toxin A. Arrows indicate PCR primers. (B) PCR genotyping results. PCR bands were compared to a 1 kb DNA ladder (NEB). Clones with successful knockin will produce a 3.8 kb PCR product, which were labeled with asterisk.

To investigate whether Cas9 was functional in these clones, I carried out a reporter assay developed in the lab using a lentiviral vector expressing BFP, GFP, and gRNA targeting the GFP sequence (Figure 3.4 (A)). An ‘empty’ vector, which expresses BFP and GFP but not the gRNA sequence, was included in the assay as a negative control. Transduced cells should be double positive when the ‘empty’ vector is used, regardless of its Cas9 function. When the vector containing gRNA targeting GFP is introduced, only BFP can be detected in Cas9-active cells, whereas both BFP and GFP will be detected in Cas9-inactive cells (Figure 3.4 (B)). Two clones were analysed for Cas9 function, in which fluorescent signal was exclusively detected in the BFP quadrant, indicating positive Cas9 activity in both clones. Clone 9, which has more consistent colony morphology, was selected for further

studies (Figure 3.4 (C)).



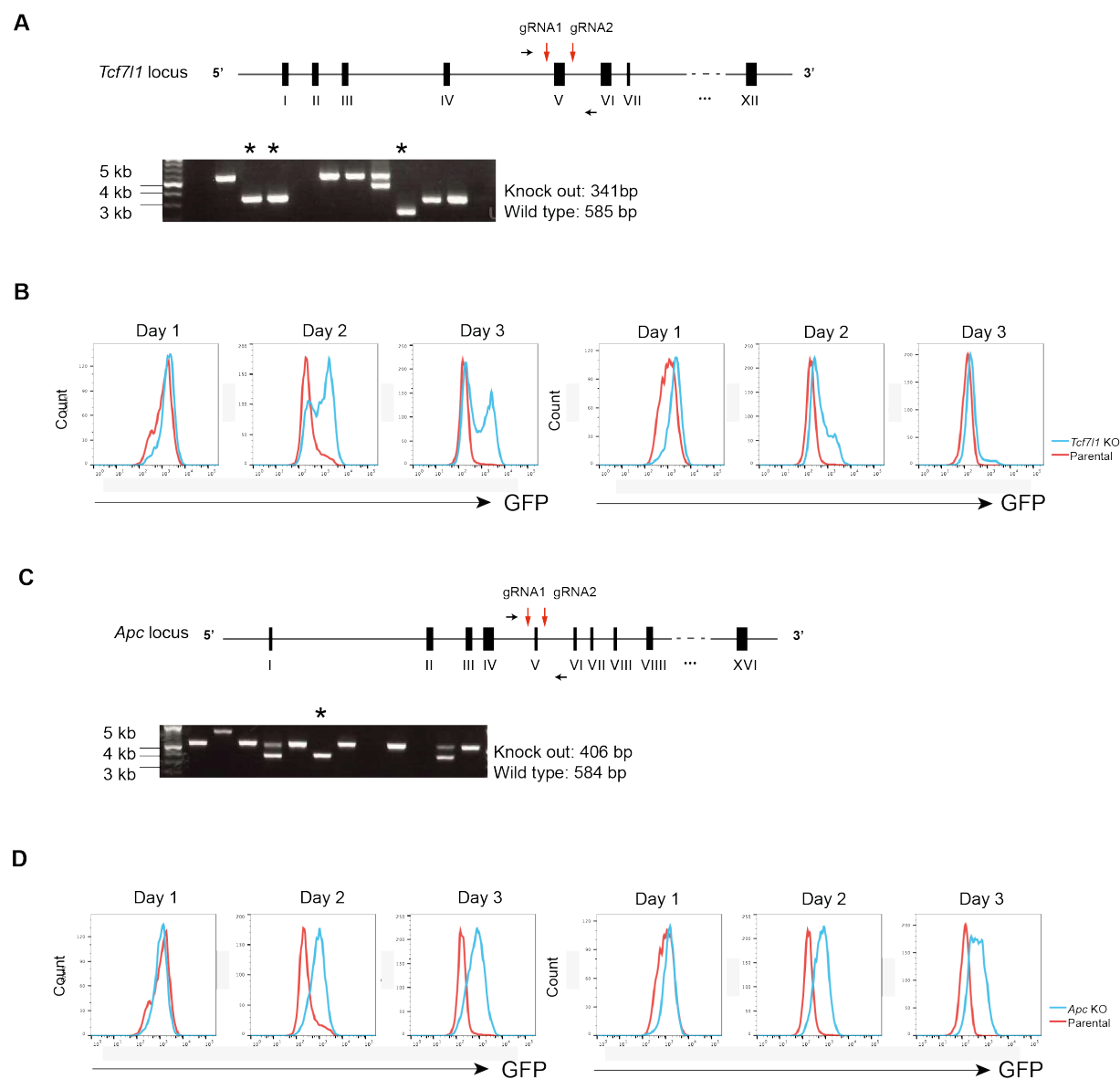
**Figure 3.4:** Cas9 function assay (A) Lentiviral vector for Cas9 function reporter assay. CMV: CMV promoter; RU5: 5' long terminal repeat; U6: U6 promoter; gGFP: gRNA targeting GFP coding sequence; Empty: The original BbsI cloning site; scaffold: gRNA scaffold; PGK: mouse phosphoglycerate kinase 1 promoter; BFP: blue fluorescent protein; GFP: green fluorescent protein; 2A, *Thossea asigna* virus 2A peptides; W: Woodchuck Hepatitis Virus posttranscriptional regulatory element;  $\Delta$ U3RU5: self-inactivating 3' long terminal repeat. Empty gRNA vector doesn't express gRNA but expresses BFP and GFP. gGFP vector expresses gRNA targeting GFP, as well as BFP and GFP. (B) The expected fluorescent expression pattern. If Cas9 nuclease was inactive, cells transduced with empty and gGFP vectors would be GFP/BFP double positive. If Cas9 was functional, cells transduced with empty vector would be double positive, whilst cells transduced with gGFP vector will express BFP only. (C) The Cas9 function of two knockin clones were analysed. Culture medium was added to mock infection instead of virus supernatant. Positive Cas9 activity was observed in both clones.

### 3.2.3 Proof of principle studies

With the establishment of the constitutive Cas9-expressing Rex1:GFPd2 cell line, I moved on to design a screening strategy which includes testing differentiation and duration conditions. To achieve that, two well-studied genes were selected as positive control genes, namely *Tcf7l1* (Tcf3) and *Apc*. Tcf3 acts as a pluripotency repressor and Apc is a subunit of the  $\beta$ -catenin degradation complex downstream of Wnt. Knocking out *Tcf7l1* and *Apc* relieves the suppression of pluripotency genes, therefore cells were expected to exhibit enhanced self-renewal and delayed differentiation phenotype.

I first sought to generate stable knockout cell lines by deleting a ‘critical exon’, which is defined as a common exon expressed in all transcript variants and when deleted, creates a frame-shift mutation. Deletion was achieved by introducing DSBs on both sides of the critical exon by the CRISPR-Cas9 system. Knockout clones were identified by PCR genotyping (Figure 3.5 (A) (C)). The selected clones were assessed under two conditions: serum-based and N2B27-based differentiation conditions. As expected, under both differentiation conditions, *Tcf7l1* knockout and *Apc* knockout showed impeded differentiation phenotype, reflected by their delayed downregulation of *Rex1* expression (Figure 3.5 (B) (D)). It appears that N2B27-based differentiation induced a more rapid decrease in Rex1GFP expression compared to serum-based differentiation, especially in *Tcf7l1* knockout. This is probably due to the distinct differentiation mechanisms. In serum-based condition, ESCs differentiate via a mixed routes towards mesoderm, endoderm and trophoderm, whereas in N2B27 without 2i/LIF, ESCs mainly differentiate towards neuroectoderm. N2B27-mediated differentiation has been adopted in large-scale RNAi- and transposon-mediated genetic screens, which can be used as reference for CRISPR-Cas9-mediated screen [19] [217]. It will also be interesting to cross compare different screening methods. For these reasons I decided to focus on N2B27-mediated differentiation.





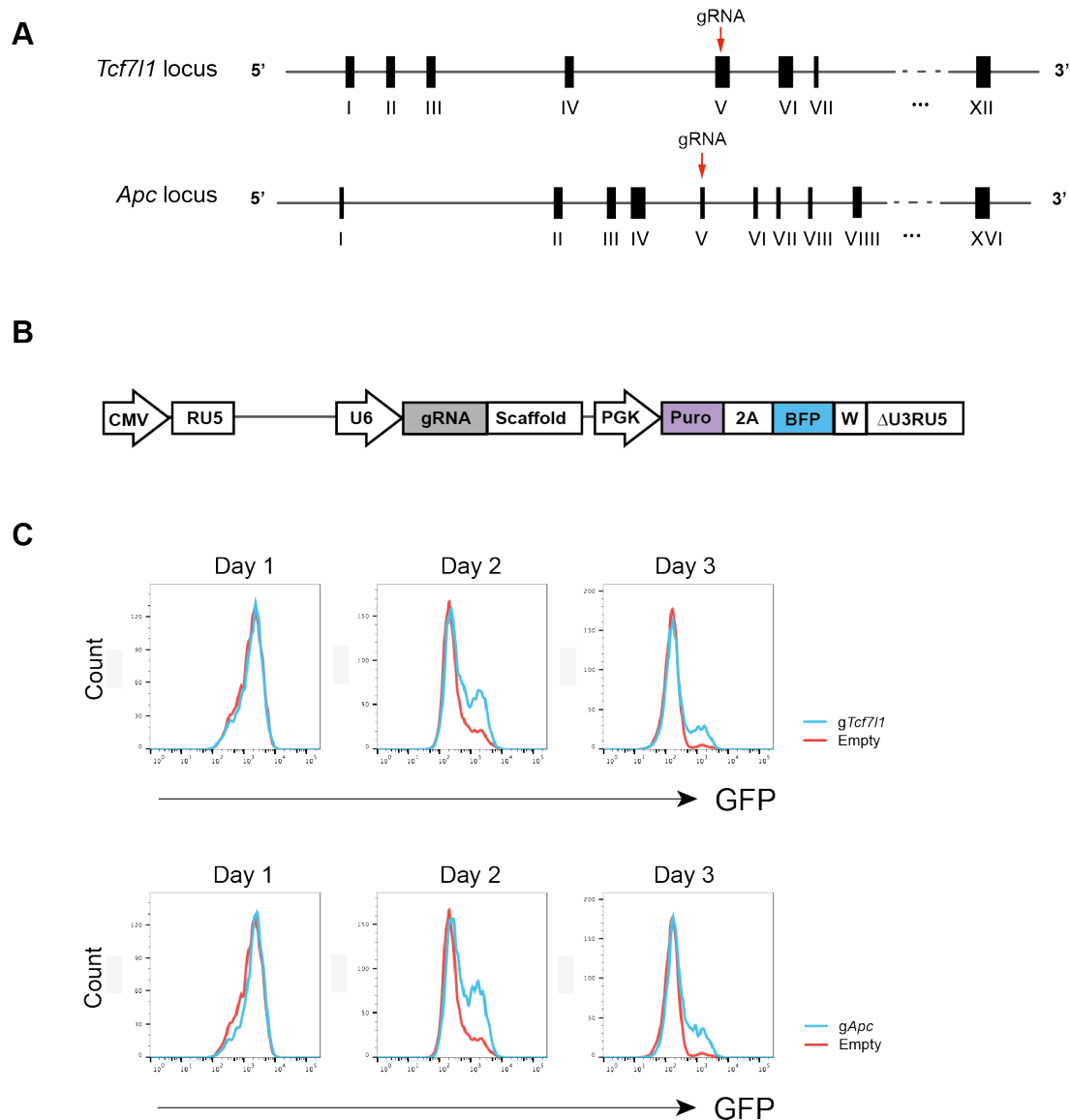
**Figure 3.5:** Positive control study with stable knockout clones. (A) Schematic diagram of critical exon deletion at the *Tcf7l1* locus and genotyping results. Black arrows indicate PCR genotyping primers. Red arrows indicate gRNA cutting sites. Asterisks indicate knockout clones. Wildtype PCR product was 585bp. Knockout PCR product was 341bp. (B) *Tcf7l1* knockout clone Rex1GFP expression profile during serum-based differentiation (left) and neuroectoderm differentiation (right). Serum-based differentiation was carried out by LIF withdrawal from serum/LIF maintenance condition. Neuroectoderm differentiation was carried out by 2i/LIF withdrawal from N2B27. Rex1GFP expression was analysed everyday for three days. Blue - *Tcf7l1*/*Apc* knockout Rex1:GFP cells; Red - wt Rex1:GFP cells. (C) Schematic diagram of critical exon deletion at *Apc* locus and genotyping result. Black arrows indicate PCR genotyping primers. Red arrows indicate gRNA cutting sites. Asterisks indicate knockout clones. Wild type PCR product was 584bp. Knockout PCR product was 406bp. (D) *Apc* knockout clone Rex1GFP expression profile during serum-based differentiation (left) and neuroectoderm differentiation (right). Serum-based differentiation was carried out by LIF withdrawal from serum/LIF maintenance condition. Neuroectoderm differentiation was carried out by 2i/LIF withdrawal from N2B27. Rex1GFP expression was analysed everyday for three days. Blue - *Tcf7l1*/*Apc* knockout Rex1:GFP cells; Red - wt Rex1:GFP cells.

In the actual screening setting, lentivirus is used as a vehicle to deliver the gRNA library and the phenotype generated by each gRNA is assessed as a mixed population with indels of various sizes. To recapitulate this, I transduced ESCs with lentivirus expressing gRNAs targeting the critical exon of *Tcf7l1* and *Apc* individually (Figure 3.6 (A) (B)), followed by differentiation in N2B27 without 2i. The resulted phenotype was compared to that of the cells transduced with gRNA-negative lentivirus. Consistent with the phenotype of pure knockouts, a clear difference in Rex1GFP expression was observed between *Tcf7l1/Apc* targeted cells and the empty control, on differentiation day 2 and day 3 (Figure 3.6 (C)). I reasoned that day 3 is the optimal time to harvest cells for gRNA representation analysis. Because vast majority of wild type cells were Rex-GFP negative on day 3, which produces a clear contrast with cells showing delayed differentiation, and a cleaner background will be obtained.

### 3.2.4 A Preliminary screen

Once the screening principle was verified by knockout of *Tcf7l1* and *Apc* via lentivirally expressed gRNAs in Rex1:GFPd2 Rosa26:Cas9 cells, I then sought to perform a preliminary screen in order to study the scale-up effect and further optimise screening conditions. Due to the reform of the manufacturing company Stem Cell Inc at the time, there was a long delay in purchasing the basal media N2B27. Therefore, I carried out the preliminary screen based on serum differentiation. Although the mechanisms of differentiation differ, the fundamental principle and design of the screen remain the same.

The Rex1:GFPd2 Rosa26:Cas9 cells were transduced with the genome-wide lentiviral gRNA library. The transduced cells were sorted according to BFP expression two days after infection, followed by three days of expansion before plating in serum-containing medium without LIF for differentiation. Differentiation medium was replenished on day two. After three days LIF withdrawal, cells retaining Rex1GFP expression were collected by FACS and subsequently pelleted for genomic DNA extraction. The gRNA sequences were amplified by PCR from genomic DNA and sent for sequencing with the Illumina HiSeq platform. A technical replicate was carried out in parallel. Sequenced gRNAs were mapped to the library and counted, and statistical analysis was performed by the computational algorithm MAGeCK.



**Figure 3.6:** Positive control study with single-gRNA knockouts. (A) Schematic diagram of gRNAs targeting the critical exon of *Tcf7l1* and *Apc*. Red arrows indicate gRNA cutting sites. The gRNA sequence is included in Chapter2 section 2.1.2.4. (B) Lentiviral vector for single gRNA knockout. CMV: CMV promoter; RU5: 5' long terminal repeat; U6: U6 promoter; gRNA: gRNA targeting *Tcf7l1* and *Apc*; PGK: mouse phosphoglycerate kinase 1 promoter; Puro: puromycin resistant gene; BFP: blue fluorescent protein; 2A, *Thosea asigna* virus 2A peptides; W: Woodchuck Hepatitis Virus posttranscriptional regulatory element;  $\Delta$ U3RU5: self-inactivating 3' long terminal repeat. (C) *Tcf7l1* and *Apc* knockout differentiation profiles. Differentiation was induced by 2i and LIF removal. Rex1GFP expression was analysed everyday for three days. Blue - *Tcf7l1/Apc* knockout Rex1:GFP cells; Red - wt Rex1:GFP cells.

Once the screening principle was verified by knockout of *Tcf7l1* and *Apc* via lentivirally expressed gRNAs in Rex1:GFPd2 Rosa26:Cas9 cells, I then sought to perform a preliminary screen in order to study the scale-up effect and further optimise screening conditions. Due to the reform of the manufacturing company Stem Cell Inc at the time, there was a long delay in purchasing the basal media N2B27. Therefore, I carried out the preliminary screen based on serum differentiation. Although the mechanisms of differentiation differ, the fundamental principle and design of the screen remain the same.

The Rex1:GFPd2 Rosa26:Cas9 cells were transduced with the genome-wide lentiviral gRNA library (Figure 3.7 (B)). The transduced cells were sorted according to BFP expression two days after infection, followed by three days of expansion before plating in serum-containing medium without LIF for differentiation. Differentiation medium was replenished on day two. After three days LIF withdrawal, cells retaining Rex1GFP expression were collected by FACS and subsequently pelleted for genomic DNA extraction. The gRNA sequences were amplified by PCR from genomic DNA and sent for sequencing with the Illumina HiSeq platform. A technical replicate was carried out in parallel. Sequenced gRNAs were mapped to the library and counted, and statistical analysis was performed by the computational algorithm MAGeCK.

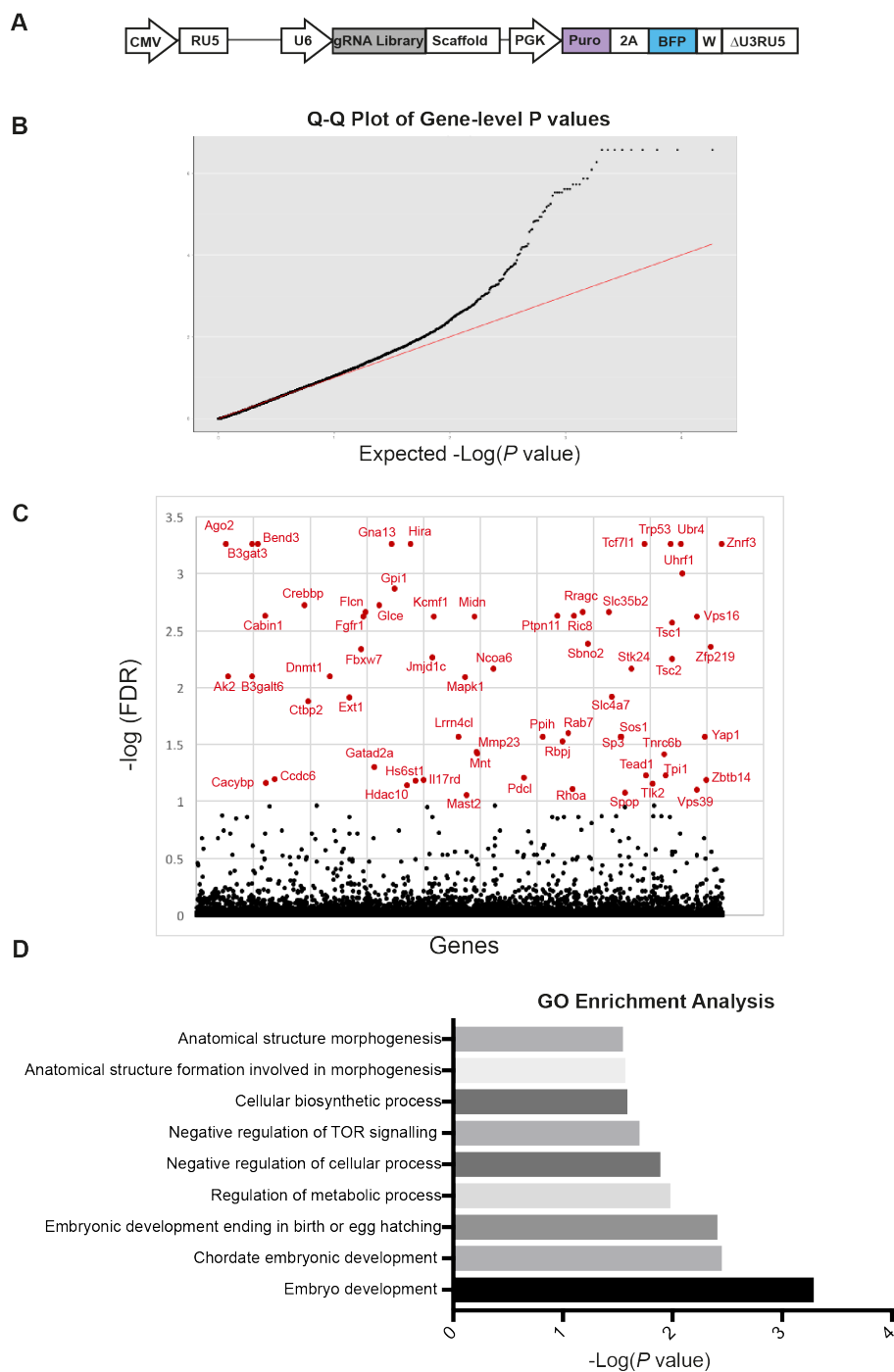
A Quantile-Quantile (Q-Q) plot was generated using the gene level  $P$ -values calculated by MAGeCK (Figure 3.7 (B)). The distribution of data points followed a curved pattern with an increasing slope that diverged from the expected  $P$ -value distribution, suggesting that a large number of genes showed statistical significance. Such distribution was expected based on the results of similar screens done in the past. The exit of pluripotency screen performed by Betschinger et al. has identified 70 genes that passed stringent  $Z$  score cut-off [19]. Two of other similar screens carried out by Yang et al. and Leeb et al. have revealed 272 and 113 hits with high confidence [217] [464]. Although only part of these candidate lists was validated, it suggests that there are many genes involved in the process of differentiation initiation. Therefore, a highly skewed  $P$ -value distribution with a large group of genes showing some degree of statistical significance would be expected. As mentioned previously, the FGF/ERK pathway and WNT/GSK3 pathway play crucial roles in the exit from pluripotency. Reassuringly, the preliminary screen identified mutants from both pathways such as *Fgfr1*, *Mapk1* from the FGF/ERK pathway, and *Tcf7l1*, *Crebbp* and *Ctbp1* from the WNT/GSK3 pathway (Figure 3.7 (C)). The recovered hits also include genes that were identified and validated from previous screens, such as *Flcn*, *Tsc2*, and *Hira*. Screening candidates that passed the statistical cut-off were analysed

for enrichment of Gene Ontology (GO) terms. The results demonstrated a significant enrichment in terms related to embryo development and regulation of gene expression, which corroborated the hyperactive transcription status of the transition period from pluripotency to differentiation (Figure 3.7 (D)).

### 3.3 Discussion and Conclusion

This chapter has described the preparation work and proof-of-principle studies for the set-up of the genome-wide CRISPR-Cas9-based exit of pluripotency screen. In verifying the knockout phenotype of *Tcf7l1* and *Apc* in the Rex1:GFPd2 Rosa26:Cas9 cell line, I was ready to perform the screen.

The proof-of-principle studies have provided valuable insights into screening design. The rapid downregulation of *Rex1* expression suggests that the dissolution of pluripotency takes place very soon after inhibitor withdrawal. To capture this fast-happening event, I reasoned that it is better to terminate the screen on day two or day three of differentiation. The Rex1GFP flow profile of *Tcf7l1* and *Apc* knockouts also suggests that the most drastic difference between knockout and wild type control occurs on day 2 or day 3. A few screens conducted by other groups have adopted the strategy of several rounds of replating and enrichment, which may not be necessary for CRISPR-Cas9-mediated screen. Given the high efficiency of gRNA and its convenience in mutant identification, the CRISPR-Cas9-mediated screen is highly sensitive to detect subtle changes in gRNA representation. Furthermore, because the gRNA counts are available which allows statistical analysis, the assessment of a specific phenotype is no longer ‘black and white’, but can be evaluated in a quantitative way. Another concern over several rounds of prolonged enrichment is the depletion of fitness genes or essential genes, which are required for ESC survival but may also be involved in pluripotency regulation. Therefore, I decided to perform a screen as a short-course one-round of differentiation, after which cells showing differentiation defects will be collected and sequenced.



**Figure 3.7:** Preliminary screen result analysis. (A) RNA library expression vector. CMV: CMV promoter; RU5: 5' long terminal repeat; U6: U6 promoter; gRNA library: gRNA sequence from the mouse V2 library; scaffold: gRNA scaffold; PGK: mouse phosphoglycerate kinase 1 promoter; Puro: puromycin resistant gene; BFP: blue fluorescent protein; 2A, *Thoesa asigna* virus 2A peptides; W: Woodchuck Hepatitis Virus posttranscriptional regulatory element;  $\Delta$ U3RU5: self-inactivating 3' long terminal repeat. (B) Gene-level  $P$ -values Quantile-Quantile (Q-Q) plot. Gene level  $P$ -values were calculated by the published algorithm MAGeCK. The  $P$ -values were sorted in ascending order, and then plotted versus quantiles calculated from a theoretical distribution. A 95% confidence interval was used. A 45-degree reference line (red) was also plotted. Points would fall along the reference line if the observed  $P$ -values were randomly distributed. The Q-Q plot was generated using the `gglot2` package from software R. (C) Gene-level false discovery rate (FDR). The FDR values were calculated by MAGeCK. X-axes represents genes ranked in alphabetical order. Y-axes represents  $-\log(\text{FDR})$ . Genes with  $\text{FDR} < 0.1$  were labelled red. (D) Gene ontology term analysis of statistically significant hits. Cutoff was defined as  $\text{FDR} < 0.1$ . The analysis was performed using the online GO analysis algorithm PANTHER (<http://pantherdb.org/>).

The preliminary screen reassured screening strategy, helped to study the scale-up effect and provided material for a data analysis run-through. An important parameter to be decided is the statistical cut-off, which is essential to reduce the number of false positives and restrict the scale of subsequent validation. Setting the cut-off at  $P$ -value  $< 0.05$  resulted in 1095 significant candidates, which is likely to include high proportion of false positives, and were obviously impractical for further validation purposes. Also, because the screening results were generated from a genome-wide study, where thousands of hypothesis tests were conducted simultaneously, small  $P$ -values may occur by chance. It is thus more accurate to use the false discovery rate (FDR) to control false positives. To set a decent threshold, the enrichment fold change in the control experiment can be used as a reference, where *Tcf7l1* and *Apc* KO cells were enriched in the Rex1GFP positive population by 4.9 times and 4.3 times respectively after three days of differentiation. Assuming the selection criteria were based on  $FDR < 0.1$ , the shortlist of potential candidate genes were reduced to 64. *Mast2*, the candidate gene at the cut off point, has an average of 1.8 times and 3.9 times enrichment in each replicate at the gRNA level, which is slightly lower compare to the enrichment effect in the positive control. However, *Tcf7l1* and *Apc* are genes with relatively strong phenotypes. Therefore, to reduce the chance of having false negatives, the cut-off should be selected to allow the test to pick up potential candidates with weaker phenotypes. Hence,  $FDR < 0.1$  would be an appropriate threshold to start with.

## Chapter 4

# Screening Result and Pathway Analysis



## 4.1 Introduction

Several large-scale RNAi-based or transposon-based loss-of-function screens have been carried out in mouse ESCs to identify genes required for maintenance of pluripotency and initiation of differentiation. Consistent with prior knowledge, these screens recovered core pluripotency genes such as *Oct4*, *Sox2* and *Nanog*, as well as components from the WNT and FGF/MAPK signalling pathways. Beyond these well-studied pathways, a number of other candidate effectors were also identified and several have been validated and investigated in detail. For example, the RNA binding protein Pum1 was identified in a *piggyBac*-mediated haploid ESC screen [217]. It was shown that Pum1 binds to the 3' untranslated regions of mRNA for major naive pluripotency factors such as *Esrrb*, *Klf2*, *Tbx3* and *Tfcp2l1*. In the absence of Pum1, those factors were upregulated and showed sustained expression upon 2i withdrawal, suggesting that Pum1 may act as a circuit limiter that potentially constrains the self-renewal machinery. Tumour suppressor Folliculin, along with its interaction partners *Fnip1* and *Fnip2*, were identified in an siRNA screen [19]. It was demonstrated that Folliculin and *Fnip1*/*Fnip2* drive differentiation by restricting nuclear localisation of the bHLH transcription factor *Tfe3*. Other candidates recovered include mediator-cohesin complexes, Paf1 complex, protein kinase C, the SWI/SNF chromatin remodeling complex. For reference, I have compiled a table of previously published pluripotency-related screens in mouse ESCs (Table:4.1).

Despite these intriguing findings, screens mentioned above were based on siRNA knock-down or *piggyBac*-mediated insertional mutations, and hence were limited by the intrinsic technical defects of mutagenesis methods. For instance, large-scale mutagenesis mediated by the transposon-based gene trap system is a relatively random process, meaning that the experimenters have no control over the integration sites and genes to be knocked out. Such characteristics render the discovery process a partial 'black box', which is unfavourable to investigators. Furthermore, there is an unavoidable bias in the probability of knocking out different genes, given the vast range of the sizes of each gene in the genome. Similarly, RNAi-based screening is often haunted by off-target effects and incomplete knockdown penetrance, which is likely to result in poor reproducibility. For example, similar studies performed by Hu et al. and Ding et al. showed little consistency, despite using the same cell line and selection marker [162] [90]. Out of 149 and 186 candidate genes ( $Zscore > 2$ ) found by each screen, only five were overlapping. Additionally, contradictory results were observed between siRNA screens. For instance, 14% showed the opposite phenotype in

the screen conducted by Ding et al [90] [19].

The development of CRISPR-Cas9 genome-wide library provides an opportunity to study the exit of pluripotency in an unprecedented depth given its high knockout efficiency and accuracy. In this chapter, I describe the screen I performed using a genome-wide gRNA library to dissect the dissolution of the naive pluripotency program. I also describe the screening results, as well as pathways recovered in the screen.

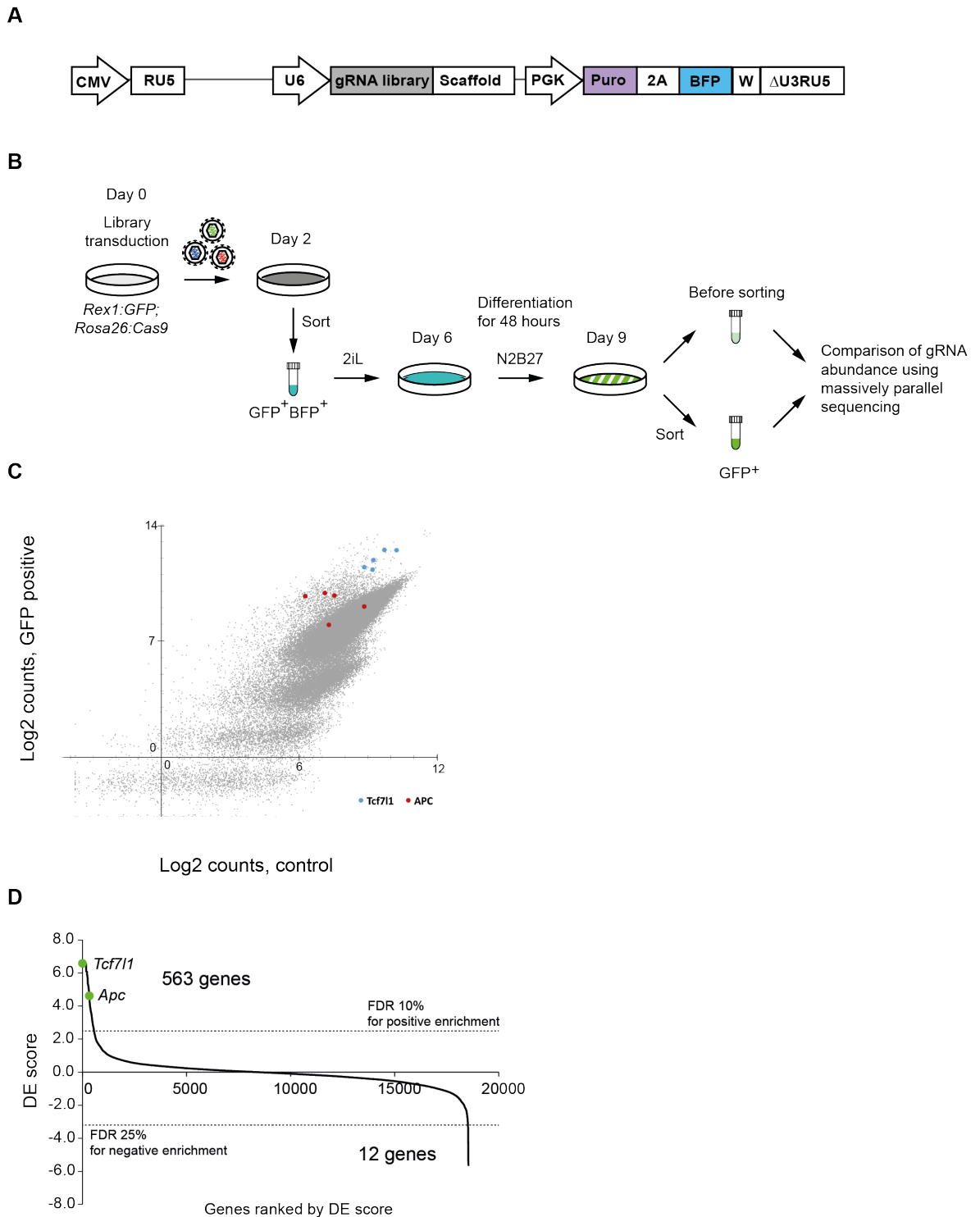
Table 4.1: Published pluripotency-related screens using mESC.

Format	Screen type	Selection	Scale	Major finding	References
Arrayed	siRNA	ESC maintenance; Colony morphology	Focused library; 1008 genes	Tip60-p400 is essential for ESC maintenance	[110]
Arrayed	siRNA	ESC maintenance; Oct4 expression	Genome-wide; 16,683 genes	Transcription regulators Cnot3 and Trim28 are essential for ESC self-renewal	[162]
Arrayed	siRNA	ESC maintenance; Oct4 expression	Genome-wide; 25,057 siRNAs	Paf1 complex plays an important role in maintaining ESC identity	[90]
Arrayed	shRNA	Retinoid acid differentiation; Nanog expression	Focused library; 312 genes	SWI/SNF complex plays an important role in facilitating mESC differentiation	[360]
Pooled	<i>piggyBac</i> ; <i>Blim</i> -deficient ESCs	Differentiation upon 2i removal; Rex1 expression	Genome-wide	Tcf3 is a potent negative regulator of pluripotency	[142]
Pooled	shRNA	Differentiation upon LIF removal; Colony outgrowth	Genome-wide; 16,000 siRNAs	Mp1 is required for mESC differentiation	[441]
Arrayed	siRNA	Differentiation upon 2i removal; Rex1 expression	Genome-wide; 16,873 siRNAs	MAP kinase phosphatase plays an important role in early embryonic stem cell fate decisions	[464]
Arrayed	siRNA	Differentiation upon 2i removal; Oct4 expression	Genome-wide; 9,900 siRNAs	Foliculin regulates mESC pluripotency via regulating subcellular localisation of Tfe3	[19]
Pooled	<i>piggyBac</i> ; haploid ESCs	Differentiation upon 2i removal; Rex1 expression; Colony formation	Genome-wide	Zfp706 and Pum1 are essential for the dissolution of mESC self-renewal circuitry	[217]
Pooled	siRNA	RA Differentiation; Nanog expression	Genome-wide; 16784 genes	Shai1 and Snai2 interplay with Nanog in the transcriptional regulation of pluripotency-associated genes	[137]

## 4.2 Results

### 4.2.1 Screening Strategy and Result Analysis

To produce a library of mutagenised ESCs, I transduced the Cas9-expressing Rex1:GFPd2 cells with the genome-wide Mouse V2 lentiviral gRNA library generated in the lab (Figure 4.1(A)(B)). The library comprises 90,230 gRNAs targeting a total of 18,424 genes [421]. Compared to the previous version of the mouse gRNA library, the performance of the V2 library has been improved by removing gRNAs with disfavoured nucleotide composition and using an optimised gRNA scaffold [421]. These improvements have largely increased the gene knockout efficiency and overall screening sensitivity, therefore laying the groundwork for a successful robust screen. To ensure a sufficient coverage of gRNA library complexity, approximately 32 million cells were used in transduction. The amount of library virus used was carefully titrated to achieve transduction efficiency between 25% to 30%. Virus infection events follow Poisson distribution, by achieving 25% to 30% transduction efficiency, the majority of the transduced cells were infected by one viral particle. This led to a total of about 8 million transduced cells, which corresponds to approximately 90X gRNA library coverage. The transduced cells were collected via cell sorting two days after infection, followed by four days of expansion under 2i/LIF condition. For induction of differentiation, the expanded cells were plated in N2B27 medium at a density of  $1 \times 10^4$  per  $cm^2$  without LIF or 2i for 48 hours before being lifted for cell sorting. These conditions are highly permissive for differentiation induced by autocrine signals. As shown in Chapter 3, the majority of the cells were differentiated and lost Rex1 expression after 48 hours upon 2i removal. Cells with persistent *Rex1* expression after 48 hours were potential mutants, whose differentiation has been interrupted by gene perturbation. Approximately 2-3 million Rex1GFP positive cells were collected from FACS sorting, followed by genomic DNA extraction. About 20 million unsorted cells were collected as a control. The screen was performed in four biological replicates.



**Figure 4.1:** Screening Strategy and Result Analysis. (A) gRNA library expression vector. CMV: CMV promoter; RU5: 5' long terminal repeat; U6: U6 promoter; gRNA library: gRNA sequence from the mouse V2 library; scaffold: gRNA scaffold; PGK: mouse phosphoglycerate kinase 1 promoter; Puro: puromycin resistant gene; BFP: blue fluorescent protein; 2A, *Thossea asigna* virus 2A peptides; W: Woodchuck Hepatitis Virus posttranscriptional regulatory element;  $\Delta$ U3RU5: self-inactivating 3' long terminal repeat. (B) Outline of the screening strategy. (C) Scatter plot comparing the gRNA abundance in Rex1GFP positive population and in unsorted control. gRNA counts were median-normalised by the MAGeCK algorithm. The mean value was taken from four biological replicates. (D) Overall summary of the screening results. Genes were ranked by DE (Depletion/Enrichment) score. The DE score was computed by taking the sum of the  $\log_{10}$  (Depletion  $P$  value) and the negative  $\log_{10}$  (Enrichment  $P$  value)

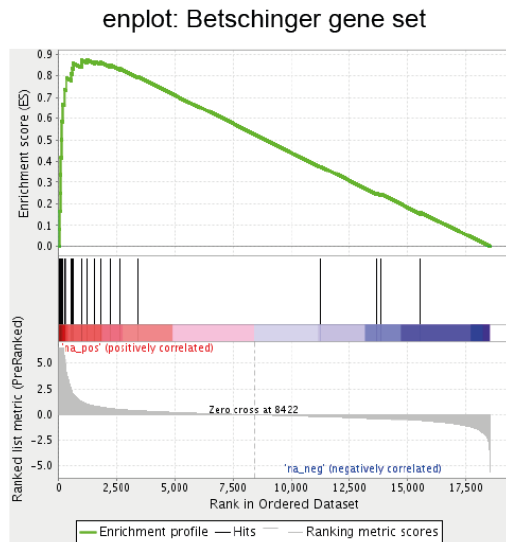
Genomic DNA from both sorted and control samples was extracted. gRNA sequences were amplified by PCR and sent for deep sequencing. The gRNA abundance was analysed using the statistical algorithm MAGeCK [225]. Figure 4.1(C) demonstrated a direct comparison of gRNA representation between control and Rex1-GFP positive populations. As expected, the majority of the gRNAs showed good concordance between two datasets. A group of gRNAs were over-represented in the Rex1GFP positive population, meaning those knockouts exhibited a delayed differentiation phenotype. Among them there were the two positive control genes *Tcf7l1* and *Apc*. Noticeably, two gRNAs targeting *Apc* did not show the expected phenotype. Nonetheless, *Apc* came up as a significant candidate at gene level. This observation suggested the potential existence of false negatives due to the presence of nonfunctional gRNAs. Such errors could be controlled to a certain extent by including multiple gRNAs and applying suitable statistical analysis. To summarise the screening results, genes were ranked by Depletion/Enrichment (DE) score, which was computed as:  $\log_{10}(\text{Depletion } P \text{ value}) + [-\log_{10}(\text{Enrichment } P \text{ value})]$  (Figure 4.1(D)). By applying a 10% false discovery rate (FDR) cut-off, 563 genes, whose mutant showed delayed differentiation, could be identified. Much fewer genes (12 genes) could be identified from the depletion side of the screen, even with a more relaxed FDR cut-off. This is probably because a much higher sensitivity is required for the detection of dropouts in a population. The gRNA scatter plot also revealed that a number of gRNAs were depleted by about four to eight fold in the Rex1GFP population (Figure 4.1(C)). However, because the read counts for those gRNAs were relatively low, the variance between each gRNA is inevitably enlarged, and therefore less likely to be calculated as statistically significant. This can be improved by optimising the design of the screen by changing the timeline in particular when to collect for cell sorting, or by choosing another selection marker gene such as *Sox1* or *Fgf5*.

#### 4.2.2 Screening Result Validation

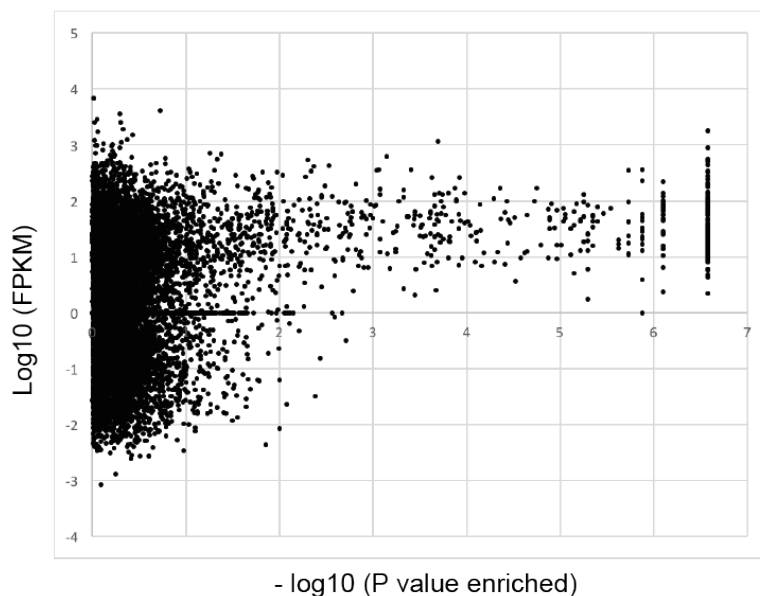
The identification of *Tcf7l1* and *Apc* confirmed the reliability of the screening results. For further validation, I carried out a Gene Set Enrichment Analysis (GSEA) using a set of genes whose knockdown has been shown to delay differentiation in a study performed by Betschinger et al. [19]. Genes present in the gRNA library were pre-ranked according to DE score. The siRNA validated gene set was highly skewed towards the enrichment side, indicating high consistency with the screening result (Figure 4.2(A)). In addition, it was observed that large numbers of gRNAs enriched in the Rex1GFP positive population were

targeting genes that were already expressed in the naive pluripotent state (fragments per kilobase of transcript per million mapped reads [FPKM] $>0.5$ ) (Figure 4.2(B)). This observation suggested that naive pluripotency is actively constrained by an ongoing network in the cell and the initiation of differentiation is triggered mainly by the dismantling of the pluripotency network rather than the induction of lineage specific genes.

A



B



**Figure 4.2:** Screening Result Validation. (A) Gene Set Enrichment Analysis (GSEA) comparing the screening result with a validated gene list [19]. The validated gene set includes *Apc*, *Hgs*, *Mapksp1*, *Tcf7l1*, *Flcn*, *Tsc2*, *Kras*, *Zfp281*, *L3mbtl3*, *Trp53*, *Rnf2*, *Josd1*, *Hand1*, *Ctbp2*, *Sox2*, *Relt*, *Smg1*, *Ewsr1*, *Klhl23*, *E2f4*, *Upf1*, *Csde1*, *Trrap*, *Raf1*, *Taf6l*, *Arid1a*, *Nedd8* and *Nf2* [19]. (B) Comparison between gene expression and statistical significance of enrichment.

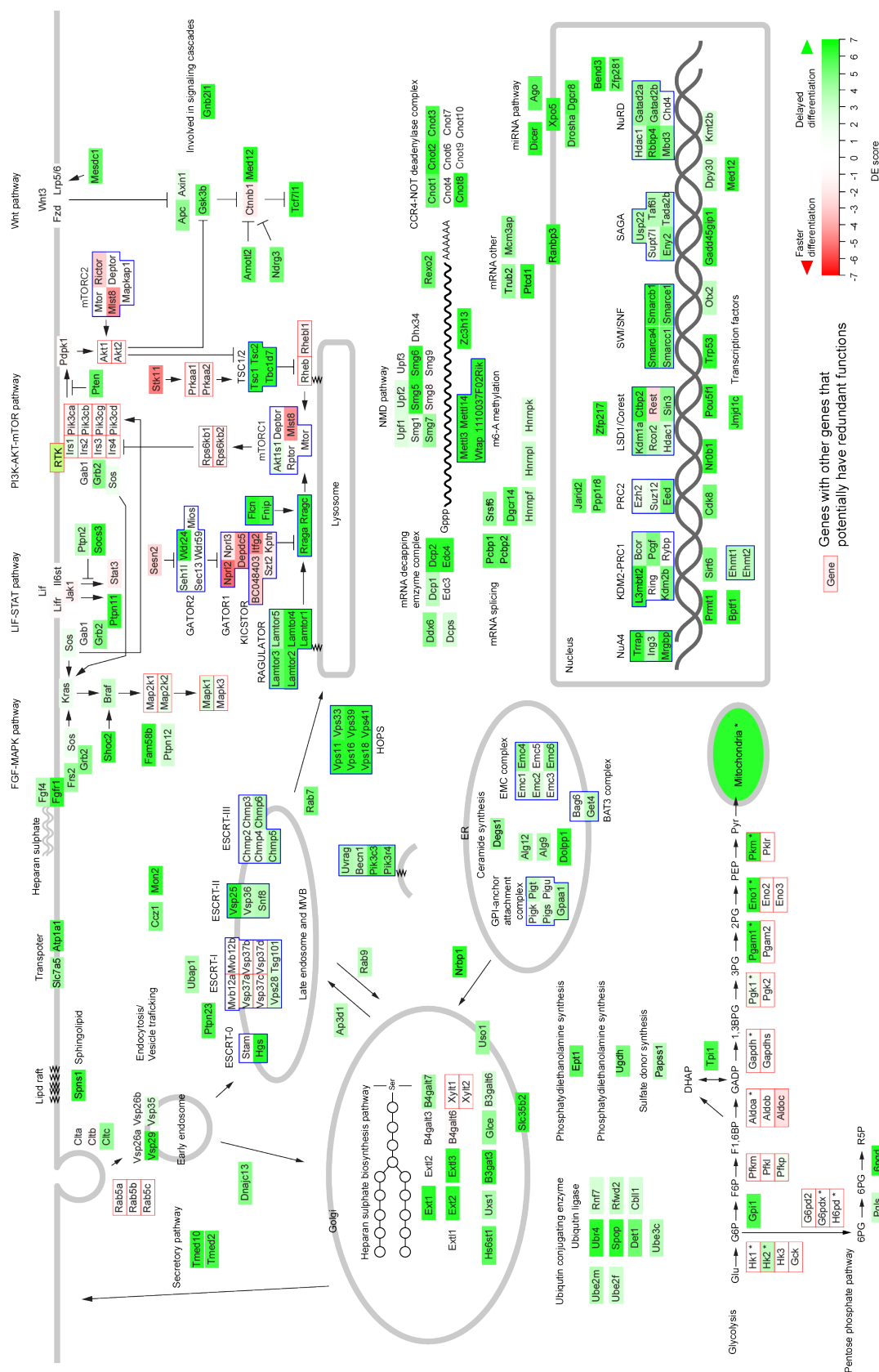
### 4.2.3 Pathways Analysis

#### 4.2.3.1 Signalling pathways

As discussed in the first chapter, abundant evidence has indicated that autocrine FGF-mediated ERK activation is one of the key triggers for mouse ESC differentiation. Indeed, the screen identified almost all the key elements of the FGF/MAPK pathway, ranging from Fgf4, Fgf receptor 1, docking/scaffold proteins such as Frs2 and Grb2, tyrosine phosphatase Shp2, Ras and Raf. Components that could not be identified in the screen were probably due to the presence of functionally redundant isoforms, for instance MEK1 and MEK2; which are encoded by *Map2k1* and *Map2k2*, respectively. Remarkably, pathways indirectly related to the FGF/MAPK pathway, such as the heparan sulphate biosynthesis pathway, have also been identified. It has been reported that heparan sulfate is required for the stable expression of Fgfr on the cell surface, binding of Fgf to Fgfr as well as its internalisation [325] [227]. Cells defective in heparan sulfate synthesis will have reduced Fgf signalling transduction, thus are deficient in response to differentiation cues.

Several key genes downstream of Wnt, which is another pathway implicated in self-renewal and differentiation, were identified in the screen. Examples include subunits of the degradation complex as well as its functional effector Tcf3. Last but not least, the screen revealed a profound role of mTORC1 in the regulation of pluripotency and differentiation, as knockout of many key components of the mTORC1-centred pathways led to an aberrant differentiation phenotype. Interestingly, opposite phenotypes were observed during differentiation from knockouts of TSC1/2 and GATOR1 complexes, although both function as mTORC1 inhibitors. To uncover the unknown mechanisms behind mTORC1-centred pluripotency regulation, I decided to focus on TSC1/2 and GATOR1 for further investigation, which will be discussed in the following chapters.





**Figure 4.3:** Cellular pathways underpinning the initiation of differentiation. Pathways were manually curated according to statistically significant screening hits. Green-coloured label: Knockouts that exhibited delayed differentiation phenotype; Red-coloured label: knockouts that differentiate in an accelerated fashion. Genes in red rectangles may have functional redundancy in the genome. Mitochondria-related genes were not illustrated in the diagram.

### 4.2.3.2 mRNA related pathways

The mRNA post-transcriptional modification and degradation appeared to be another important module that regulates the resolution of naive pluripotency. Pathways enriched included the mRNA nonsense-mediated decay (NMD), m<sup>6</sup>A mRNA methylation, deadenylation and decapping, alternative splicing, as well as the microRNA pathway.

Some of them have been reported in the literature in relation to ESC maintenance and differentiation, such as pathways regulating mRNA stability and degradation. Wang and colleagues have shown that depleting major NMD factors compromises differentiation and proposed the role of NMD as a licensing pathway for ESC differentiation [224]. Recently, Geula et al. identified *Mettl3*, an N<sup>6</sup>A-methyladenosine (m<sup>6</sup>A) transferase, as a regulator which facilitates the termination of naive pluripotency [132]. Consistent with Geula's finding, *Mettl3* appeared to be one of the top hits in the screen, together with *Mettl14*, *Wtap* and *1110037F02Rik*, which collectively form a complex with *Mettl3* and mediates the m<sup>6</sup>A deposition on nuclear RNA. Among the top hits, there were several components from the CCR4-NOT complex. As a multi-functional complex, it is involved in transcription regulation, protein modification, and deadenylation of mRNA. Although it has been demonstrated that CCR4-NOT positively regulates planarian stem cell differentiation, the molecular mechanism of its role in mouse ESCs remains to be investigated. Given the above, it was not surprising to identify the mRNA decapping associated genes, which also regulates the mRNA stability, and work cooperatively with pathways such as NMD.

mRNA splicing serves as another important regulatory step that contributes to the functional repertoire of different cell types and ESC differentiation [256]. Ng and colleagues have identified the spliceosome-associated factor SON as a regulator of pluripotency in hESCs [237]. Lu et al. also demonstrated that the splicing factor SFRS2 stabilises pluripotency by regulating alternative splicing of the methyl-CpG binding protein MDB2 [238]. *Pcbp*, also known as heterogeneous ribonucleoproteins (hnRNPs), are RNA binding proteins that were reported to serve as splice enhancers in erythrocyte differentiation [173]. An *in vivo* study conducted by Ghanem et al. confirmed that *Pcbp1* and *Pcbp2* are independently important for murine embryogenesis and fetal survival, though the molecular mechanism remains unclear [134]. Several genes encoding heterogeneous ribonucleoprotein particles, which are important for mRNA splicing and maturation, appeared in the screen to have a relatively mild effect on the delay of differentiation. However, elucidating the detailed mechanism of their function requires further mechanistic investigation.

### 4.2.3.3 Chromatin modifiers

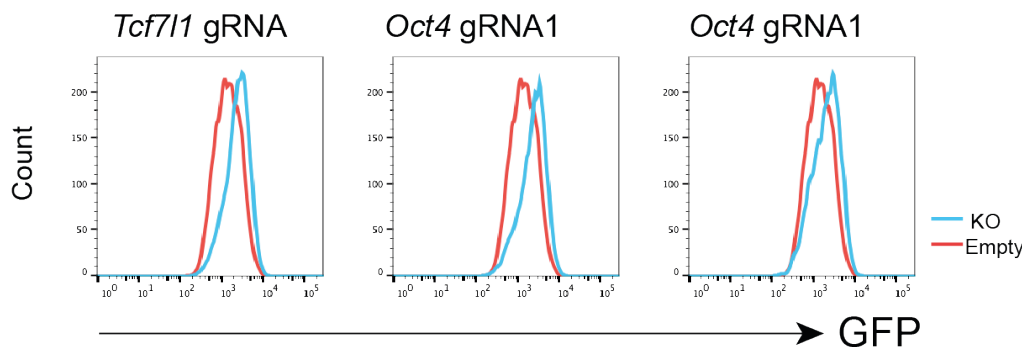
The chromatin environment and epigenetic landscape are essential for the correct preservation of ESC identity, as well as lineage specification during the transition towards differentiation. Screening identified a number of chromatin modifiers, which confirmed the prominent role of epigenetic regulation in embryonic development. One of the most well-studied epigenetic regulators in ESCs is the Polycomb group proteins (PcG), which can be further divided into two major classes, namely Polycomb repressive complex 1 (PRC1) and Polycomb repressive complex 2 (PRC2) [266]. The core PRC2 complex contains Eed, Suz12 and the histone methyltransferase Ezh1/2 – all were identified in the screen, although Ezh2 demonstrated a slightly higher FDR value. A number of PRC2 interaction partners was also identified, including Jarid2, Hira and Nipp1. Conversely, only Pcgf1 from canonical PRC1 was significantly enriched in the Rex1GFP positive population. Knockout of Ring1B showed moderate phenotype but did not pass the threshold of 10% FDR. Instead, non-canonical PRC1 partners such as Kdm2b and L3mbtl2 were identified from the screen. Interestingly, knockout of Cbx7 was significantly depleted, indicating an accelerated differentiation profile. This evidence suggests that canonical and non-canonical PRC1 might play different roles in the dissolution of pluripotency. Moreover, validation of screening result as well as further molecular investigation is required to draw solid conclusions.

Another renowned category of pluripotency-related epigenetic regulators is the HDAC1- and HDAC2-containing complexes, which carry out chromatin-mediated transcription repression via histone deacetylation or other chromatin-modifying depending on its subunits. The most well-studied HDAC1/2-containing complexes are SIN3, NuRD, LSD1 and CoREST, which were all present in the screening output. Notably, almost all of the NuRD complex subunits were identified in the screen. Consistent with the screening result, it has been reported that the deletion of Mbd3, an essential subunit, which maintains the integrity of NuRD, led to an aberrantly stabilised pluripotency status in ESCs and failure in lineage commitment [183]. Reynolds et al. also demonstrated that loss of Mbd3 results in upregulation of pluripotency related genes such as *Zfp42*, *Tbx3*, *Klf4* and *Klf5* [332]. Collectively, the findings above suggested that the MBD3/NuRD complex deacetylates histone lysine residues at pluripotency genes and contribute to the formation of a differentiation permissive status. Similarly, it has been reported that LSD1, a histone H3K4/K9 demethylase, is required for the repression of pluripotency genes by decommissioning ESC-specific enhancers [442].

In addition to PcGs and HDAC1/2-containing complexes, a set of other chromatin regulatory proteins have arisen from the screen. For example, the SAGA complex is a highly conserved multisubunit complex, which mainly acts as a co-activator [66]. Genes identified in the screen cover almost all of its functional multisubunits, such as the histone acetyltransferase (*Tada2b* and *Supt7l*), transcription factor (*Taf6l*) and histone deubiquitinases (*Eny2* and *Usp22*). Although it has been reported that Usp22 regulates several pluripotency factors, including c-Myc and Sox2 [484] [397], the function of other SAGA subunits remain unclear. It has also been demonstrated that the chromatin remodeling complex esBAF is required for the repression of *Nanog* and other pluripotency-related genes upon differentiation [360]. Consistent with this finding, four of the esBAF subunits were identified from the screen. Another nucleosome remodelling complex, NURF, was found to negatively regulate differentiation. One of its component, Bptf, has been reported to be essential for the activation of genes required for the development of three germ layers. Three subunits of the histone acetyltransferase NuA4 complex, namely Trrap, Ing3 and Mrgbp, were ranked within the top range of the enriched genes in the Rex1GFP-positive population. An RNAi screen revealed that certain subunits of HAT including Trrap are involved in the maintenance of ESC identity [110]. Sawan et al. also reported that Trrap is essential for ESC self-renewal and differentiation restriction [356]. However, in my screen, Trrap knockout showed endured expression of *Rex1* upon differentiation, indicating a more persistent pluripotency network and self-renewal ability. Such discrepancy might be explained by the difference in the assessment of pluripotency, or the technologies used, although validation and further investigation needs to be carried out for a better understanding. The role of other NuA4 complex subunits ING3 and MRGBP has not been well-studied in the stem cell context.

#### 4.2.3.4 Transcription factors and other transcriptional regulatory proteins

In addition to chromatin modifiers, the screen also identified transcription factors and other transcriptional regulatory proteins. Among them were well-studied differentiation initiation regulators such as Oct4, Otx2 and Zfp281. Oct4 was previously implicated in early differentiation because cells with reduced Oct4 expression exhibited enhanced self-renewal and delayed differentiation kinetics [328] [188]. This has been validated with single gRNA-mediated knockout ESCs as shown in Figure 4.4. Otx2 was reported to recruit Oct4 to enhancers that are associated with genes induced during differentiation [463] [45]. Zfp281 was identified as a repressor, which down-regulates expression of many pluripotency genes including *Nanog*. Therefore, deletion of Zfp281 resulted in stabilised pluripotency status and compromised differentiation. In addition, Tgif1 and Nr0b1 were also reported to counterbalance the activities of core pluripotency factors [216] [396].



**Figure 4.4:** Rex1GFP differentiation profile of Oct4 knockout mESCs. ESCs were transduced with lentivirus expressing gRNAs targeting Oct4. Differentiation was induced by 2i/LIF removal. Rex1GFP was measured at 29 hours differentiation. Blue - Rex1:GFP cells transduced with gRNAs targeting indicated genes; Red - Rex1:GFP cells transduced with gRNA free vectors.

Several genes that have not been studied in ESCs demonstrated relatively strong delayed differentiation phenotype during the screen. For instance, *Cdk8* and *Med12*, which encode subunits of the Cdk8 mediator complex, were both identified as significantly enriched genes from the screen. *Cdk8* has been reported as part of the canonical Wnt/ $\beta$ -catenin pathway that drives colon tumorigenesis and transformation [115] [199]. However,  $\beta$ -catenin regulates transcription rather differently in colon cancer cells and ESCs. In the proposed model in colon cancer cells, *Cdk8* positively regulates the expression of  $\beta$ -catenin-driven downstream targets, which cannot explain its knockout phenotype observed in ESCs. Therefore, there is likely to be an unknown molecular mechanism behind the *Cdk8* mediator complex's role in regulating pluripotency that worth further investigation. Examples of other understudied genes include *Gadd45gip1*, which is negatively regulated by NAC1 [175], and

Zfp161, which is a transcription activator/repressor. Both of them showed high statistical confidence but have not been previously related to pluripotency and differentiation. Bend3 was another gene showing a strong knockout phenotype. As it has been reported to actively recruit NuRD complex to hypomethylated DNA regions [344], its knockout phenotype could be the consequence of insufficient downregulation of pluripotency genes, in line with that of the Mbd3 knockout. Finally, there were genes whose knockout generated a relatively weak phenotype, albeit statistically significant. Examples of those genes include *Foxi3* and *Foxd3* in the Forkhead box transcription factor family, *Hinfp*, *Ddx5* and *Tfdp1*. Those genes might be directly or indirectly involved in pluripotency regulation, the mechanisms may be worth exploring but at low priority.

#### 4.2.3.5 Mitochondria-related pathways

Remarkably, about half of the candidate genes enriched in the Rex1GFP positive population were mitochondria-related. For instance, almost all (more than 90%) of the mitochondrial ribosomal protein (MRP) genes, have an FDR value less than 0.2, indicating the indispensable requirement of an integrate healthy mitochondria population in the early differentiation stage. Similarly, GSEA revealed enriched distribution of mitochondria morphogenesis genes as well as oxidative phosphorylation genes, further confirming the functional importance of mitochondria and its related pathways (Figure 4.5 (A)(B)).

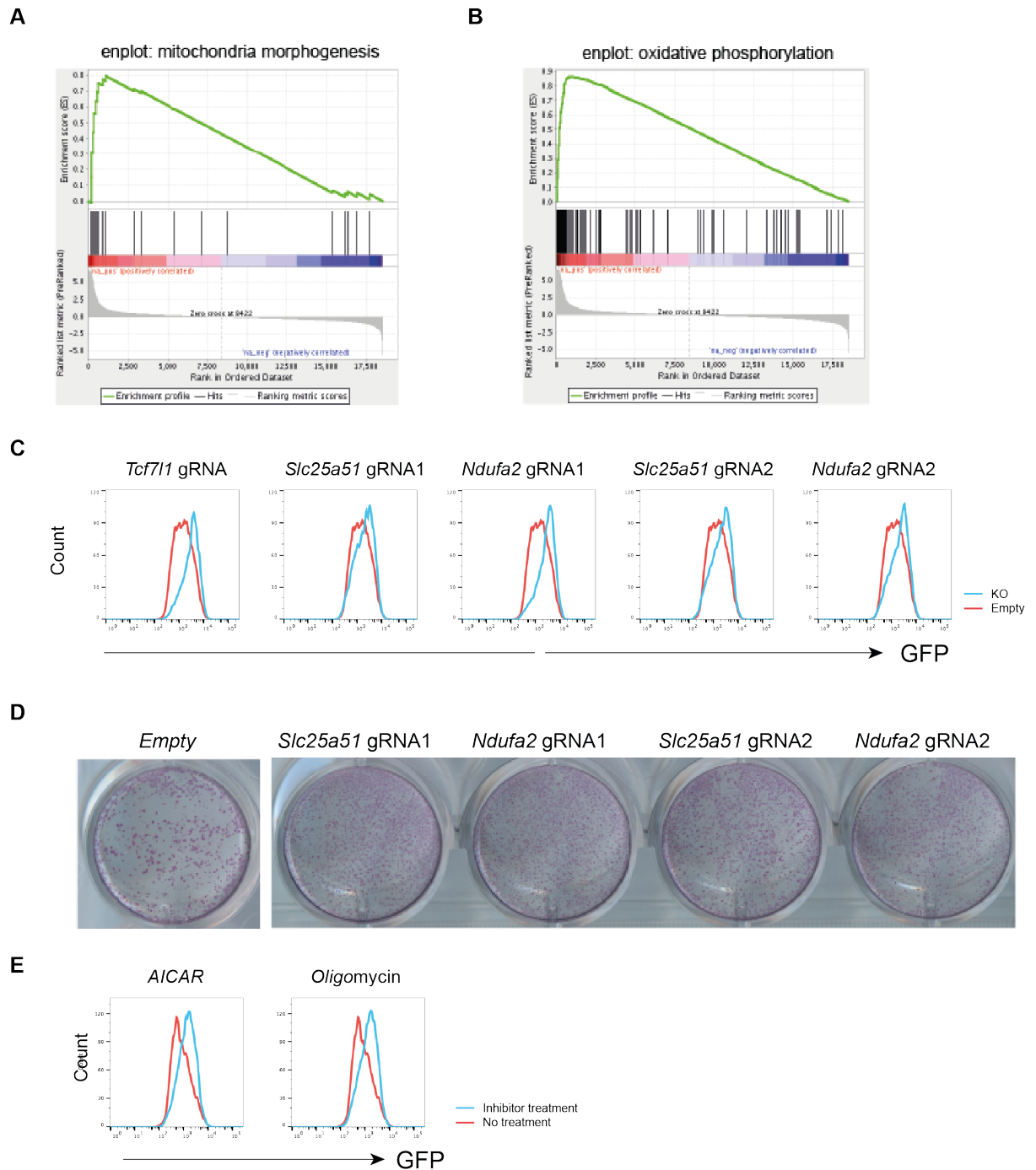
To validate the above, I selected two top mitochondria-related hits, namely *Slc25a51* and *Ndufa2*, for further analysis. Slc25a51 is a carrier protein located in the inner membrane of the mitochondria, while the Ndufa2 protein is a subunit of the NADH dehydrogenase, whose knockout interrupts the electron transport chain. To investigate their function in pluripotency and differentiation, I used two gRNAs, which target two different exons, for each gene. Consistent with the screening result, the RexGFP expression profile of *Slc25a51* and *Ndufa2* knockout clearly demonstrated a delayed differentiation pattern compared to that of the empty vector infected ESCs (Figure 4.5 (C)). However, as GFP degradation is an ATP-dependent process, the Rex1GFP flow profile might not be faithfully reflective to the pluripotency state of the cell with defective mitochondria function. To rule out such possibility, I performed colony forming assay with cells transduced with lentivirus expressing gRNA targeting *Slc25a51* and *Ndufa2*. Twenty four hours after 2i withdrawal, single cells were plated into 2i/LIF at clonal density, allowed to proliferate for seven days, and stained by alkaline phosphatase. The number of colonies obtained represents the proportion of naive pluripotent cells in the culture after 24 hours, as 2i/LIF is a

highly selective culture condition in which cells that have exited naive pluripotency fail to proliferate and subsequently die. Consistent with the Rex1GFP flow profile, *Slc25a51* and *Ndufa2* knockout ESCs generated significantly more AP positive colonies compared to cells infected with a control virus, confirming that the phenotype observed was truly biologically related rather than a result of lagging GFP degradation (Figure 4.5 (D)). Notably, the average colony size of mitochondria-related knockouts is much smaller, suggesting growth defects caused by deficiency in energy supply. These results indicate that mitochondria play an important role in the initiation of differentiation.

#### 4.2.3.6 Endocytosis and vesicle trafficking

Interestingly, genes related to the endocytosis and vesicle trafficking pathway appeared to stand out from the screening candidates. The function of those genes in relation to pluripotency and differentiation has not been well studied so far. However, existing evidence suggests that the endosomal trafficking pathway indirectly regulates embryonic development as part of the signalling transduction cascade, particularly the FGF/MAPK pathway. Upon binding to Fgf, the activated Fgfrs undergo clathrin mediated endocytosis and subsequent trafficking to a series of intracellular compartments [25]. The signalling cascade persists along the endocytic pathway until the internalised FGFR complex are sorted into the degradative compartment, where proteolysis takes place and signal transduction is terminated. The endosomal trafficking pathway controls the intensity, duration and specificity of the signalling pathways. Therefore, any related defects compromise appropriate FGF/MAPK signalling and result in persistent expression of pluripotency factors such as *Rex1* [5] [98].

In addition, it was reported that the endocytic pathway, especially the late endosome is essential for mTORC1 signalling, due to the fact that the interaction and binding of mTORC1 with its activator Rheb occurs on late endosomes [117]. The screen identified a group of candidates required for the early/late endosome conversion with relatively high confidence, such as Rab7 and subunits of the HOPS complex. This observation hints that the endocytic pathway could be involved in differentiation regulation through the mTORC1 signalling. Future work is required to interrogate this hypothesis and investigate the underlying molecular mechanism.



**Figure 4.5:** Validation of genes in mitochondria-related pathways. (A) and (B) Gene Set Enrichment Analysis using predefined datasets: mitochondria morphogenesis gene set (A) and oxidative phosphorylation gene set (B). Screening hits were pre-ranked according to the Depletion/Enrichment (DE) score, which was computed as:  $\log_{10}$  (Depletion  $P$  value) +  $[-\log_{10}$  (Enrichment  $P$  value)]. The gene sets were downloaded from the Molecular Signatures Database (MSigDB). The GSEA analysis was performed using the online algorithm 'GSEAPreranked' developed and maintained by the Broad Institute. (C) Rex1GFP differentiation profile of mitochondria-related gene knockouts. Differentiation was induced by 2i/LIF removal. Rex1GFP was measured at 29 hours differentiation. Blue - Rex1:GFP cells transduced with gRNAs targeting indicated genes; Red - Rex1:GFP cells transduced with gRNA free vectors. (D) Commitment assay after mitochondria-related gene knockout. Cells were kept in N2B27 medium before being replated in N2B27 supplemented with 2i/LIF. Alkaline phosphatase (AP) staining was used to visualise ESC colonies one week after replating.



## 4.2.4 Discussion and Conclusion

### 4.2.4.1 Functional genomics with CRISPR-Cas9

The developmental progression from naive pluripotency entails complex regulation mechanisms. Given the success of a few large-scale screens, a group of proteins have been implicated in the exit of pluripotency, among which some have been investigated in detail. However, those identified factors were often analysed as individual cases. For a better understanding of this complicated transition process, we need a comprehensive dissection of cellular pathways in an unbiased manner. The results described in this Chapter have demonstrated the utility of CRISPR-Cas9 genome-wide screen in addressing this need and identify molecular drivers and facilitators of the transition from pluripotency to differentiation.

By applying a statistical cut-off of FDR  $<0.1$ , 563 genes whose mutants produced a delayed differentiation phenotype could be identified. Among were well-studied genes and pathways such as the FGF/MAPK pathway, Wnt signalling pathway, /NuRD complex and OTX2. Such observation demonstrated great consistency with prior knowledge, which provided confidence in the screening results and confirmed the powerfulness of CRISPR-Cas9-mediated genetic screening. It is noteworthy that Oct4 came up in the screen as a factor that controls the dismantling of pluripotency and lineage commitment. This has been confirmed and investigated in detail by two independent studies conducted by Karwacki-Neisius et al. and Radzishewska et al., but was missed in all the previous screens [188] [328]. Being able to capture Oct4, CRISPR-Cas9-mediated screen has demonstrated unprecedented sensitivity. In addition to the identification of genes whose knockout hampered the priming and competence of differentiation, the screen could also detect genes whose knockout tipped the balance towards differentiation. In other words, with carefully optimised screening design, one can assess whether a perturbation facilitates or inhibits a biological process and look at the objective from two opposite angles. The candidates from the enriched and depleted ends should reassure and compensate for each other, thus facilitating the development of a complete perspective.

#### 4.2.4.2 Integration of the regulatory network

The screen generated a panoramic picture of pathways involved in the dissolution of naive pluripotency. As discussed previously, each pathway plays an important role in safeguarding the ability to exit pluripotency. This intricate pathway map indicated that instead of working independently, these pathways intercalate with each other and exist in a dynamic equilibrium. Any disturbance to this equilibrium will cause the balance to collapse and results in hampered pluripotency or aberrant differentiation phenotype. In particular, it was noticed that the initiation of differentiation is predominantly under the control of the repression machineries in the cell, which govern the shut-down of the pluripotency network prior to any lineage specification process.

LIF/inhibitor removal results in loss of positive pluripotency input, upon which the repressive functions of various regulators dominate, downregulating the expression of pluripotency genes. Perhaps the most immediate repression comes from the transcription repressors that directly counterbalance and restrict the activities of pluripotency factors. Examples from the screening result include *Zfp281*, *Tgif1*, *Nr0b1* and the most well-studied transcription factor, *Tcf3*. Being the main downstream effector of WNT signalling in ESC differentiation, *Tcf3* co-localises with *Oct4* and directly represses *Klf4*, *Nanog* and *Esrrb* [316] [468] [259]. Another layer of repression effect comes from the epigenetic machineries, such as PcGs, LSD1 and NuRD, which act to repress pluripotency gene expression via histone deacetylation, deposition of repressive histone modifications such as H3K27me3 and H2AK119ub, permissive histone methylation removal, as well as histone remodelling. Finally, the dissolution of pluripotency is safeguarded by post-transcriptional regulation. Cells deficient in mRNA decay and stability regulation are resilient to differentiation, probably because the pluripotency-related transcripts need to be degraded rapidly for differentiation to take place. These pathways co-exist and cross-regulate each other, in order to cooperatively achieve the dismantling of pluripotency.

#### 4.2.4.3 Metabolic regulation

It has been reported in mice and humans that there is a reduction in mitochondrial respiration when ESCs transit from naive to primed pluripotency [489] [143] [402]. In particular, mouse ESCs are metabolically bivalent with both glycolysis and mitochondrial respiration in place, whereas primed cells, such as EpiSC or human ESC as well as epiblast in the post-implantation embryo, have low mitochondrial respiratory capacity, despite having a more developed mitochondrial content [489]. Kallan et al. have recently reported an overall decrease in transcripts encoding components of mitochondrial respiratory complexes after 25 hours differentiation [184]. These studies demonstrate that oxidative phosphorylation is rather trivial in regards to energy production in primed pluripotent stem cells. However, despite this, healthy mitochondria seem to be required for cells to be able to enter the differentiation stage as mouse ESCs with malfunctioning mitochondria failed to differentiate at a normal pace. Therefore, it can be postulated that mitochondria biosynthesis and maturation is one of the prerequisites of the initiation of differentiation, even though they remain dormant in the next stage of development.

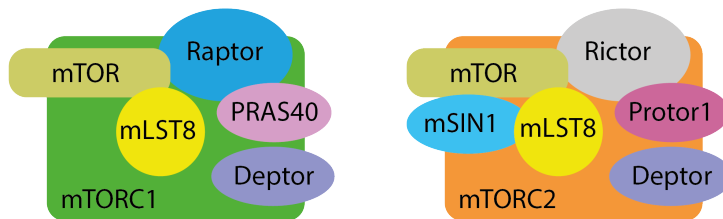
Notably, screening results revealed that several genes from the glycolysis pathway also showed a delayed differentiation phenotype when knocked out, indicating the indispensable role of glycolysis in the early developmental transition. Given that glycolysis is another major energy generation pathway and serves as one of the initial steps in oxidative phosphorylation, this leads us to the hypothesis that the dismantling of the pluripotency network as well as the changes in the transcriptome and methylome at the onset of differentiation creates an energy barrier which a cell needs to overcome to complete the transition. Furthermore, it has been reported that metabolism integrates with genetic and epigenetic programs to regulate stem cell function and fate [482]. For instance, acetyl-CoA generated by fatty acid metabolism, glycolysis and in mitochondria can be transported to the nucleus for histone acetylation [440] [48]. Therefore, undisrupted energy supply is a crucial requirement for the initiation of differentiation. Further investigation is worth carrying out to understand the mechanism of how metabolic state affects ESC identity, function and lineage determination.

## Chapter 5

# The Role of mTOR-related Pathways in Pluripotency/Differentiation Regulation

## 5.1 Introduction

Mammalian target of rapamycin (mTOR) is a serine/threonine kinase discovered in the 1990s as the direct target of rapamycin-FKBP12 complex in mammals [40] [341]. It is a crucial signalling node that integrates both intracellular and extracellular signals to regulate various cellular processes such as cell metabolism, growth, proliferation and survival. The mTOR protein functions in two distinct complexes, mTOR complex 1 (mTORC1) and mTOR complex 2 (mTORC2) [141]. mTORC1 comprises five components: mTOR, Raptor, mLST8, PRAS40 and Deptor [194] [196] [150] (Figure 5.1). It has been shown that Raptor facilitates the recruitment downstream targets to mTORC1 [300] [359]. PRAS40 and Deptor physically associate with mTORC1 and negatively regulate its kinase activity [319] [347]. Upon activation, mTORC1 directly phosphorylates Deptor and PRAS40, resulting in reduced physical association and full activation of mTORC1. mLST8 acts to stabilise the kinase activation loop of mTORC1, although genetic studies suggested that it may be dispensable for mTORC1 activity [461] [124]. mTORC2 is made of six components, some of them are common to mTORC1, such as mTOR, mLST8 and Deptor [124]. The mTORC2-specific subunits are mSIN1, Rictor and Protor1 [124] (Figure 5.1). Unlike the dispensable role in mTORC1, mLST8 is functionally essential for mTORC2 [124]. Rictor and mSIN1 stabilise each other and establish the structural foundation of mTORC2 [215]. Similar to its role in mTORC1, Deptor negatively regulates mTORC2 activity [319]. mTORC2 is characterised for its insensitivity to rapamycin treatment; however, it was demonstrated that prolonged rapamycin treatment abrogates mTORC2 signalling [212] [351]. Compared to mTORC1, mTORC2-related biology is less characterised.

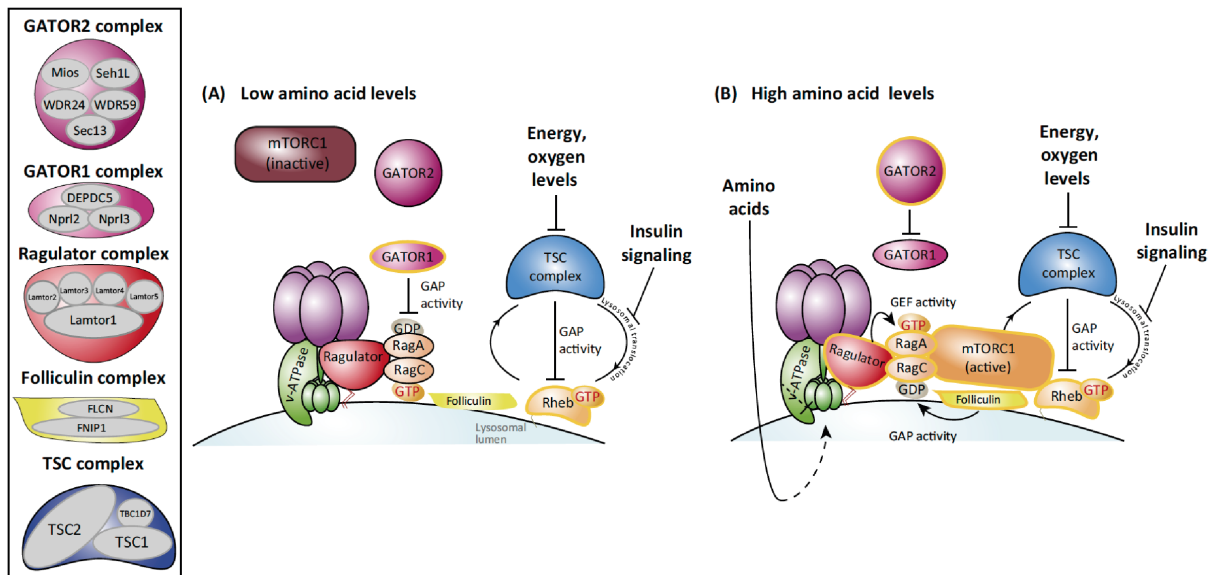


**Figure 5.1:** Structure of mTORC1 and mTORC2. mTORC1 consists of three core components: mTOR, Raptor and mLST8. It also contains two inhibitory subunits PRAS40 and Deptor. Like mTORC1, mTORC2 also contains mTOR, Deptor and mLST8. Components unique to mTORC2 are Rictor, mSin1 and Protor1.

mTORC1 is a central regulator for cell growth and metabolism. It positively regulates cell proliferation by promoting anabolic processes such as protein and lipid biosynthesis [215]. S6K1 and 4EBP1 are two of its key downstream effectors [215]. 4EBP inhibits translation through binding eIF4E to prevent the formation of eIF4F. mTORC1 phosphorylates 4EBP, which leads to its dissociation from eIF4E, allowing the cap-dependent translation to take place [44] [138]. mTORC1 also directly phosphorylates and activates S6K1, which in turn phosphorylates and activates several substrates that promotes ribosome synthesis, such as S6 and eIF4B [95] [159]. Furthermore, activated S6K promotes lipid and nucleotides synthesis via phosphorylating and activating Srebp and Cad, respectively [320] [15]. In addition to activating anabolic pathways, mTORC1 suppresses protein catabolism, such as autophagy [215]. mTORC1 was shown to phosphorylate Ulk1, which prevents it to form the autophagy driver complex with Atg13, Fip2000 and Atg101 [196]. Furthermore, a recent study demonstrated that mTORC1 regulates protein turnover by inhibiting protein ubiquitination or reducing the amount of proteasomal chaperones by inhibition of Erk5 [338].

mTORC1 integrates major metabolism related signals such as growth factors, amino acids, oxygen and energy levels to regulate downstream processes [215]. Growth factors activate mTORC1 mainly via stimulating PI(3)K-Akt pathway. Binding of insulin or other growth factors such as Igf-1 and Egf to their cell surface receptors promotes the tyrosine kinase activity of the receptors and leads to the recruitment of Irs1 [152]. Irs1 in turn activates PI(3)K and produces PtdIns(3,4,5)P3. A key downstream effector of PtdIns(3,4,5)P3 is Akt, which is recruited to the cell membrane and activated by Pdk1 [152]. Once activated, Akt moves to the cytoplasm, where it phosphorylates and inhibits the tuberous sclerosis complex (TSC) [165] [253]. TSC is a heterotrimeric complex comprising Tsc1, Tsc2 and Tbc1d7 [87]. It functions as a GTPase-activating protein (GAP) for the small Ras-related GTPase (Rheb), which directly interacts with mTORC1 and stimulates its activity via unknown mechanism [408] [235]. Notably, S6k1, activated by mTORC1, phosphorylates Irs1 and attenuates signal transduction from receptor to PI(3)K [252]. This S6K1-dependent negative feedback loop has been shown to be deregulated in many diseases including cancer and metabolic diseases [252]. The TSC complex is also involved in response to changes in energy status and oxygen levels. Under energy deficit condition, AMPK, a major energy sensor, is activated and positively regulates Tsc2, which leads to the reduction of mTORC1 activity [166]. In response to hypoxia, Redd1 is activated and negatively regulates mTORC1 by disrupting the association of Tsc2 and 14-3-3 protein

[41] [85].



**Figure 5.2:** Structure of mTOR regulating complexes and mechanistic target of mTORC1 amino acid sensing pathway. (A) Under amino acid deprived condition, Ragulator complex and V-ATPase remain in an inhibitory state. GATOR1 complex exerts GTPase-activating protein (GAP) activity towards RagA, keeping as inactive from recruiting mTORC1 to the lysosomal membrane. TSC complex is translocated to the lysosomal surface where it inhibits Rheb through its GAP activity. (B) Upon amino acid stimulation, Ragulator and V-ATPase undergo conformational changes. Together with folliculin, Ragulator and V-ATPase catalyse the activation of Rags complex, which recruits mTORC1 to the lysosomal membrane where it interacts with Rheb. Akt phosphorylation of TSC2 drives the TSC complex off the lysosomal surface, allowing Rheb to activate mTORC1. Image taken and adapted from Bar-Peled and Sabatini, 2014 [11].

The amount of amino acids is another strong signal that regulates the mTORC1 pathway (Figure 5.2). It was shown that the activation of mTORC1 by amino acids is independent of TSC complex, as S6K1 phosphorylation is protected from amino acid withdrawal in *Tsc2*-knockdown cells [299]. Amino acid sensing by mTORC1 largely relies on the heterodimeric Rag GTPases, which consist of RagA or RagB with RagC or RagD [194] [346]. The Rags complex is located on the lysosomal membrane through association with the Regulator complex, which comprises five subunits from Lamtor1 to Lamtor5 [345] [346]. Amino acid stimulation activates the Rags complex, which allows it to recruit mTORC1 to the lysosomal surface via binding of Raptor. Once recruited to the lysosomal surface, mTORC1 interacts with Rheb and becomes activated [346]. It was shown that mTORC1 senses cytosolic and lysosomal amino acids through different mechanisms. Lysosomal amino acids activate mTORC1 through transporter SLC38A9, which interacts with the Rag-Ragulator-v-ATPase complex and promotes the activation of the Rag complex [180]

[330] [432]. Cytosolic amino acids signal to mTORC1 through GATOR1 and GATOR2 complexes [10]. GATOR1 is a GAP for RagA/B and acts as an indirect inhibitor of mTORC1 [10]. GATOR2 positively regulates mTORC1 via inhibiting GATOR1 [314]. Recently, a complex named KICSTOR was identified as another negative regulator that binds and recruits GATOR1 to the lysosomal membrane [448]. Sestrin2 has been identified as the primary leucine sensor for mTORC1. In the absence of leucine, Sestrin2 binds and inhibits GATOR2, which activates GATOR1, leading to mTORC1 downregulation. Another amino acid sensor is CASTOR1 (Cellular Arginine Sensor for mTORC1) [358]. Similar to Sestrin2, CASTOR1 binds and inhibits GATOR2 in the absence of arginine and dissociates upon arginine binding to enable the activation of mTORC1 [60].

mTOR deregulation is associated with a range of human diseases, such as cancer, obesity, neurodegeneration and type 2 diabetes. Although it has been shown that mTOR pathway is required in pluripotency and development, its exact function remains largely undefined. Recently, Bulut-Karslioglu et al. showed that inhibition of mTOR induces reversible pausing of mouse blastocyst development and allows prolonged culture *ex vivo*, indicating that mTOR regulates developmental timing at the peri-implantation stage [46]. Sabatini and colleagues reported that abrogation of mTORC1, by genetic loss of Rag GTPase, leads to embryonic lethality, indicating the essentiality of mTORC1 in embryonic development [101]. From a large-scale siRNA screen, Betschinger et al. found that knockdown of the mTOR regulators *Folliculin* (*Flcn*) and *Tsc2* resulted in delayed differentiation [19]. The authors also showed that Folliculin, together with its binding partners Fnip1 and Fnip2 drives differentiation by regulating the subcellular location of the transcription factor Tfe3. Genome-wide location and functional analysis showed that Tfe3 promotes the pluripotency circuitry through up-regulation of *Esrrb*.

Consistent with Betschinger's siRNA screening result, *Flcn* and *Tsc2* were identified in the CRISPR-Cas9 genome-wide screen as statistically significant genes whose loss impeded differentiation initiation. Additionally, a number of other mTORC1/2 regulators were found to be essential for the onset of differentiation, including RagA and RagC from the Rags complex, Wdr24 from GATOR2 complex and all the 5 subunits of Regulator complex. Furthermore, many of the negative regulators or components of the mTORC1/2 complexes were identified as genes whose loss accelerate mESC differentiation. This include Nprl2 and Depdc5 from GATOR1 complex, Itfg2 from KICSTOR complex, and Mlst8, which is a common component of mTORC1/2 complex. Sestrin2 and Rictor were also detected, albeit at a less significant level. One of the surprising findings from the

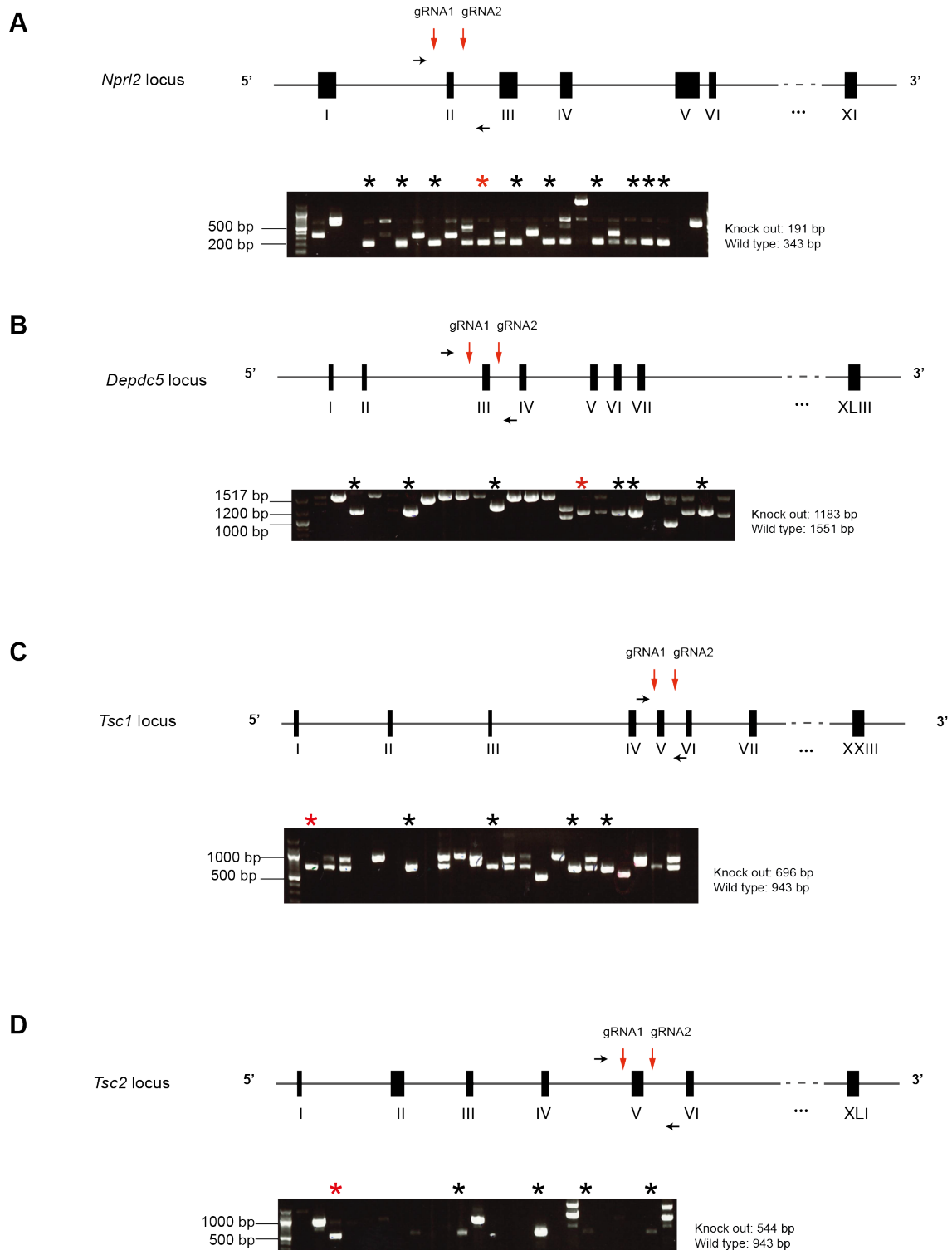


screen is that knockouts of mTORC1 regulators from amino acid sensing and growth factor stimulation pathways led to opposite outcomes on ESC differentiation. From the amino acid sensing aspect, mTORC1 upregulation accelerated ESC differentiation, as seen in the phenotype of GATOR1 knockout. In contrast, knockout of TSC complex, the key factor mediating the growth factor stimulation pathway, resulted in delayed differentiation phenotype. These results suggested unknown mechanisms in relation to mTORC1/2-mediated pluripotency and differentiation regulation. The following experiments in this Chapter aimed to investigate the molecular mechanisms underlying mTORC1/2-mediated pluripotency/differentiation regulation. Firstly, I verified the knockout phenotypes of GATOR1 and TSC complexes. Secondly, I compared the knockout effect of GATOR1 and TSC complexes on upstream regulators and downstream effectors of mTORC1. And finally, I analysed the transcriptomic changes of GATOR1 and TSC knockouts.

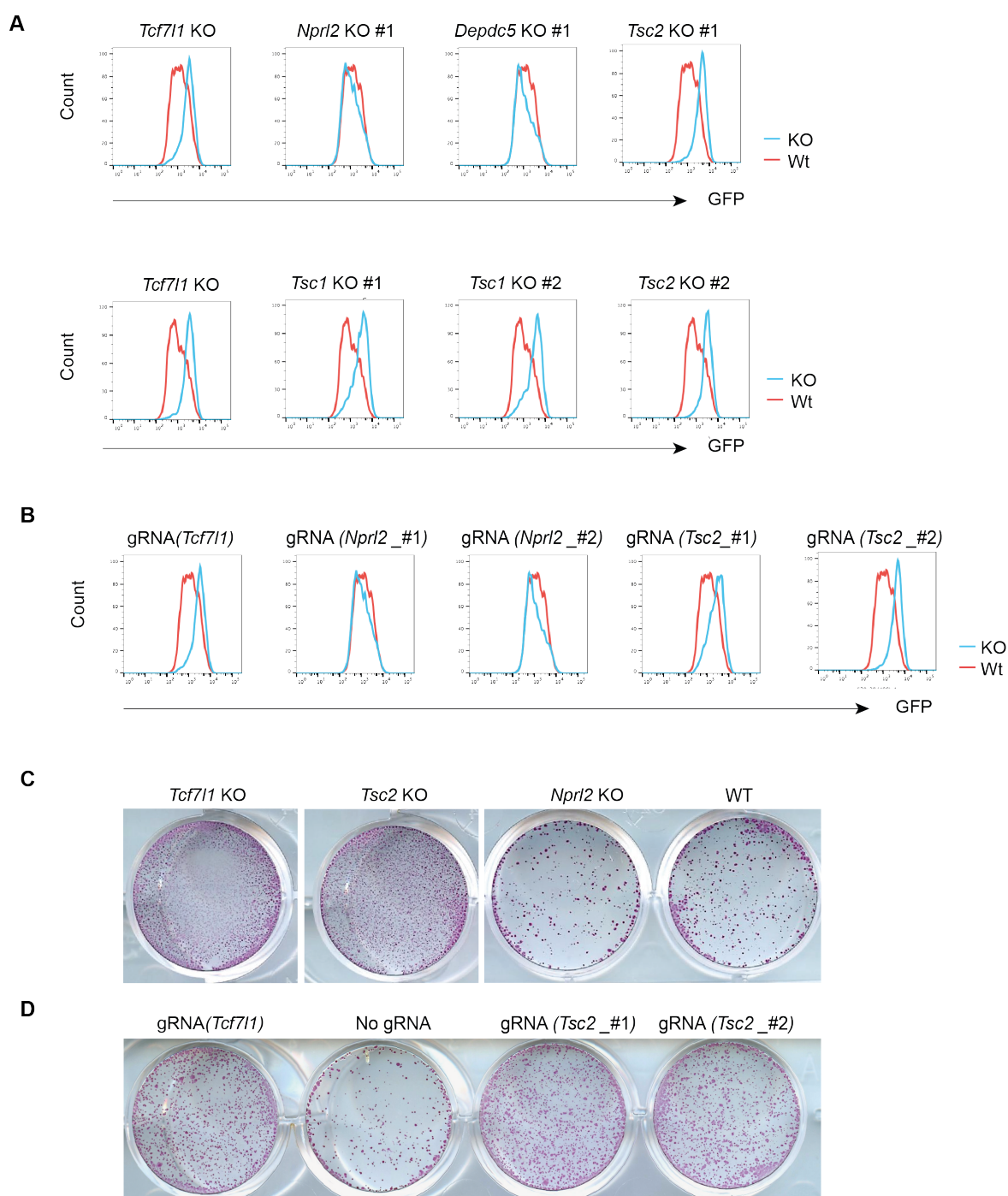
## 5.2 Results

### 5.2.1 Generation of knockout cell lines and phenotype validation

To decipher the role of mTORC1 in pluripotency and differentiation regulation, I decided to focus on the two mTORC1 inhibitory complexes: GATOR1 complex and TSC complex, whose knockouts exhibited completely different phenotypes upon differentiation from my screen. I first generated knockout ESC lines for GATOR1 components *Nprl2* and *Depdc5*, which is directly inhibited by GATOR1, as well as the TSC complex subunits *Tsc1* and *Tsc2* (Figure 5.3). Same targeting strategy was applied as discussed in Chapter 3, where *Tcf7l1* and *Apc* knockouts were described. The knockout cell lines were subsequently assessed for differentiation ability by plating in N2B27 for 27 hours followed by flow cytometry analysis. As shown in Figure 5.4 (A), phenotypes observed in each knockout cell lines agreed with screening results: *Nprl2* and *Depdc5* knockout cell lines lost Rex1GFP expression faster than parental cell line, whereas *Tsc1* and *Tsc2* knockout cell lines retained greater RexGFP positive population. For further validation, the Rex1GFPd2 cells were transduced with lentivirus expressing gRNAs targeting *Nprl2* and *Tsc2*, and plated in differentiation condition. As shown in Figure 5.4 (B), the observed single gRNA knockout phenotypes were consistent with stable knockout cell lines. To assess the differentiation progress at the cellular level, I replated cells at a low density into 2i/LIF medium after 28-hours differentiation in N2B27, and cultured them for a week before AP staining. As shown in Figure 5.4 (C), the number of AP-positive colonies was well correlated with the Rex1-GFP profiles. Thus, the screening results were confirmed with individually targeted knockout clones. Before any further investigation, it was important to make sure that the knockout phenotypes were not specific to the genetic background of E14 cells, which is the parental cell line which Rex1GFPd2 cell line was established from. To test that, I performed commitment assay using the Cas9-expressing JM8 cells transduced with gRNAs targeting *Tsc2*. As shown in Figure 5.4 (D), *Tsc2* knockout JM8 cells exhibited significantly more colonies compared to wild type, suggesting that the phenotype of delayed differentiation was not related to genetic background.



**Figure 5.3:** Generation of stable KO cell lines for mTORC1 regulating genes. (A) to (D) Schematic targeting strategy and genotyping results for *Tsc2*, *Tsc1*, *Nprl2* and *Depdc5* KO cell lines. Two gRNAs were designed to flank a critical exon and transfected into Rex1:GFPd2 cell line. A critical exon was defined as a common exon expressed in all transcript variants and when deleted, creates a frame-shift mutation. Transfected cells were plated in clonal density and single cell colonies were picked and genotyped by PCR. Black arrows indicate PCR genotyping primers. Red arrows indicate gRNA cutting sites. gRNA and PCR primer sequences are listed in Chapter 2 Table:2.1 and Table:2.2. Asterisk indicates KO clones. Red asterisk indicates KO clones used in analysis. Roman numbers indicate exon numbers.



**Figure 5.4:** mTOR-related gene knockout phenotype validation. (A) Rex1GFP differentiation profile of indicated stable knockout ESC clones (knockout strategy shown in (Figure 5.1)). Differentiation was induced by 2i and LIF removal. Rex1GFP was measured at 27 hours differentiation. Blue - knockout Rex1:GFP clones; Red - wt Rex1:GFP cells. (B) Rex1GFP differentiation profile of cells expressed with gRNAs targeting indicated genes. Differentiation was induced by 2i and LIF removal. Rex1GFP was measured at 27 hours differentiation. Blue -Rex1:GFP cells transduced with indicated gRNAs; Red -Rex1:GFP cells transduced with empty vector. (C) Commitment assay of stable knockout clones. Cells were kept in N2B27 medium before being replated in N2B27 supplemented with 2i/LIF. Alkaline phosphatase (AP) staining was used to visualise ESC colonies one week after replating. (D) Commitment assay using Cas9-expressing JM8 cells. Cells were transduced with Lentiviral vectors expressing indicated gRNAs.

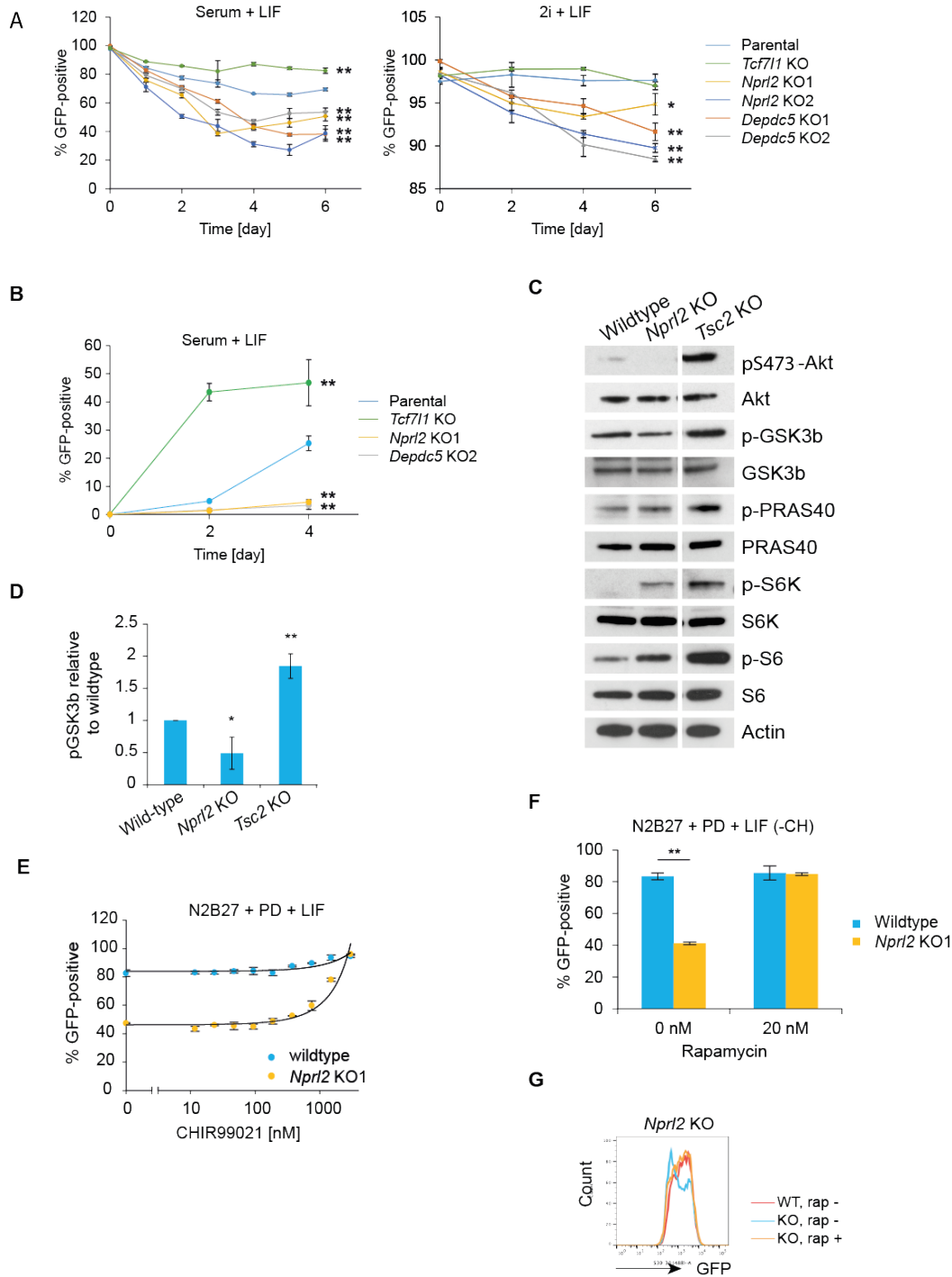
## 5.2.2 Disruption of GATOR1 complex induces mTORC1-activated negative feedback loop

As discussed previously, mESCs cultured in serum/LIF condition exhibit heterogeneous expression of naive pluripotency markers. These two subpopulations exist in a dynamic equilibrium and can be re-established in a few passages after purifying either population by cell sorting [57]. To investigate the self-renewal ability of GATOR1 knockout cells, I sorted Rex1-GFP positive population and plated them back to serum/LIF and monitored the percentage of GFP positive population for six days. As shown in Figure 5.5 (A), wild type cells gradually lost GFP positive population for about four days until a plateau was reached at around 75%. *Tcf7l1* knockout cells hardly lost any GFP positive population: around 90% cells were expressing Rex1 at day six, indicating its enhanced self-renewal ability. In contrast, both *Nprl2* and *Depdc5* knockout clones lost Rex1-GFP expression at a much faster rate than wild type cells, and plateaued at around 40% to 50% at day six, which is significantly lower than wild type cells. Strikingly, GATOR1 knockout cells struggled to retain GFP positive population even in 2i/LIF condition. They lost 5-10% of GFP-positive cells in 6 days, whilst wild type and *Tcf7l1* knockout cells retained high GFP-positive fraction. Furthermore, the Rex-GFP negative population of GATOR1-deficient cells failed to re-establish the GFP-positive naive pluripotent cells under serum/LIF condition, whereas GFP-negative wild type and *Tcf7l1* knockout cells could re-acquire Rex1 expression to approximately 30% and 50%, respectively, in four days (Figure 5.5 (B)). These observations indicate compromised self-renewal ability in GATOR1 knockout cells.

I next analysed the knockout effect of *Nprl2* and *Tsc2* on other key regulators upstream and downstream of mTORC1 by focusing on their phosphorylation status (Figure 5.5 (C)). In line with previous findings, *Nprl2* and *Tsc2* knockout cells exhibited higher phosphorylation of S6K and S6, indicating increased mTORC1 activity [165] [10]. It is known that activation of mTORC1 induces negative feedback loop through phosphorylation of Irs1 by activated S6K, which results in reduced signal transduction from receptor to Akt [252]. This is well-correlated with the phenotype of *Nprl2* knockout, where phosphorylation of Akt on S473 is considerably weaker compared to that in wild type, suggesting down-regulated Akt activity as expected. Surprisingly, *Tsc2* knockout cells exhibited strongly up-regulation of phosphorylation at S473, indicating higher Akt activity. I then moved on to analyse the downstream targets of Akt such as GSK3 and PRAS40. Consistent with Akt S473 phosphorylation, PRAS40 was highly phosphorylated in *Tsc2* knockout cells, but *Nprl2* knockout cells showed no obvious difference from wild type cells. The

phosphorylation pattern of GSK3 in each knockout cell line is well correlated with Akt phosphorylation pattern, as well as the observed differentiation phenotypes (Figure 5.5 (D)). Because GSK3 is an important node of the Wnt signalling pathway in regulating pluripotency and differentiation, I hypothesised that the opposite differentiation phenotypes of GATOR1 knockout and TSC knockout cells were caused by the difference in GSK3 regulation.

For further investigation of the GSK3 involvement in *Nprl2* knockout phenotype, *Nprl2* knockout cells were plated in 'li' condition (N2B27 with Mek inhibitor PD0325901 and LIF) supplemented with serially diluted GSK3 inhibitor CHIR99021 (Figure 5.5 (E)) (This experiment was performed by Kosuke Yusa). In the complete absence of CHIR99021, wild type cells retained around 80% Rex1-GFP positive population, whereas *Nprl2* knockout cells has only approximately 45% Rex1-GFP positive cells. Furthermore, *Nprl2* knockouts appeared to be more responsive to the changes of CHIR99021 concentration and more dependent on CHIR99021 to maintain Rex1-GFP expression, in line with the elevated Gsk3 activity indicated by western blot. Since it was speculated that the elevation of Gsk3 activity in *Nprl2* knockout cells is most likely through the negative feedback mechanism by hyperactive mTORC1 and the resulted negative feedback loop, inhibition of mTORC1 would reverse the phenotype. Indeed, as shown in Figure 5.5 (F) and (G), treating *Nprl2* knockout cells with 20nM rapamycin restored the ability of pluripotency maintenance in li/LIF (This experiment was performed by Kosuke Yusa), and rescued the accelerated differentiation phenotype.



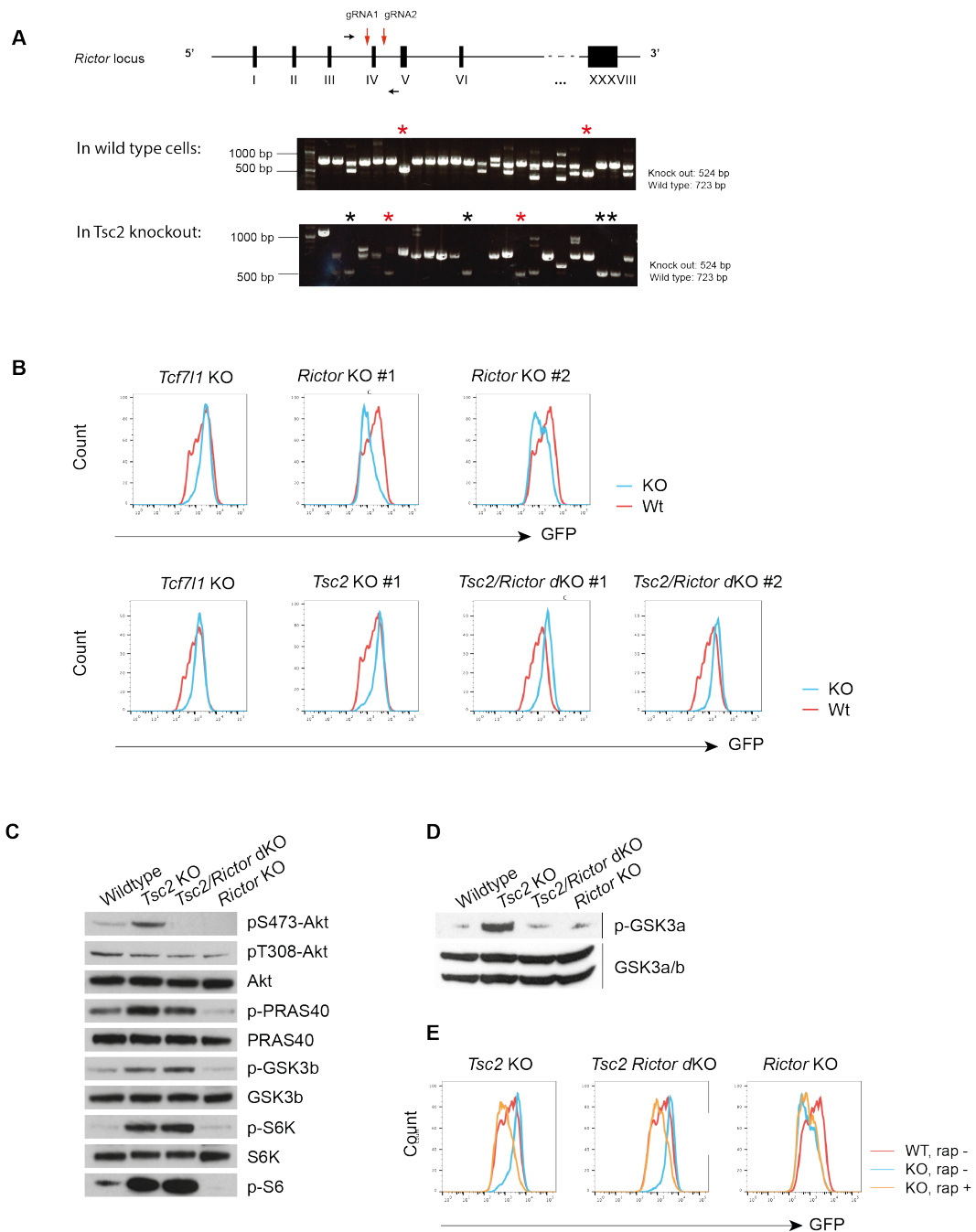
**Figure 5.5:** GATOR1 KO phenotype is resulted from activated GSK3. (A) Rex1-GFP positive population of GATOR1 key component (*Nprl2* and *Depdc5*) KO cells was sorted and maintained for six days in serum/LIF and 2i/LIF conditions. Rex1-GFP expression was monitored by flow cytometry every day. (B) Rex1-GFP negative population of *Nprl2* and *Depdc5* KO cells were plated in serum/LIF. Rex1-GFP expression was monitored by flow cytometry every day for four days. (C) WT, *Nprl2* KO and *Tsc2* KO cell lines were starved in basal medium for 2 hours and stimulated with N2B27 for 10 minutes. Cell lysate was probed with indicated antibodies. (D) Quantification of Gsk3 phosphorylation shown in (C). (E) *Nprl2* KO and WT cells were plated in N2B27+PD0325901+LIF supplemented with serially diluted GSK3 inhibitor concentration. The percentage of Rex1-GFP positive population was measured by flow cytometry. (F) Rapamycin treatment of *Nprl2* KO cells in N2B27+PD0325901+LIF. The percentage of Rex1-GFP positive population was measured by flow cytometry. (G) Rapamycin treatment rescued accelerated differentiation phenotype of *Nprl2* KO. Differentiation was induced by 2i/LIF withdrawal from N2B27. 20 nM rapamycin was added when cells were plated for differentiation. Rex1GFP expression was analysed after 27 hrs.

### 5.2.3 *Tsc2* knockout demonstrated unconventional cross-talk between mTOR and PI(3)K-Akt pathways

As mentioned in the previous section, *Tsc2* knockout mESC cells showed unconventional Akt phosphorylation pattern, namely upregulation of S473 phosphorylation (Figure 5.5 (C)). This observation suggests that, despite up-regulated mTORC1 activity, the S6K-mediated negative feedback loop does not seem to be operational: Akt activity was substantially up-regulated as shown by hyper-phosphorylated GSK3 and PRAS40 was observed. To investigate the role of hyperactivated Akt in *Tsc2* knockout mESCs, I generated *Rictor* knockout and *Tsc2*\ *Rictor* double knockout mESC lines (Figure 5.6 (A)). Rictor is an essential component of mTORC2, which has been identified as the kinase that phosphorylates Akt at S473 and fully activates Akt. In my screen, *Rictor* was identified as a significant hit whose knockout was depleted from the Rex1-GFP population, indicating that Rictor knockout accelerates differentiation. Consistent with the screening results, *Rictor* knockout ESCs demonstrated accelerated differentiation phenotype (Figure 5.6 (B)). However, *Tsc2*\ *Rictor* double knockout cells exhibited Rex1-GFP profile similar to *Tsc2* knockout, indicating that hyperactive Akt was not the causative factor for the delayed differentiation phenotype of *Tsc2* knockout.

To further dissect the molecular functions of mTORC2 and Akt, I analysed the phosphorylation status of the key components of Akt-mTOR pathway in *Rictor* knockout, *Tsc2* knockout and *Tsc2*\ *Rictor* double knockout. As shown in Figure 5.6 (C), loss of mTORC2 activity did not affect the up-regulated mTORC1 activity in *Tsc2* knockout background as indicated by highly phosphorylated S6K and S6 in *Tsc2*\ *Rictor* double knockout cells. Deletion of Rictor completely abolished Akt phosphorylation at S473 in both *Rictor* knockout and *Tsc2*\ *Rictor* double knockout. Not much change was observed on Akt phosphorylation at T308. An Akt kinase assay was performed by Jason Yu in the lab using GSK3a as the substrate and showed that Akt activity in *Tsc2*\ *Rictor* double knockout cells was almost abolished (Figure 5.6 (D)). However, interestingly, two Akt downstream targets PRAS40 and GSK3 retained high levels of phosphorylation in double knockout ESCs comparable to that of the *Tsc2* knockout cells, which indicates that the phosphorylation of PRAS40 and GSK3 in *Tsc2*\ *Rictor* double knockout background is resulted from some unknown kinases other than Akt.





**Figure 5.6:** The unconventional cross-talk between mTOR and PI(3)K-Akt pathways in *Tsc2* KO. (A) Schematic targeting strategy and genotyping results of *Rictor* KO. *Rictor* KO was generated using WT and *Tsc2* KO cells. Two gRNAs were designed to flank a critical exon and transfected into Rex1:GFPd2 cell line. Transfected cells were plated in clonal density and single cell colonies were picked and genotyped by PCR. Black arrows indicate PCR genotyping primers. Red arrows indicate gRNA cutting sites. gRNA and PCR primer sequences are listed in Chapter 2 Table:2.1 and Table:2.2. Asterisk indicate KO clones. Red asterisk indicates KO clones used in analysis. Roman numbers indicate exon numbers.(B) Rex1-GFP profile of *Rictor* KO, *Tsc2* KO and *Tsc2\Rictor* double KO after 27 hours differentiation in N2B27. Blue - indicated KO cells; Red - wt Rex1:GFP cells.(C) Immunoblotting result of WT, *Rictor* KO, *Tsc2* KO and *Tsc2\Rictor* double KO. Cell lysate was probed with indicated antibodies. (D) Akt kinase assay using GSK3a as a substrate (Performed by Jason Yu). (E) Rapamycin treatment rescued the impeded differentiation phenotypes of *Tsc2* KO and *Tsc2\Rictor* double KO. Differentiation was induced by 2i/LIF withdrawal from N2B27. 20 nM rapamycin was added when cells were plated for differentiation. Rex1GFP expression was analysed after 27 hrs. Red - wt rapamycin non-treated Rex1:GFP cells; Blue - indicated KO rapamycin non-treated; Yellow - indicated KO rapamycin treated.

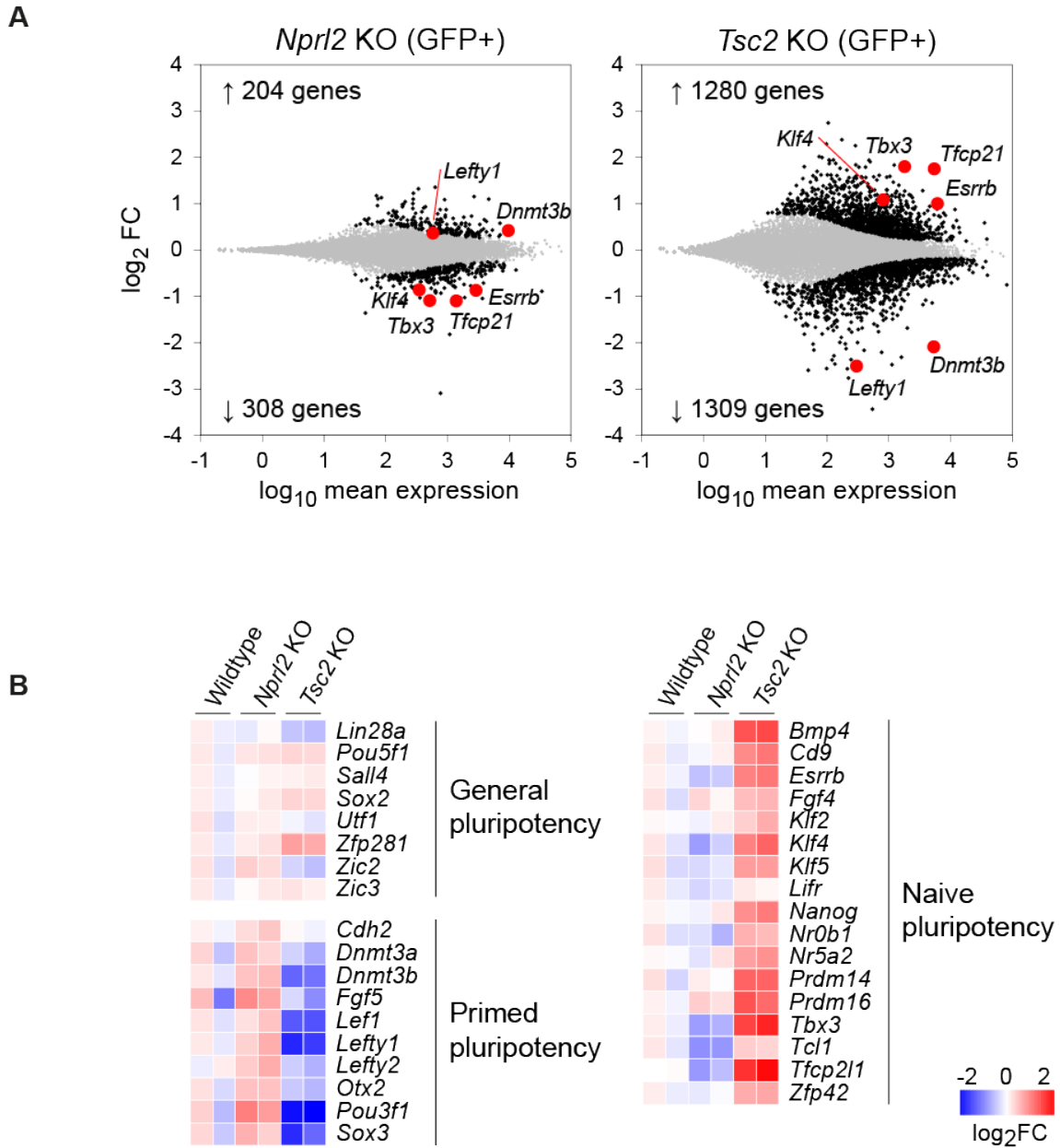
Given that depletion of *Tsc2* strongly activates mTORC1 activity, one of the possibilities is that the kinase responsible for GSK3 phosphorylation lies downstream of the hyper-activated mTORC1. In order to test this possibility, I treated cells with rapamycin and analysed Rex1GFP profiles at 27 hours differentiation. As shown in Figure 5.6 (E), delayed differentiation phenotype in *Tsc2* knockout and *Tsc2*\Rictor double knockout ESCs was completely rescued by rapamycin treatment. And these knockout cells exhibited differentiation profiles comparable to that of the wild type cells.

#### 5.2.4 Transcriptional changes in *Nprl2* and *Tsc2* knockouts

To further explore the downstream effect of *Nprl2* knockout and *Tsc2* knockout, I performed RNA-seq analysis and examined the global expression differences between these two genotypes. As these knockout cells exhibit different levels of heterogeneity of naive pluripotent cells, I purified Rex1-GFP positive populations from *Nprl2*, *Tsc2* knockout and wild type ESCs by cell sorting and performed differential gene expression analysis (Figure 5.7 (A)). At the cutoff threshold at adjusted *P*-value less than 0.1, the analysis revealed 204 up-regulated genes and 308 down-regulated genes in *Nprl2* knockout naive ESCs. *Tsc2* knockout naive ESCs exhibited substantially more differentially expressed genes: 1280 genes were up-regulated and 1390 genes were down-regulated. Surprisingly, despite the fact that only Rex1-GFP positive population was analysed, genes associated with naive pluripotency such as *Klf4*, *Tbx3*, *Esrrb* and *Tfcp2l1* were down-regulated in *Nprl2* knockout and upregulated in *Tsc2* knockout, whereas genes in relation to primed pluripotency such as *Lefty1* and *Dnmt3b* were found to be up-regulated in *Nprl2* knockout and down-regulated in *Tsc2* knockout.

To further analyse the pluripotency status of each knockout, I used curated panels of markers for general, naive and primed pluripotency[184], and examined their expression level in the RNA-Seq datasets (Figure 5.7 (B)). It was found that general pluripotency markers were expressed at a similar level across the three genotypes. Naive pluripotency markers were up-regulated in *Tsc2* knockout and primed pluripotency markers, which are associated with post-implantation epiblast, were down-regulated in *Tsc2* knockout, strongly indicating a more stabilised naive pluripotency status in *Tsc2* knockout than wild type. Complete opposite expression pattern was observed in *Nprl2* knockout, indicating that cells existed in a more primed pluripotency state despite Rex1-GFP expression. This finding indicates that although *Nprl2* and *Tsc2* both function as mTORC1 inhibitors, the alteration of these two genes resulted in completely different outcomes in pluripotency-

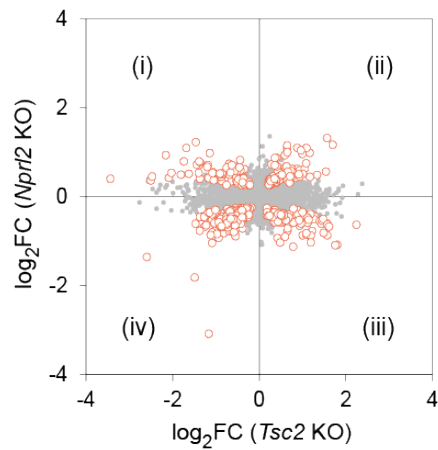
related gene expression.



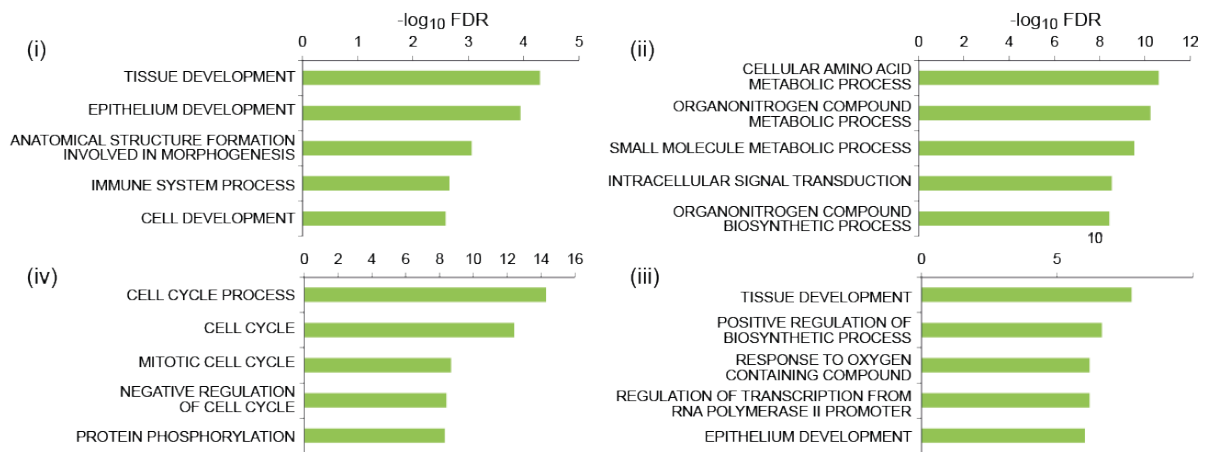
**Figure 5.7:** Transcriptional changes in *Nprl2* and *Tsc2* knockouts. (A) Global gene expression changes in Rex1-GFP positive populations of *Nprl2* and *Tsc2* knockout ESCs measured by RNA-Seq. FC: Fold change. Black dots represent differentially expressed genes. Several pluripotency/differentiation related genes were highlighted in red as examples. (B) Expression of selected general, naive and primed pluripotency markers from two independent replicates of wild type, *Nprl2* knockout and *Tsc2* knockout ESCs. Scale bar represents Log2 fold change compared to the mean TMP value of wild type replicates.

To dissect the differences in *Nprl2* and *Tsc2* knockout gene expression profiles, I divided the differentially expressed genes into four categories (Figure 5.8 (A)) and performed pathway enrichment analysis. As shown in Figure 5.8 (B), the analysis revealed that highly up-regulated genes in both genotypes were associated with metabolic processes such as amino acid metabolic process and organonitrogen metabolic process, probably as a result of stimulated mTORC1 activity. Genes that were commonly down-regulated were associated with regulation of cell cycle, notably, negative regulation of cell cycle. I postulated that constitutive active mTORC1 promotes cell proliferation thus inhibits genes that negatively regulates cell cycle. Genes that were up-regulated in *Nprl2* knockout but down-regulated in *Tsc2* knockout were associated with developmental process, which highly agrees with its accelerated differentiation phenotype. Pathways enriched from genes that were up-regulated in *Tsc2* knockout but down-regulated in *Nprl2* knockout include 'Response to oxygen containing compound', which is in line with the fact that TSC complex is involved in oxygen sensing. Interestingly, transcription-related pathway was also enriched from genes that were uniquely up-regulated in *Tsc2* knockout ESCs, which indicates that *Tsc2* might be involved in transcription regulation. Whether this is through mTORC1 inhibition remain to be tested.

A



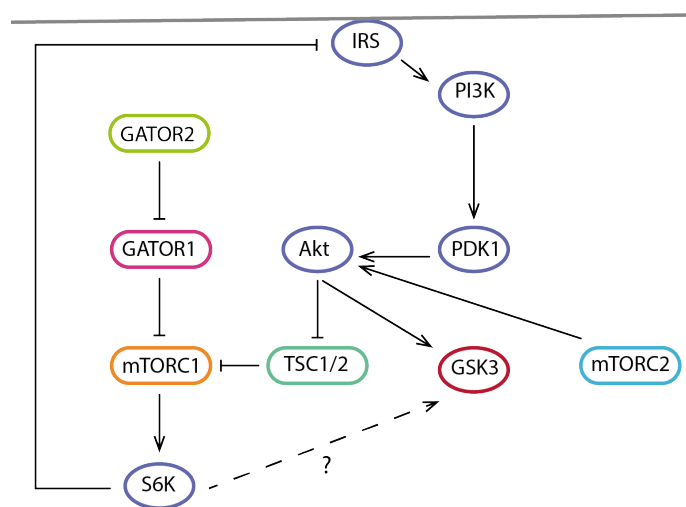
B



**Figure 5.8:** Pathway analysis of differentially expressed gene sets. (A) Differentially expressed genes were grouped into four categories: up-regulated in *Nprl2* knockout but down-regulated in *Tsc2* knockout (i), up-regulated in both genotypes (ii), up-regulated in *Tsc2* knockout but down-regulated in *Nprl2* knockout (iii), down-regulated in both genotypes (iv). (B) Enriched pathways of differentially expressed gene sets in each category.

### 5.3 Discussion and conclusion

The discovery of many regulators on the mTOR axis in relation to the exit from pluripotency is unprecedented, owing to the power of CRISPR-Cas9-mediated screen and the recent advance in knowledge on mTOR pathway. Although the role of mTOR and its related proteins in development has been studied previously [46] [19] [101], its exact function and the mechanism of interaction with its upstream regulators and downstream targets to make an impact on pluripotency and differentiation regulation remain a mystery. Interestingly, the screen hits covered a broad range of pathways upstream mTORC1, including growth factor stimulation, amino acid sensing and oxygen level regulation. It is noteworthy that almost all the major complexes involved in amino acid sensing were recovered and the outcome of each knockout correlated very well with its role in the pathway. Through studying the knockout phenotype of GATOR1 components, I demonstrated that the amino acid sensing-related mTORC1 regulators are involved in pluripotency regulation via influencing the Gsk3 activity by Akt (summarised in Figure 5.9).



**Figure 5.9:** Summary of major relevant signalling components upstream and downstream of mTORC1 and mTORC2. mTORC1 regulation by amino acid sensing is operated via the GATOR complexes: GATOR1 negatively regulates mTORC1, whereas GATOR2 functions as a positive regulator of mTORC1 from its inhibition of GATOR1. Knocking out key components (Nprl2 and Depdc5) of GATOR1 complex confers delayed differentiation in mESCs. This is due to the increased phosphorylation of S6K resulted from mTORC1 activation, which creates a negative feedback loop to down-regulate Akt via IRS-PI3K-PDK1. The weakened Akt leads to reduced Gsk3 phosphorylation, which impairs the pluripotency status through destabilised downstream  $\beta$ -catenin. Knocking out TSC1/2 complex, which is another major inhibitor of mTORC1, also showed increased phosphorylation of S6K. However, conventional negative feedback loop was not operating as shown by increased Akt activity. The up-regulated Akt does not contribute to the delayed differentiation in Tsc2 knockout mESCs as knocking out Rictor (key component of mTORC2, which phosphorylates Akt and is responsible for its full activation) in Tsc2 deficient mESCs did not rescue the phenotype. There are evidence suggesting S6K rather than Akt phosphorylates and inactivates Gsk3 in MEFs [481], but whether this is the case in mESCs remains to be tested.

Gsk3 is a key negative regulator of the canonical Wnt pathway by phosphorylating and promoting degradation of  $\beta$ -catenin. As discussed in Chapter 1, for naive pluripotency maintenance, the major role of  $\beta$ -catenin is to abrogate the transcriptional repressor function of Tcf711. Gsk3 inhibition thus positively contributes to sustaining naive pluripotency and suppressing differentiation. Figure 5.5 (E) and (F) clearly indicates that *Nprl2* knockout mESCs have an elevated Gsk3 activity and thus are more dependent on Gsk3 inhibition to maintain Rex1GFP expression. The delayed differentiation phenotype could be rescued by rapamycin (Figure 5.5 (G)), which indicates an operative negative feedback loop in *Nprl2* knockout ESCs from hyperactive S6K to Akt via IRS.

One of the striking findings from the screen was that *Tsc2* knockout cells demonstrated strong delayed exit of pluripotency which is opposite of *Nprl2* knockout despite both act as mTORC1 inhibitors. This is probably also due to hyper-phosphorylated Gsk3 as shown in Figure 5.5 (C). Although there must be distinct mechanisms in the control of pluripotency regulation underlying the opposite phenotype between *Tsc2* and *Nprl2* knockout ESCs. Additionally, phosphorylation analysis revealed unexpected Akt activity in *Tsc2* knockout. Firstly, Akt appeared to be up-regulated compared to wild type cells evidenced by the hyper-phosphorylated downstream targets Pras40 and the result of Akt kinase assay Figure 5.6 (C) (D), which is contradicting the established knowledge in cancer cells that activated mTORC1 induces negative feedback loop and down-regulates Akt [252]. Furthermore, Huang et al. demonstrated that TSC complex physically interacts and activates mTORC2, which is required for the phosphorylation of Akt at S473. Mouse embryonic fibroblast cells deficient of *Tsc2* lack of Akt S473 phosphorylation, which is in sharp contrast with what we observed in mouse ESCs [163] (Figure 5.6 (C)). These findings raised a possibility that the stabilised pluripotency status is resulted from upregulated Akt. However, *Tsc2*\Rictor double knockout did not rescue the phenotype observed in *Tsc2* knockout Figure 5.6 (D), suggesting that Akt is not involved in regulating Gsk3 in *Tsc2* knockout cells and that Gsk3 is rather regulated by other kinases. Given that rapamycin could rescue the phenotype caused by *Tsc2* knockout, the kinase that phosphorylates Gsk3 is most likely downstream of mTORC1. It has been described that in *Tsc2* knockout MEFs, S6K rather than Akt phosphorylates and inactivates Gsk3, which might also be the case in mouse ESCs [481]. To test this, S6k inhibitor or gRNAs targeting S6k could be used in *Tsc2* knockout ESCs.

## Chapter 6

# Conclusions and Future Perspectives



This thesis describes the use of CRISPR-Cas9-based genome-wide knockout library to study the transition from pluripotency to differentiation in mouse embryonic stem cells. It encompasses three major sections of work. The first part focused on the preparation work and proof-of-principle studies, including generation of constitutive Cas9 expressing Rex1:GFPd2 cell line and analysis of the differentiation profiles of Tcf7l1 and Apc knockout mESCs. These studies established screening foundations and provided valuable insights into screening design optimisation. The second part describes a detailed analysis of the screening result. Recovery of several well-studied pathways and factors confirmed that the screen was effective in identifying drivers of commitment. The discovery of previously unknown pathways and pluripotency regulators suggested the powerfulness and sensitivity of the screen. The final section focused on the mechanistic studies of mTOR-related pathway in regulation of pluripotency and differentiation. Through detailed dissection of the knockout phenotypes of *Nprl2* and *Tsc2*, I demonstrated that GATOR1 and TSC2 complexes regulate pluripotency via distinct routes, albeit both function as mTORC1 inhibitors.

## **6.1 The immeasurable potential of CRISPR-Cas9-mediated genetic screen in stem cell biology**

Since the establishment of ESC cultures more than three decades ago, studies have revealed a number of genetic and epigenetic machineries regulating pluripotency and differentiation. However, the mechanism behind the transition from pluripotency to differentiation has not been fully understood. This is partly due to the lack of scalable and high-throughput genetic methodology that allows comprehensive investigation of genes causing a specific phenotype. So far stem cell research has been typically carried out with resource-intensive hypothesis-driven approaches. The hypothesis-free forward genetics approach has been applied in yeast, *Drosophila* and *Caenorhabditis elegans* and provided deep insights into diverse biological processes [119] [178] [387]. A few large-scale screens were performed in stem cells (summarised in Table:4.1), but being limited by the technology available at the time, they were shown to be inefficient and lack of power.

The advent of CRISPR-Cas9 technology provided unprecedented opportunities to perform genetic screens and investigate genetic pathways underlying cellular processes in mammalian cells. The results described in this thesis demonstrated immense power and remarkable sensitivity in identifying genes involved in the regulation of differentiation

initiation. The CRISPR-Cas9-mediated screen is a step forward from RNAi-mediated screens and insertion-mediated screens in regard to knockout efficiency, off-target effect and simplicity in identifying causative mutations. It is also advantageous in offering the opportunity for a truly unbiased genome-wide comprehensive analysis of a particular biological process.

Most recently published CRISPR-Cas9 screens in mammalian cell culture used cell growth as a model [422] [153] [435]. Unlike those, the screen describe in this thesis used FACS-based reporter gene expression as a readout for positive selection. This mode of genetic screening is relatively less explored, possibly due to the difficulties in maintaining library complexity during cell sorting. To overcome such technical challenges, a series of optimisation experiments were performed for the best outcome of large-scale cell sorting. Under optimised conditions (described in Chapter 2), I could collect at least  $5 \times 10^6$  cells per replicate for downstream genomic DNA extraction and gRNA amplification. This achieves 50X coverage per gRNA, which is approximately half of the coverage in negative-selected cancer cell growth studies.

Furthermore, as shown in Chapter 4 and 5, by carefully choosing screening parameters, one can possibly uncover both enriched and depleted gRNAs from a screen and identify both positive and negative regulators of the process. Additionally, as the phenotype is assessed based on the gRNA abundance in control and treated samples, statistical analysis can be performed to quantitatively evaluate the phenotype caused by the knockout of a particular gene. The statistical significant value obtained from the screen is to some extent reflecting the level of association of a specific gene to the biological process, which provides a useful guidance for investigators to narrow down the candidate list and set up priorities for further investigation. The screening result described in this thesis demonstrated high sensitivity, which is well exemplified by uncovering genes whose mutation caused loss of Rex1-GFP expression. Although the screening conditions were optimised for positive selection to capture genes whose loss lead to sustained Rex1-GFP expression, 12 genes could be identified from negative enrichment with a more relaxed cut-off of FDR=25%. Even so, *Rictor*, which was ranked at 37 in the negative selection with an FDR of 51%, could be validated by a separate experiment in knockout mESCs (Figure 5.6). This indicates that, due to higher noise in negative selection, some genes failed to pass the statistical cutoff can still be meaningful and worth further investigation. Another gene *Sesn2*, encoding Sestrin2, is known to function as a leucine sensor whose loss results in continuous mTORC1 upregulation. *Sesn2* was ranked 46 in the negative selection with

an FDR of 0.61. Although it did not pass the statistical cutoff, the phenotype of *Sesn2* might be true, subject to further validation experiments. Re-designing the screen with optimised conditions for negative enrichment could be worthwhile to uncover the biology of mechanisms facilitates the initiation of differentiation.

The success of CRISPR-Cas9 genome-wide screen in unraveling genes involved in the exit of pluripotency using mouse ESCs demonstrated its potential in stem cell biology and opened up several lines of research work that could be performed in the near future. One of the apparent projects is to apply the same screen in human ESCs. Although mouse ESCs remain as invaluable materials in developmental biology, it is an imperfect model for studying human embryonic development [121]. Recent studies have developed several strategies to establish and maintain human ESCs in the naive state by over-expression of certain transcription factors or chemically manipulate the culture system [58] [128] [403] [437] [412]. CRISPR-Cas9-based genetic screening could facilitate the identification of genes that are required to safeguard human naive pluripotency state, as well as genes that drive the exit from pluripotency.

The discovery of iPSCs has provided a revolutionary platform for human disease modelling, stem cell-based therapy and drug discovery. Great progress has been made since its birth a decade ago. However, several important issues remain to be addressed: the current reprogramming process occurs stochastically with low efficiency and the resulted reprogrammed cells often exhibit variations in differentiation potential [371]. CRISPR-Cas9-based genetic screening could open a new avenue for understanding the reprogramming mechanism and basic biology of iPSCs to develop methods that facilitate the efficient generation of standardised human iPSCs. Another challenge for the research and therapeutic applications of iPSCs and ESCs is to differentiate them into specific cell types. Recent studies have been successful in discovering the key events during cell fate decisions; however, the currently available differentiation methods often result in heterogeneous population with low efficiency in derivation of desired cell types [427] [231] [31]. CRISPR-Cas9-based genetic screening holds great potential in understanding the genetic and molecular control that dictates the specification to distinct cellular identities, which facilitates the development of better strategies for directed differentiation of pluripotent cell to a specific cell type.

## 6.2 The complex regulatory network that drives dissolution of pluripotency

The screening result described in this thesis demonstrated a panoramic picture of cellular pathways that facilitate progression out of naive pluripotency. These include known signal transduction pathways such as Wnt, FGF/MAPK and PI3K pathways. The screenings hits also include processes that are indirectly involved in signal transductions such as the heparan sulphate biosynthesis pathway and endocytosis. As a positive regulator of the binding of Fgf and Fgfr, deficiency in heparan sulphate biosynthesis results in reduced FGF/MAPK signalling therefore negatively regulating the onset of differentiation. The endosomal trafficking pathway controls the intensity, duration and specificity of the signalling pathways. Of various vesicle trafficking and endocytosis complexes were identified, all 6 genes that compose the HOPS complex demonstrated noticeable differentiation defects. Other published pluripotency regulating pathways that were enriched in the screen include mRNA degradation and miRNA biogenesis pathways. To be specific, genes involved in mRNA decapping, deadenylation, nonsense-mediated decay and m<sup>6</sup>A mRNA methylation appeared to be significant in the screen, which is consistent with published studies [224] [132] [14]. miRNAs are also known to regulate differentiation [187] [378]. The screen has identified genes involved in miRNA biogenesis such as *Dicer1* and *Drosha*, as well as the miRNA nuclear exporter *Xpo5*. The effect of transcriptional regulatory proteins and chromatin modifiers on regulation of pluripotency and differentiation has been extensively studied. A large number of them has been captured in the screens, such as *Otx2*, *Zfp281*, polycomb repressive complexes and HDAC1/2 containing complexes. Interestingly, the screen has identified *Pou5f1* (protein: Oct4) as a gene required for differentiation. As a well-established core pluripotency gene, this finding seems somewhat contradicting. However, it has been shown recently that the effect of Oct4 is sensitive to its protein level and ESCs with heterozygous *Oc4* expression exhibited enhanced self-renewal and delayed differentiation kinetics [328] [188]. As CRISPR-Cas9 is a DSB-based technology which generates various alleles, there are probably a subpopulation of heterozygously edited cells in the mutant library which demonstrated delayed differentiation phenotype, while homozygous *Pou5f1* knockout ESCs were depleted from the population.

The screen has not only identified the majority of the pluripotency and differentiation regulators, but also revealed a large number of previously unknown factors, which broadened the landscape of factors regulating ESC identify and differentiation. This dataset provides

an important resource for the scientific community and raised several interesting biological questions for further investigation, such as the role of SAGA complex and NuA4 complex, which are highly conserved chromatin regulatory complexes whose function in the stem cell context has not been well-studied. Also, compare to genetic and epigenetic regulations, the role of metabolism in pluripotent stem cells is ill defined. Nearly half of the 563 screening hits were nucleus-encoded mitochondrial genes, indicating the functional importance of mitochondria and oxidative phosphorylation in the transition from naive to primed pluripotency. However, it was widely accepted that primed pluripotent cells mainly rely on glycolysis as the source of energy, and mitochondria respiration is reduced compare to naive pluripotent cells [489] [402] [184]. It is therefore a mystery whether ATP production is directly taking part in regulating the exit of naive pluripotency, and if so, what is the underlying molecular mechanism. Another possible explanation is that metabolism pathways integrate with genetic and epigenetic programs to regulate stem cell function and fate. Certain byproduct or intermediate product of metabolism such as acetyl-CoA can be transported into the nucleus for epigenetic modification. The screen revealed a crucial role of the under-appreciated metabolic pathways in pluripotency and differentiation, whose exact function and mechanism require further investigation. Another interesting module that worth further study is the ESCRT machinery and HOPS complex. Although likely to be involved in sustaining external signal transmission such as FGF-MAPK pathway, in the last few years, studies have revealed that late endosome and lysosome function not only in protein sorting and degradation, but also act as a platform for mTORC1 signalling pathway [355] [393]. It was also shown that the level of amino acids inside the lysosome lumen directly modulates mTORC1 activity [492]. It might be worth investigating any additional roles of the ESCRT and HOPS complexes.

Collectively, these pathways demonstrated strong interconnectivity and formed a functionally interdependent network. The dissolution of pluripotency is dependent on differentiation cues, which transmit into the cell via signalling cascades. One of the downstream effects of differentiation input is the repression of naive pluripotency network, which is achieved cooperatively by transcription repressors and chromatin modifiers. To ensure a rapid response to differentiation signals, the mRNA modification and degradation pathways are activated, which promotes the down-regulation of pluripotent transcripts. As the naive pluripotency network is dismantled, the cells initiate a new transcription programme which resembles the early post implantation epiblast. This transition process is supported by metabolic pathways, possibly through providing energy and intermediate metabolites

required for the function of genetic and epigenetic pathways.

### **6.3 The prominent role of mTOR and its related proteins in pluripotency and differentiation regulation**

This thesis highlighted the power of CRISPR-Cas9 mediated genetic screen in identifying novel pathways underlying biological processes in stem cells. Among the novel pathways I focused on showing the pivotal role of the mTOR-centred pathway in controlling the exit of pluripotency. mTOR regulators were overrepresented in the candidate list: there were 19 mTORC1 regulators and 2 components of mTORC2. This was especially true among the negative selection gene hits, where 5 of the top 10 significant genes are mTOR related. Surprisingly, deficiency in TSC complex and GATOR1 complex resulted in opposite phenotypes during ESC differentiation despite both function to inhibit mTORC1 activities. Through knocking out two GATOR1 components (Nprl2 and Depdc5), I demonstrated that mTORC1 activation by loss of mTORC1 negative regulators increases Gsk3 activity and destabilise naive pluripotency. I have also demonstrated through Rictor knockout mESCs that mTORC2 deficiency causes reduction of Akt activity and consequently increases Gsk3 activity, thereby destabilising naive pluripotency.

Conversely, mTORC1 upregulation by TSC complex deficiency resulted in the opposite phenotype, that is, delayed differentiation. Data in this thesis showed that Tsc2 loss in mESCs led a unique phosphorylation pattern: Akt in Tsc2 knockout exhibited higher activity evidenced by the hyper-phosphorylated downstream targets Pras40 and the result of Akt kinase assay. This is contradicting the established knowledge in cancer cells that activated mTORC1 induces negative feedback loop and down-regulates Akt[252]. Knock-out of Rictor in Tsc2 knockout cells demonstrated that high Akt activity and this was due to ectopically activated mTORC2. However, the activated Akt in Tsc2 KO mESCs does not play a major role in Gsk3 phosphorylation nor phenotypic outcome. It is still unclear why activated Akt does not phosphorylate its conventional target, Gsk3, and what the molecular basis of this phosphorylation rewiring is. S6K activation due to Nprl2 deficiency does not result in phosphorylation on Gsk3. As Tsc2 physically interact with the  $\beta$ -catenin degradation complex [246], TSC complex might serve as a molecular scaffold and recruit Akt to phosphorylate Gsk3. Upon loss of Tsc2, the  $\beta$ -catenin degradation complex is no longer attached to this platform and Gsk3 can now be accessed and phosphorylated by S6K. It would be interesting to further investigate these hypotheses.

Several mTORC1 regulators including Folliculin and Tsc2 have been identified as drivers of commitment in a siRNA-mediated screen, and a model has been proposed in which Folliculin acts downstream of mTOR and drives differentiation by restricting nuclear localisation [19]. This model was limited by the scarcity of knowledge of the biological function of Folliculin at the time. It was found later that the Folliculin-Fnip2 complex acts as a GAP for RagC/D that stimulates mTORC1 activity in the presence of amino acids [420] [321]. This agrees with GATOR1 phenotype which inhibits mTORC1 via acting as a GAP for RagA/B. Although there was a link between Folliculin-Fnip2 complex to the subcellular location of Tfe3, it is not clear how Folliculin-Fnip2 caused the nuclear exclusion of Tfe3 and how does mTORC1 involved this process. Therefore more work is to be done to unravel the detailed molecular mechanisms of mTOR-related pathways. Several questions were raised in this thesis, one of which is whether mTORC1 is directly involved in exit of pluripotency. Although many mTORC1 regulators were identified in the screen, knockout of mTORC1 itself does not seem to alter the differentiation profile given that none of its essential components exhibited any phenotype in the screen. Furthermore, it was observed that treating wild type cells with rapamycin did not affects their differentiation progress [19]. However, rapamycin treatment could rescue the aberrant differentiation phenotype of *Tsc2* knockout and *Nprl2* knockout. For further investigation, it may be worth adopting other perturbation approaches other than rapamycin such as conditional knockout or knockdown. It has also been reported two point mutations on mTOR confer constitutive activation of mTOR signalling even under starvation conditions [354]. These activated mutants could be helpful to reveal the role of mTOR by direct comparison with *Nprl2* knockout and *Tsc2* knockout. Another interesting finding that would be interesting to follow up is how *Tsc2* knockout resulted in mTORC2 activation and unconventional Akt regulation in mouse ESCs. It is hypothesised that TSC complex plays additional roles other than mTORC1 inhibitor. A promising approach is to perform CRISPR-Cas9-mediated genome-wide screen with *Tsc2* knockout ESCs.

# Bibliography

- [1] D C Ambrosetti, C Basilio, 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. *Mol Cell Biol*, 17(11):6321–9, Nov 1997.
- [2] D C Ambrosetti, H R Schöler, L Dailey, and C Basilio. Modulation of the activity of multiple transcriptional activation domains by the dna binding domains mediates the synergistic action of sox2 and oct-3 on the fibroblast growth factor-4 enhancer. *J Biol Chem*, 275(30):23387–97, Jul 2000.
- [3] Yen-Sin Ang, Su-Yi Tsai, Dung-Fang Lee, Jonathan Monk, Jie Su, Kajan Ratnakumar, Junjun Ding, Yongchao Ge, Henia Darr, Betty Chang, Jianlong Wang, Michael Rendl, Emily Bernstein, Christoph Schaniel, and Ihor R Lemischka. Wdr5 mediates self-renewal and reprogramming via the embryonic stem cell core transcriptional network. *Cell*, 145(2):183–97, Apr 2011.
- [4] Sebastian J Arnold and Elizabeth J Robertson. Making a commitment: cell lineage allocation and axis patterning in the early mouse embryo. *Nat Rev Mol Cell Biol*, 10(2):91–103, Feb 2009.
- [5] Giulio Auciello, Debbie L Cunningham, Tulin Tatar, John K Heath, and Joshua Z Rappoport. Regulation of fibroblast growth factor receptor signalling and trafficking by src and eps8. *J Cell Sci*, 126(Pt 2):613–24, Jan 2013.
- [6] Ariel A Avilion, Silvia K Nicolis, Larysa H Pevny, Lidia Perez, Nigel Vivian, and Robin Lovell-Badge. Multipotent cell lineages in early mouse development depend on sox2 function. *Genes Dev*, 17(1):126–40, Jan 2003.
- [7] Pedro Aza-Blanc, Christopher L Cooper, Klaus Wagner, Serge Batalov, Quinn L Deveraux, and Michael P Cooke. Identification of modulators of trail-induced apoptosis



- via rnai-based phenotypic screening. *Mol Cell*, 12(3):627–37, Sep 2003.
- [8] Véronique Azuara, Pascale Perry, Stephan Sauer, Mikhail Spivakov, Helle F Jørgensen, Rosalind M John, Mina Gouti, Miguel Casanova, Gary Warnes, Matthias Merckenschlager, and Amanda G Fisher. Chromatin signatures of pluripotent cell lines. *Nat Cell Biol*, 8(5):532–8, May 2006.
- [9] Sangsu Bae, Jeongbin Park, and Jin-Soo Kim. Cas-offinder: a fast and versatile algorithm that searches for potential off-target sites of cas9 rna-guided endonucleases. *Bioinformatics*, 30(10):1473–5, May 2014.
- [10] Liron Bar-Peled, Lynne Chantranupong, Andrew D Cherniack, Walter W Chen, Kathleen A Ottina, Brian C Grabiner, Eric D Spear, Scott L Carter, Matthew Meyerson, and David M Sabatini. A tumor suppressor complex with gap activity for the rag gtpases that signal amino acid sufficiency to mtorc1. *Science*, 340(6136):1100–6, May 2013.
- [11] Liron Bar-Peled and David M Sabatini. Regulation of mtorc1 by amino acids. *Trends Cell Biol*, 24(7):400–6, Jul 2014.
- [12] Rodolphe Barrangou, Christophe Fremaux, Hélène Deveau, Melissa Richards, Patrick Boyaval, Sylvain Moineau, Dennis A Romero, and Philippe Horvath. Crispr provides acquired resistance against viruses in prokaryotes. *Science*, 315(5819):1709–12, Mar 2007.
- [13] C Basilico, D Ambrosetti, D Fraidenraich, and L Dailey. Regulatory mechanisms governing fgf-4 gene expression during mouse development. *J Cell Physiol*, 173(2):227–32, Nov 1997.
- [14] Pedro J Batista, Benoit Molinie, Jinkai Wang, Kun Qu, Jiajing Zhang, Lingjie Li, Donna M Bouley, Ernesto Lujan, Bahareh Haddad, Kaveh Daneshvar, Ava C Carter, Ryan A Flynn, Chan Zhou, Kok-Seong Lim, Peter Dedon, Marius Wernig, Alan C Mullen, Yi Xing, Cosmas C Giallourakis, and Howard Y Chang. m(6)a rna modification controls cell fate transition in mammalian embryonic stem cells. *Cell Stem Cell*, 15(6):707–19, Dec 2014.
- [15] Issam Ben-Sahra, Gerta Hoxhaj, Stéphane J H Ricoult, John M Asara, and Brendan D Manning. mtorc1 induces purine synthesis through control of the mitochondrial tetrahydrofolate cycle. *Science*, 351(6274):728–733, Feb 2016.

- 
- [16] Katrien Berns, E Marielle Hijmans, Jasper Mullenders, Thijn R Brummelkamp, Arno Velds, Mike Heimerikx, Ron M Kerkhoven, Mandy Madiredjo, Wouter Nijkamp, Britta Weigelt, Reuven Agami, Wei Ge, Guy Cavet, Peter S Linsley, Roderick L Beijersbergen, and René Bernards. A large-scale rna screen in human cells identifies new components of the p53 pathway. *Nature*, 428(6981):431–7, Mar 2004.
- [17] Bradley E Bernstein, Tarjei S Mikkelsen, Xiaohui Xie, Michael Kamal, Dana J Huebert, James Cuff, Ben Fry, Alex Meissner, Marius Wernig, Kathrin Plath, Rudolf Jaenisch, Alexandre Wagschal, Robert Feil, Stuart L Schreiber, and Eric S Lander. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell*, 125(2):315–26, Apr 2006.
- [18] E Bernstein, A A Caudy, S M Hammond, and G J Hannon. Role for a bidentate ribonuclease in the initiation step of rna interference. *Nature*, 409(6818):363–6, Jan 2001.
- [19] Joerg Betschinger, Jennifer Nichols, Sabine Dietmann, Philip D Corrin, Patrick J Paddison, and Austin Smith. Exit from pluripotency is gated by intracellular redistribution of the bhlh transcription factor tfe3. *Cell*, 153(2):335–47, Apr 2013.
- [20] Nidhi Bhutani, Jennifer J Brady, Mara Damian, Alessandra Sacco, Stéphane Y Corbel, and Helen M Blau. Reprogramming towards pluripotency requires aid-dependent dna demethylation. *Nature*, 463(7284):1042–7, Feb 2010.
- [21] M Bibikova, D Carroll, D J Segal, J K Trautman, J Smith, Y G Kim, and S Chandrasegaran. Stimulation of homologous recombination through targeted cleavage by chimeric nucleases. *Mol Cell Biol*, 21(1):289–97, Jan 2001.
- [22] Marina Bibikova, Kelly Beumer, Jonathan K Trautman, and Dana Carroll. Enhancing gene targeting with designed zinc finger nucleases. *Science*, 300(5620):764, May 2003.
- [23] Marina Bibikova, Mary Golic, Kent G Golic, and Dana Carroll. Targeted chromosomal cleavage and mutagenesis in drosophila using zinc-finger nucleases. *Genetics*, 161(3):1169–75, Jul 2002.
- [24] Steve Bilodeau, Michael H Kagey, Garrett M Frampton, Peter B Rahl, and Richard A Young. Setdb1 contributes to repression of genes encoding developmental regulators and maintenance of es cell state. *Genes Dev*, 23(21):2484–9, Nov 2009.

- [25] Naomi Bishop, Alistair Horman, and Philip Woodman. Mammalian class e vps proteins recognize ubiquitin and act in the removal of endosomal protein-ubiquitin conjugates. *J Cell Biol*, 157(1):91–101, Apr 2002.
- [26] J Bitinaite, D A Wah, A K Aggarwal, and I Schildkraut. Foki dimerization is required for dna cleavage. *Proc Natl Acad Sci U S A*, 95(18):10570–5, Sep 1998.
- [27] Jens Boch and Ulla Bonas. Xanthomonas avrbs3 family-type iii effectors: discovery and function. *Annu Rev Phytopathol*, 48:419–36, 2010.
- [28] Jens Boch, Heidi Scholze, Sebastian Schornack, Angelika Landgraf, Simone Hahn, Sabine Kay, Thomas Lahaye, Anja Nickstadt, and Ulla Bonas. Breaking the code of dna binding specificity of tal-type iii effectors. *Science*, 326(5959):1509–12, Dec 2009.
- [29] Alexander Bolotin, Benoit Quinquis, Alexei Sorokin, and S Dusko Ehrlich. Clustered regularly interspaced short palindrome repeats (crisprs) have spacers of extrachromosomal origin. *Microbiology*, 151(Pt 8):2551–61, Aug 2005.
- [30] Thorsten Boroviak, Remco Loos, Paul Bertone, Austin Smith, and Jennifer Nichols. The ability of inner-cell-mass cells to self-renew as embryonic stem cells is acquired following epiblast specification. *Nat Cell Biol*, 16(6):516–28, Jun 2014.
- [31] Malgorzata Borowiak, René Maehr, Shuibing Chen, Alice E Chen, Weiping Tang, Julia L Fox, Stuart L Schreiber, and Douglas A Melton. Small molecules efficiently direct endodermal differentiation of mouse and human embryonic stem cells. *Cell Stem Cell*, 4(4):348–58, Apr 2009.
- [32] Hicham Bouabe and Klaus Okkenhaug. Gene targeting in mice: a review. *Methods Mol Biol*, 1064:315–36, 2013.
- [33] Laurie A Boyer, Tong Ihn Lee, Megan F Cole, Sarah E Johnstone, Stuart S Levine, Jacob P Zucker, Matthew G Guenther, Roshan M Kumar, Heather L Murray, Richard G Jenner, David K Gifford, Douglas A Melton, Rudolf Jaenisch, and Richard A Young. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell*, 122(6):947–56, Sep 2005.
- [34] A Bradley, M Evans, M H Kaufman, and E Robertson. Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines. *Nature*, 309(5965):255–6, 1984.

- 
- [35] Abraham L Brass, Derek M Dykxhoorn, Yair Benita, Nan Yan, Alan Engelman, Ramnik J Xavier, Judy Lieberman, and Stephen J Elledge. Identification of host proteins required for hiv infection through a functional genomic screen. *Science*, 319(5865):921–6, Feb 2008.
- [36] Nicolas L Bray, Harold Pimentel, Páll Melsted, and Lior Pachter. Near-optimal probabilistic rna-seq quantification. *Nat Biotechnol*, 34(5):525–7, May 2016.
- [37] Adrian W Briggs, Xavier Rios, Raj Chari, Luhan Yang, Feng Zhang, Prashant Mali, and George M Church. Iterative capped assembly: rapid and scalable synthesis of repeat-module dna such as tal effectors from individual monomers. *Nucleic Acids Res*, 40(15):e117, Aug 2012.
- [38] I Gabrielle M Brons, Lucy E Smithers, Matthew W B Trotter, Peter Rugg-Gunn, Bowen Sun, Susana M Chuva de Sousa Lopes, Sarah K Howlett, Amanda Clarkson, Lars Ahrlund-Richter, Roger A Pedersen, and Ludovic Vallier. Derivation of pluripotent epiblast stem cells from mammalian embryos. *Nature*, 448(7150):191–5, Jul 2007.
- [39] Stan J J Brouns, Matthijs M Jore, Magnus Lundgren, Edze R Westra, Rik J H Slijkhuis, Ambrosius P L Snijders, Mark J Dickman, Kira S Makarova, Eugene V Koonin, and John van der Oost. Small crispr rnas guide antiviral defense in prokaryotes. *Science*, 321(5891):960–4, Aug 2008.
- [40] E J Brown, M W Albers, T B Shin, K Ichikawa, C T Keith, W S Lane, and S L Schreiber. A mammalian protein targeted by g1-arresting rapamycin-receptor complex. *Nature*, 369(6483):756–8, Jun 1994.
- [41] James Brugarolas, Kui Lei, Rebecca L Hurley, Brendan D Manning, Jan H Reiling, Ernst Hafen, Lee A Witters, Leif W Ellisen, and William G Kaelin, Jr. Regulation of mtor function in response to hypoxia by redd1 and the tsc1/tsc2 tumor suppressor complex. *Genes Dev*, 18(23):2893–904, Dec 2004.
- [42] Thijn R Brummelkamp, René Bernards, and Reuven Agami. Stable suppression of tumorigenicity by virus-mediated rna interference. *Cancer Cell*, 2(3):243–7, Sep 2002.
- [43] Thijn R Brummelkamp, René Bernards, and Reuven Agami. A system for stable expression of short interfering rnas in mammalian cells. *Science*, 296(5567):550–3, Apr 2002.

- [44] G J Brunn, C C Hudson, A Sekulić, J M Williams, H Hosoi, P J Houghton, J C Lawrence, Jr, and R T Abraham. Phosphorylation of the translational repressor phas-i by the mammalian target of rapamycin. *Science*, 277(5322):99–101, Jul 1997.
- [45] Christa Buecker, Rajini Srinivasan, Zhixiang Wu, Eliezer Calo, Dario Acampora, Tiago Faial, Antonio Simeone, Minjia Tan, Tomasz Swigut, and Joanna Wysocka. Reorganization of enhancer patterns in transition from naive to primed pluripotency. *Cell Stem Cell*, 14(6):838–53, Jun 2014.
- [46] Aydan Bulut-Karslioglu, Steffen Biechele, Hu Jin, Trisha A Macrae, Miroslav Hejna, Marina Gertsenstein, Jun S Song, and Miguel Ramalho-Santos. Inhibition of mtor induces a paused pluripotent state. *Nature*, 540(7631):119–123, 12 2016.
- [47] T Burdon, C Stracey, I Chambers, J Nichols, and A Smith. Suppression of shp-2 and erk signalling promotes self-renewal of mouse embryonic stem cells. *Dev Biol*, 210(1):30–43, Jun 1999.
- [48] Ling Cai, Benjamin M Sutter, Bing Li, and Benjamin P Tu. Acetyl-coa induces cell growth and proliferation by promoting the acetylation of histones at growth genes. *Mol Cell*, 42(4):426–37, May 2011.
- [49] Maurice A Canham, Alexei A Sharov, Minoru S H Ko, and Joshua M Brickman. Functional heterogeneity of embryonic stem cells revealed through translational amplification of an early endodermal transcript. *PLoS Biol*, 8(5):e1000379, May 2010.
- [50] Matthew C Canver, Elenoe C Smith, Falak Sher, Luca Pinello, Neville E Sanjana, Ophir Shalem, Diane D Chen, Patrick G Schupp, Divya S Vinjamur, Sara P Garcia, Sidinh Luc, Ryo Kurita, Yukio Nakamura, Yuko Fujiwara, Takahiro Maeda, Guo-Cheng Yuan, Feng Zhang, Stuart H Orkin, and Daniel E Bauer. Bcl11a enhancer dissection by cas9-mediated in situ saturating mutagenesis. *Nature*, 527(7577):192–7, Nov 2015.
- [51] Iara D Carbery, Diana Ji, Anne Harrington, Victoria Brown, Edward J Weinstein, Lucy Liaw, and Xiaoxia Cui. Targeted genome modification in mice using zinc-finger nucleases. *Genetics*, 186(2):451–9, Oct 2010.
- [52] Jan E Carette, Carla P Guimaraes, Malini Varadarajan, Annie S Park, Irene Wuethrich, Alzbeta Godarova, Maciej Kotecki, Brent H Cochran, Eric Spooner, Hidde L Ploegh, and Thijn R Brummelkamp. Haploid genetic screens in human cells identify host factors used by pathogens. *Science*, 326(5957):1231–5, Nov 2009.

- [53] Peter Cartwright, Cameron McLean, Allan Sheppard, Duane Rivett, Karen Jones, and Stephen Dalton. Lif/stat3 controls es cell self-renewal and pluripotency by a myc-dependent mechanism. *Development*, 132(5):885–96, Mar 2005.
- [54] Tomas Cermak, Erin L Doyle, Michelle Christian, Li Wang, Yong Zhang, Clarice Schmidt, Joshua A Baller, Nikunj V Somia, Adam J Bogdanove, and Daniel F Voytas. Efficient design and assembly of custom talen and other tal effector-based constructs for dna targeting. *Nucleic Acids Res*, 39(12):e82, Jul 2011.
- [55] Stormy J Chamberlain, Della Yee, and Terry Magnuson. Polycomb repressive complex 2 is dispensable for maintenance of embryonic stem cell pluripotency. *Stem Cells*, 26(6):1496–505, Jun 2008.
- [56] Ian Chambers, Douglas Colby, Morag Robertson, Jennifer Nichols, Sonia Lee, Susan Tweedie, and Austin Smith. Functional expression cloning of nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell*, 113(5):643–55, May 2003.
- [57] Ian Chambers, Jose Silva, Douglas Colby, Jennifer Nichols, Bianca Nijmeijer, Morag Robertson, Jan Vrana, Ken Jones, Lars Grotewold, and Austin Smith. Nanog safeguards pluripotency and mediates germline development. *Nature*, 450(7173):1230–4, Dec 2007.
- [58] Yun-Shen Chan, Jonathan Göke, Jia-Hui Ng, Xinyi Lu, Kevin Andrew Uy Gonzales, Cheng-Peow Tan, Wei-Quan Tng, Zhong-Zhi Hong, Yee-Siang Lim, and Huck-Hui Ng. Induction of a human pluripotent state with distinct regulatory circuitry that resembles preimplantation epiblast. *Cell Stem Cell*, 13(6):663–75, Dec 2013.
- [59] Srinivasan Chandrasegaran and Dana Carroll. Origins of programmable nucleases for genome engineering. *J Mol Biol*, 428(5 Pt B):963–89, Feb 2016.
- [60] Lynne Chantranupong, Sonia M Scaria, Robert A Saxton, Melanie P Gygi, Kuang Shen, Gregory A Wyant, Tim Wang, J Wade Harper, Steven P Gygi, and David M Sabatini. The castor proteins are arginine sensors for the mtorc1 pathway. *Cell*, 165(1):153–164, Mar 2016.
- [61] Raj Chari, Prashant Mali, Mark Moosburner, and George M Church. Unraveling crispr-cas9 genome engineering parameters via a library-on-library approach. *Nat Methods*, 12(9):823–6, Sep 2015.

- [62] Alejandro Chavez, Jonathan Scheiman, Suhani Vora, Benjamin W Pruitt, Marcelle Tuttle, Eswar P R Iyer, Shuailiang Lin, Samira Kiani, Christopher D Guzman, Daniel J Wiegand, Dmitry Ter-Ovanesyan, Jonathan L Braff, Noah Davidsohn, Benjamin E Housden, Norbert Perrimon, Ron Weiss, John Aach, James J Collins, and George M Church. Highly efficient cas9-mediated transcriptional programming. *Nat Methods*, 12(4):326–8, Apr 2015.
- [63] Claire Chazaud, Yojiro Yamanaka, Tony Pawson, and Janet Rossant. Early lineage segregation between epiblast and primitive endoderm in mouse blastocysts through the grb2-mapk pathway. *Dev Cell*, 10(5):615–24, May 2006.
- [64] Baohui Chen, Luke A Gilbert, Beth A Cimini, Joerg Schnitzbauer, Wei Zhang, Gene-Wei Li, Jason Park, Elizabeth H Blackburn, Jonathan S Weissman, Lei S Qi, and Bo Huang. Dynamic imaging of genomic loci in living human cells by an optimized crispr/cas system. *Cell*, 155(7):1479–91, Dec 2013.
- [65] Sidi Chen, Neville E Sanjana, Kaijie Zheng, Ophir Shalem, Kyunghoon Lee, Xi Shi, David A Scott, Jun Song, Jen Q Pan, Ralph Weissleder, Hakho Lee, Feng Zhang, and Phillip A Sharp. Genome-wide crispr screen in a mouse model of tumor growth and metastasis. *Cell*, 160(6):1246–60, Mar 2015.
- [66] Taiping Chen and Sharon Y R Dent. Chromatin modifiers and remodellers: regulators of cellular differentiation. *Nat Rev Genet*, 15(2):93–106, Feb 2014.
- [67] Taiping Chen, Yoshihide Ueda, Jonathan E Dodge, Zhenjuan Wang, and En Li. Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by dnmt3a and dnmt3b. *Mol Cell Biol*, 23(16):5594–605, Aug 2003.
- [68] Xi Chen, Han Xu, Ping Yuan, Fang Fang, Mikael Huss, Vinsensius B Vega, Eleanor Wong, Yuriy L Orlov, Weiwei Zhang, Jianming Jiang, Yui-Han Loh, Hock Chuan Yeo, Zhen Xuan Yeo, Vipin Narang, Kunde Ramamoorthy Govindarajan, Bernard Leong, Atif Shahab, Yijun Ruan, Guillaume Bourque, Wing-Kin Sung, Neil D Clarke, Chia-Lin Wei, and Huck-Hui Ng. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell*, 133(6):1106–17, Jun 2008.
- [69] A M Cheng, T M Saxton, R Sakai, S Kulkarni, G Mbamalu, W Vogel, C G Torrice, R D Cardiff, J C Cross, W J Muller, and T Pawson. Mammalian grb2

- regulates multiple steps in embryonic development and malignant transformation. *Cell*, 95(6):793–803, Dec 1998.
- [70] Albert W Cheng, Haoyi Wang, Hui Yang, Linyu Shi, Yarden Katz, Thorold W Theunissen, Sudharshan Rangarajan, Chikdu S Shivalila, Daniel B Dadon, and Rudolf Jaenisch. Multiplexed activation of endogenous genes by crispr-on, an rna-guided transcriptional activator system. *Cell Res*, 23(10):1163–71, Oct 2013.
- [71] Seung Woo Cho, Sojung Kim, Jong Min Kim, and Jin-Soo Kim. Targeted genome engineering in human cells with the cas9 rna-guided endonuclease. *Nat Biotechnol*, 31(3):230–2, Mar 2013.
- [72] Seung Woo Cho, Sojung Kim, Yongsub Kim, Jiyeon Kweon, Heon Seok Kim, Sangsu Bae, and Jin-Soo Kim. Analysis of off-target effects of crispr/cas-derived rna-guided endonucleases and nickases. *Genome Res*, 24(1):132–41, Jan 2014.
- [73] Jiho Choi, Aaron J Huebner, Kendell Clement, Ryan M Walsh, Andrej Savol, Kaixuan Lin, Hongcang Gu, Bruno Di Stefano, Justin Brumbaugh, Sang-Yong Kim, Jafar Sharif, Christopher M Rose, Arman Mohammad, Junko Odajima, Jean Charron, Toshi Shioda, Andreas Gnirke, Steven Gygi, Haruhiko Koseki, Ruslan I Sadreyev, Andrew Xiao, Alexander Meissner, and Konrad Hochedlinger. Prolonged mek1/2 suppression impairs the developmental potential of embryonic stem cells. *Nature*, 548(7666):219–223, Aug 2017.
- [74] Michelle Christian, Tomas Cermak, Erin L Doyle, Clarice Schmidt, Feng Zhang, Aaron Hummel, Adam J Bogdanove, and Daniel F Voytas. Targeting dna double-strand breaks with tal effector nucleases. *Genetics*, 186(2):757–61, Oct 2010.
- [75] E H Chu. Mammalian cell genetics. 3. characterization of x-ray-induced forward mutations in chinese hamster cell cultures. *Mutat Res*, 11(1):23–34, Jan 1971.
- [76] Hans Clevers. Wnt/beta-catenin signaling in development and disease. *Cell*, 127(3):469–80, Nov 2006.
- [77] Megan F Cole, Sarah E Johnstone, Jamie J Newman, Michael H Kagey, and Richard A Young. Tcf3 is an integral component of the core regulatory circuitry of embryonic stem cells. *Genes Dev*, 22(6):746–55, Mar 2008.
- [78] Le Cong, F Ann Ran, David Cox, Shuailiang Lin, Robert Barretto, Naomi Habib, Patrick D Hsu, Xuebing Wu, Wenyan Jiang, Luciano A Marraffini, and Feng Zhang.



- Multiplex genome engineering using crispr/cas systems. *Science*, 339(6121):819–23, Feb 2013.
- [79] Neal G Copeland and Nancy A Jenkins. Harnessing transposons for cancer gene discovery. *Nat Rev Cancer*, 10(10):696–706, Oct 2010.
- [80] Thomas J Cradick, Eli J Fine, Christopher J Antico, and Gang Bao. Crispr/cas9 systems targeting  $\beta$ -globin and ccr5 genes have substantial off-target activity. *Nucleic Acids Res*, 41(20):9584–92, Nov 2013.
- [81] Elitza Deltcheva, Krzysztof Chylinski, Cynthia M Sharma, Karine Gonzales, Yanjie Chao, Zaid A Pirzada, Maria R Eckert, Jörg Vogel, and Emmanuelle Charpentier. Crispr rna maturation by trans-encoded small rna and host factor rnae iii. *Nature*, 471(7340):602–7, Mar 2011.
- [82] Dong Deng, Chuangye Yan, Xiaojing Pan, Magdy Mahfouz, Jiawei Wang, Jian-Kang Zhu, Yigong Shi, and Nieng Yan. Structural basis for sequence-specific recognition of dna by tal effectors. *Science*, 335(6069):720–3, Feb 2012.
- [83] Sergei Denisov, Helmut Hofemeister, Hendrik Marks, Andrea Kranz, Giovanni Ciotta, Sukhdeep Singh, Konstantinos Anastassiadis, Hendrik G Stunnenberg, and A Francis Stewart. Mll2 is required for h3k4 trimethylation on bivalent promoters in embryonic stem cells, whereas mll1 is redundant. *Development*, 141(3):526–37, Feb 2014.
- [84] Amit M Deshpande, Yan-Shan Dai, Yong Kim, Jeffrey Kim, Lauren Kimlin, Kai Gao, and David T Wong. Cdk2ap1 is required for epigenetic silencing of oct4 during murine embryonic stem cell differentiation. *J Biol Chem*, 284(10):6043–7, Mar 2009.
- [85] Maurice Phillip DeYoung, Peter Horak, Avi Sofer, Dennis Sgroi, and Leif W Ellisen. Hypoxia regulates tsc1/2-mtor signaling and tumor suppression through redd1-mediated 14-3-3 shuttling. *Genes Dev*, 22(2):239–51, Jan 2008.
- [86] Yarui Diao, Bin Li, Zhipeng Meng, Inkyung Jung, Ah Young Lee, Jesse Dixon, Lenka Maliskova, Kun-Liang Guan, Yin Shen, and Bing Ren. A new class of temporarily phenotypic enhancers identified by crispr/cas9-mediated genetic screening. *Genome Res*, 26(3):397–405, Mar 2016.
- [87] Christian C Dibble, Winfried Elis, Suchithra Menon, Wei Qin, Justin Klekota, John M Asara, Peter M Finan, David J Kwiatkowski, Leon O Murphy, and Bren-

- dan D Manning. Tbc1d7 is a third subunit of the tsc1-tsc2 complex upstream of mtorc1. *Mol Cell*, 47(4):535–46, Aug 2012.
- [88] Ross A Dickins, Michael T Hemann, Jack T Zilfou, David R Simpson, Ingrid Ibarra, Gregory J Hannon, and Scott W Lowe. Probing tumor phenotypes using stable and regulated synthetic microRNA precursors. *Nat Genet*, 37(11):1289–95, Nov 2005.
- [89] Junjun Ding, Huilei Xu, Francesco Faiola, Avi Ma’ayan, and Jianlong Wang. Oct4 links multiple epigenetic pathways to the pluripotency network. *Cell Res*, 22(1):155–67, Jan 2012.
- [90] Li Ding, Maciej Paszkowski-Rogacz, Anja Nitzsche, Mikolaj Michal Slabicki, Anne-Kristin Heninger, Ingrid de Vries, Ralf Kittler, Magno Junqueira, Andrej Shevchenko, Herbert Schulz, Norbert Hubner, Michael Xavier Doss, Agapios Sachinidis, Juergen Hescheler, Roberto Iacone, Konstantinos Anastassiadis, A Francis Stewart, M Teresa Pisabarro, Antonio Caldarelli, Ina Poser, Mirko Theis, and Frank Buchholz. A genome-scale RNAi screen for Oct4 modulators defines a role of the Paf1 complex for embryonic stem cell identity. *Cell Stem Cell*, 4(5):403–15, May 2009.
- [91] Bradley W Doble, Satish Patel, Geoffrey A Wood, Lisa K Kockeritz, and James R Woodgett. Functional redundancy of GSK-3 $\alpha$  and GSK-3 $\beta$  in Wnt/ $\beta$ -catenin signaling shown by using an allelic series of embryonic stem cell lines. *Dev Cell*, 12(6):957–71, Jun 2007.
- [92] John G Doench, Nicolo Fusi, Meagan Sullender, Mudra Hegde, Emma W Vaimberg, Katherine F Donovan, Ian Smith, Zuzana Tothova, Craig Wilen, Robert Orchard, Herbert W Virgin, Jennifer Listgarten, and David E Root. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat Biotechnol*, 34(2):184–191, Feb 2016.
- [93] John G Doench, Ella Hartenian, Daniel B Graham, Zuzana Tothova, Mudra Hegde, Ian Smith, Meagan Sullender, Benjamin L Ebert, Ramnik J Xavier, and David E Root. Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nat Biotechnol*, 32(12):1262–7, Dec 2014.
- [94] T Doetschman, R G Gregg, N Maeda, M L Hooper, D W Melton, S Thompson, and O Smithies. Targeted correction of a mutant hprt gene in mouse embryonic stem cells. *Nature*, 330(6148):576–8, 1987.

- 
- [95] N Valerio Dorrello, Angelo Peschiaroli, Daniele Guardavaccaro, Nancy H Colburn, Nicholas E Sherman, and Michele Pagano. S6k1- and betatrcp-mediated degradation of pdcd4 promotes protein translation and cell growth. *Science*, 314(5798):467–71, Oct 2006.
- [96] Yair Dorsett and Thomas Tuschl. sirnas: applications in functional genomics and potential as therapeutics. *Nat Rev Drug Discov*, 3(4):318–29, Apr 2004.
- [97] Yannick Doyon, Vivian M Choi, Danny F Xia, Thuy D Vo, Philip D Gregory, and Michael C Holmes. Transient cold shock enhances zinc-finger nuclease-mediated gene disruption. *Nat Methods*, 7(6):459–60, Jun 2010.
- [98] Aurelie Doyotte, Matthew R G Russell, Colin R Hopkins, and Philip G Woodman. Depletion of tsg101 forms a mammalian "class e" compartment: a multicisternal early endosome with multiple sorting defects. *J Cell Sci*, 118(Pt 14):3003–17, Jul 2005.
- [99] A J Dupuy, S Fritz, and D A Largaespada. Transposition and gene disruption in the male germline of the mouse. *Genesis*, 30(2):82–8, Jun 2001.
- [100] Christophe J Echeverri and Norbert Perrimon. High-throughput rnai screening in cultured cells: a user's guide. *Nat Rev Genet*, 7(5):373–84, 05 2006.
- [101] Alejo Efeyan, Lawrence D Schweitzer, Angelina M Bilate, Steven Chang, Oktay Kirak, Dudley W Lamming, and David M Sabatini. Raga, but not ragb, is essential for embryonic development and adult mice. *Dev Cell*, 29(3):321–9, May 2014.
- [102] S M Elbashir, J Harborth, W Lendeckel, A Yalcin, K Weber, and T Tuschl. Duplexes of 21-nucleotide rnas mediate rna interference in cultured mammalian cells. *Nature*, 411(6836):494–8, May 2001.
- [103] S M Elbashir, W Lendeckel, and T Tuschl. Rna interference is mediated by 21- and 22-nucleotide rnas. *Genes Dev*, 15(2):188–200, Jan 2001.
- [104] Ulrich Elling, Jasmin Taubenschmid, Gerald Wirnsberger, Ronan O'Malley, Simon-Pierre Demers, Quentin Vanhaelen, Andrey I Shukalyuk, Gerald Schmauss, Daniel Schramek, Frank Schnuetgen, Harald von Melchner, Joseph R Ecker, William L Stanford, Johannes Zuber, Alexander Stark, and Josef M Penninger. Forward and reverse genetics through derivation of haploid mouse embryonic stem cells. *Cell Stem Cell*, 9(6):563–74, Dec 2011.

- [105] M Ernst, A Oates, and A R Dunn. Gp130-mediated signal transduction in embryonic stem cells involves activation of jak and ras/mitogen-activated protein kinase pathways. *J Biol Chem*, 271(47):30136–43, Nov 1996.
- [106] Daniel Esch, Juha Vahokoski, Matthew R Groves, Vivian Pogenberg, Vlad Cojocaru, Hermann Vom Bruch, Dong Han, Hannes C A Drexler, Marcos J Araújo-Bravo, Calista K L Ng, Ralf Jauch, Matthias Wilmanns, and Hans R Schöler. A unique oct4 interface is crucial for reprogramming to pluripotency. *Nat Cell Biol*, 15(3):295–301, Mar 2013.
- [107] M J Evans and M H Kaufman. Establishment in culture of pluripotential cells from mouse embryos. *Nature*, 292(5819):154–6, Jul 1981.
- [108] Bastiaan Evers, Katarzyna Jastrzebski, Jeroen P M Heijmans, Wipawadee Grenrum, Roderick L Beijersbergen, and Rene Bernards. Crispr knockout screening outperforms shrna and crispr in identifying essential genes. *Nat Biotechnol*, 34(6):631–3, Jun 2016.
- [109] Fernando Faunes, Penelope Hayward, Silvia Muñoz Descalzo, Sujash S Chatterjee, Tina Balayo, Jamie Trott, Andrew Christoforou, Anna Ferrer-Vaquero, Anna-Katerina Hadjantonakis, Ramanuj Dasgupta, and Alfonso Martinez Arias. A membrane-associated  $\beta$ -catenin/oct4 complex correlates with ground-state pluripotency in mouse embryonic stem cells. *Development*, 140(6):1171–83, Mar 2013.
- [110] Thomas G Fazio, Jason T Huff, and Barbara Panning. An rna screen of chromatin proteins identifies tip60-p400 as a regulator of embryonic stem cell identity. *Cell*, 134(1):162–74, Jul 2008.
- [111] Nicola Festuccia, Rodrigo Osorno, Florian Halbritter, Violetta Karwacki-Neisius, Pablo Navarro, Douglas Colby, Frederick Wong, Adam Yates, Simon R Tomlinson, and Ian Chambers. Esrrb is a direct nanog target gene that can substitute for nanog function in pluripotent cells. *Cell Stem Cell*, 11(4):477–90, Oct 2012.
- [112] Gabriella Ficz, Miguel R Branco, Stefanie Seisenberger, Fátima Santos, Felix Krueger, Timothy A Hore, C Joana Marques, Simon Andrews, and Wolf Reik. Dynamic regulation of 5-hydroxymethylcytosine in mouse es cells and during differentiation. *Nature*, 473(7347):398–402, May 2011.
- [113] Gabriella Ficz, Timothy A Hore, Fátima Santos, Heather J Lee, Wendy Dean, Julia Arand, Felix Krueger, David Oxley, Yu-Lee Paul, Jörn Walter, Simon J Cook, Simon

- Andrews, Miguel R Branco, and Wolf Reik. Fgf signaling inhibition in escs drives rapid genome-wide demethylation to the epigenetic ground state of pluripotency. *Cell Stem Cell*, 13(3):351–9, Sep 2013.
- [114] A Fire, S Xu, M K Montgomery, S A Kostas, S E Driver, and C C Mello. Potent and specific genetic interference by double-stranded rna in *caenorhabditis elegans*. *Nature*, 391(6669):806–11, Feb 1998.
- [115] Ron Firestein, Adam J Bass, So Young Kim, Ian F Dunn, Serena J Silver, Isil Guney, Ellen Freed, Azra H Ligon, Natalie Vena, Shuji Ogino, Milan G Chheda, Pablo Tamayo, Stephen Finn, Yashaswi Shrestha, Jesse S Boehm, Supriya Jain, Emeric Bojarski, Craig Mermel, Jordi Barretina, Jennifer A Chan, Jose Baselga, Josep Taberner, David E Root, Charles S Fuchs, Massimo Loda, Ramesh A Shivdasani, Matthew Meyerson, and William C Hahn. Cdk8 is a colorectal cancer oncogene that regulates beta-catenin activity. *Nature*, 455(7212):547–51, Sep 2008.
- [116] S E Fischer, E Wienholds, and R H Plasterk. Regulated transposition of a fish transposon in the mouse germ line. *Proc Natl Acad Sci U S A*, 98(12):6759–64, Jun 2001.
- [117] Rory J Flinn, Ying Yan, Sumanta Goswami, Peter J Parker, and Jonathan M Backer. The late endosome is essential for mtorc1 signaling. *Mol Biol Cell*, 21(5):833–41, Mar 2010.
- [118] Ariel Forrai, Kristy Boyle, Adam H Hart, Lynne Hartley, Steven Rakar, Tracy A Willson, Ken M Simpson, Andrew W Roberts, Warren S Alexander, Anne K Voss, and Lorraine Robb. Absence of suppressor of cytokine signalling 3 reduces self-renewal and promotes differentiation in murine embryonic stem cells. *Stem Cells*, 24(3):604–14, Mar 2006.
- [119] S L Forsburg. The art and design of genetic screens: yeast. *Nat Rev Genet*, 2(9):659–68, Sep 2001.
- [120] Kara M Foshay, Timothy J Looney, Sheila Chari, Frank Fuxiang Mao, Jae Hyun Lee, Li Zhang, Croydon J Fernandes, Samuel W Baker, Kayla L Clift, Jedidiah Gaetz, Chun-Guang Di, Andy Peng Xiang, and Bruce T Lahn. Embryonic stem cells induce pluripotency in somatic cell fusion through biphasic reprogramming. *Mol Cell*, 46(2):159–70, Apr 2012.

- [121] F Fougereousse, P Bullen, M Herasse, S Lindsay, I Richard, D Wilson, L Suel, M Durand, S Robson, M Abitbol, J S Beckmann, and T Strachan. Human-mouse differences in the embryonic expression patterns of developmental control genes and disease genes. *Hum Mol Genet*, 9(2):165–73, Jan 2000.
- [122] M J Fraser, L Cary, K Boonvisudhi, and H G Wang. Assay for movement of lepidopteran transposon ifp2 in insect cells using a baculovirus genome as a target dna. *Virology*, 211(2):397–407, Aug 1995.
- [123] Johannes M Freudenberg, Swati Ghosh, Brad L Lackford, Sailu Yellaboina, Xiaofeng Zheng, Ruifang Li, Suresh Cuddapah, Paul A Wade, Guang Hu, and Raja Jothi. Acute depletion of tet1-dependent 5-hydroxymethylcytosine levels impairs lif/stat3 signaling and results in loss of embryonic stem cell identity. *Nucleic Acids Res*, 40(8):3364–77, Apr 2012.
- [124] Maria A Frias, Carson C Thoreen, Jacob D Jaffe, Wayne Schroder, Tom Sculley, Steven A Carr, and David M Sabatini. msin1 is necessary for akt/pkb phosphorylation, and its isoforms define three distinct mtorc2s. *Curr Biol*, 16(18):1865–70, Sep 2006.
- [125] Richard L Frock, Jiazhi Hu, Robin M Meyers, Yu-Jui Ho, Erina Kii, and Frederick W Alt. Genome-wide detection of dna double-stranded breaks induced by engineered nucleases. *Nat Biotechnol*, 33(2):179–86, Feb 2015.
- [126] Yanfang Fu, Jennifer A Foden, Cyd Khayter, Morgan L Maeder, Deepak Reyon, J Keith Joung, and Jeffrey D Sander. High-frequency off-target mutagenesis induced by crispr-cas nucleases in human cells. *Nat Biotechnol*, 31(9):822–6, Sep 2013.
- [127] Yanfang Fu, Jeffrey D Sander, Deepak Reyon, Vincent M Cascio, and J Keith Joung. Improving crispr-cas nuclease specificity using truncated guide rnas. *Nat Biotechnol*, 32(3):279–84, Mar 2014.
- [128] Ohad Gafni, Leehee Weinberger, Abed AlFatah Mansour, Yair S Manor, Elad Chomsky, Dalit Ben-Yosef, Yael Kalma, Sergey Viukov, Itay Maza, Asaf Zviran, Yoach Rais, Zohar Shipony, Zohar Mukamel, Vladislav Krupalnik, Mirie Zerbib, Shay Geula, Inbal Caspi, Dan Schneir, Tamar Shwartz, Shlomit Gilad, Daniela Amann-Zalcenstein, Sima Benjamin, Ido Amit, Amos Tanay, Rada Massarwa, Noa Novershtern, and Jacob H Hanna. Derivation of novel human ground state naive pluripotent stem cells. *Nature*, 504(7479):282–6, Dec 2013.

- [129] Josiane E Garneau, Marie-Ève Dupuis, Manuela Villion, Dennis A Romero, Rodolphe Barrangou, Patrick Boyaval, Christophe Fremaux, Philippe Horvath, Alfonso H Magadán, and Sylvain Moineau. The crispr/cas bacterial immune system cleaves bacteriophage and plasmid dna. *Nature*, 468(7320):67–71, Nov 2010.
- [130] Giedrius Gasiunas, Rodolphe Barrangou, Philippe Horvath, and Virginijus Siksnys. Cas9-crRNA ribonucleoprotein complex mediates specific dna cleavage for adaptive immunity in bacteria. *Proc Natl Acad Sci U S A*, 109(39):E2579–86, Sep 2012.
- [131] D P Gearing, C J Thut, T VandeBos, S D Gimpel, P B Delaney, J King, V Price, D Cosman, and M P Beckmann. Leukemia inhibitory factor receptor is structurally related to the il-6 signal transducer, gp130. *EMBO J*, 10(10):2839–48, Oct 1991.
- [132] Shay Geula, Sharon Moshitch-Moshkovitz, Dan Dominissini, Abed AlFatah Mansour, Nitzan Kol, Mali Salmon-Divon, Vera Hershkovitz, Eyal Peer, Nofar Mor, Yair S Manor, Moshe Shay Ben-Haim, Eran Eyal, Sharon Yunger, Yishay Pinto, Diego Adhemar Jaitin, Sergey Viukov, Yoach Rais, Vladislav Krupalnik, Elad Chomsky, Mirie Zerbib, Itay Maza, Yoav Rechavi, Rada Massarwa, Suhair Hanna, Ido Amit, Erez Y Levanon, Ninette Amariglio, Noam Stern-Ginossar, Noa Novershtern, Gideon Rechavi, and Jacob H Hanna. Stem cells. m6a mrna methylation facilitates resolution of naïve pluripotency toward differentiation. *Science*, 347(6225):1002–6, Feb 2015.
- [133] Aron M Geurts, Gregory J Cost, Yevgeniy Freyvert, Bryan Zeitler, Jeffrey C Miller, Vivian M Choi, Shirin S Jenkins, Adam Wood, Xiaoxia Cui, Xiangdong Meng, Anna Vincent, Stephen Lam, Mieczyslaw Michalkiewicz, Rebecca Schilling, Jamie Foeckler, Shawn Kalloway, Hartmut Weiler, Séverine Ménoret, Ignacio Anegón, Gregory D Davis, Lei Zhang, Edward J Rebar, Philip D Gregory, Fyodor D Urnov, Howard J Jacob, and Roland Buelow. Knockout rats via embryo microinjection of zinc-finger nucleases. *Science*, 325(5939):433, Jul 2009.
- [134] Louis R Ghanem, Andrew Kromer, Ian M Silverman, Priya Chatterji, Elizabeth Traxler, Alfredo Penzo-Mendez, Mitchell J Weiss, Ben Z Stanger, and Stephen A Liebhaber. The poly(c) binding protein pcbp2 and its retrotransposed derivative pcbp1 are independently essential to mouse development. *Mol Cell Biol*, 36(2):304–19, Nov 2015.
- [135] Luke A Gilbert, Max A Horlbeck, Britt Adamson, Jacqueline E Villalta, Yuwen

- Chen, Evan H Whitehead, Carla Guimaraes, Barbara Panning, Hidde L Ploegh, Michael C Bassik, Lei S Qi, Martin Kampmann, and Jonathan S Weissman. Genome-scale crispr-mediated control of gene repression and activation. *Cell*, 159(3):647–61, Oct 2014.
- [136] Luke A Gilbert, Matthew H Larson, Leonardo Morsut, Zairan Liu, Gloria A Brar, Sandra E Torres, Noam Stern-Ginossar, Onn Brandman, Evan H Whitehead, Jennifer A Doudna, Wendell A Lim, Jonathan S Weissman, and Lei S Qi. Crispr-mediated modular rna-guided regulation of transcription in eukaryotes. *Cell*, 154(2):442–51, Jul 2013.
- [137] Julian A Gingold, Miguel Fidalgo, Diana Guallar, Zerlina Lau, Zhen Sun, Hongwei Zhou, Francesco Faiola, Xin Huang, Dung-Fang Lee, Avinash Waghray, Christoph Schaniel, Dan P Felsenfeld, Ihor R Lemischka, and Jianlong Wang. A genome-wide rna screen identifies opposing functions of *snai1* and *snai2* on the *nanog* dependency in reprogramming. *Mol Cell*, 56(1):140–52, Oct 2014.
- [138] A C Gingras, S P Gygi, B Raught, R D Polakiewicz, R T Abraham, M F Hoekstra, R Aebersold, and N Sonenberg. Regulation of 4e-bp1 phosphorylation: a novel two-step mechanism. *Genes Dev*, 13(11):1422–37, Jun 1999.
- [139] Aaron D Goldberg, Laura A Banaszynski, Kyung-Min Noh, Peter W Lewis, Simon J Elsaesser, Sonja Stadler, Scott Dewell, Martin Law, Xingyi Guo, Xuan Li, Duancheng Wen, Ariane Chapgier, Russell C DeKolver, Jeffrey C Miller, Ya-Li Lee, Elizabeth A Boydston, Michael C Holmes, Philip D Gregory, John M Grealley, Shahin Rafii, Chingwen Yang, Peter J Scambler, David Garrick, Richard J Gibbons, Douglas R Higgs, Ileana M Cristea, Fyodor D Urnov, Deyou Zheng, and C David Allis. Distinct factors control histone variant h3.3 localization at specific genomic regions. *Cell*, 140(5):678–91, Mar 2010.
- [140] J Groden, Y Nakamura, and J German. Molecular evidence that homologous recombination occurs in proliferating human somatic cells. *Proc Natl Acad Sci U S A*, 87(11):4315–9, Jun 1990.
- [141] David A Guertin and David M Sabatini. Defining the role of mtor in cancer. *Cancer Cell*, 12(1):9–22, Jul 2007.
- [142] Ge Guo, Yue Huang, Peter Humphreys, Xiaozhong Wang, and Austin Smith. A piggybac-based recessive screening method to identify pluripotency regulators. *PLoS*



- One*, 6(4):e18189, Apr 2011.
- [143] Ge Guo, Ferdinand von Meyenn, Fatima Santos, Yaoyao Chen, Wolf Reik, Paul Bertone, Austin Smith, and Jennifer Nichols. Naive pluripotent stem cells derived directly from isolated cells of the human inner cell mass. *Stem Cell Reports*, 6(4):437–46, Apr 2016.
- [144] Ge Guo, Wei Wang, and Allan Bradley. Mismatch repair genes identified using genetic screens in blm-deficient embryonic stem cells. *Nature*, 429(6994):891–5, Jun 2004.
- [145] Ankit Gupta, Ryan G Christensen, Amy L Rayla, Abirami Lakshmanan, Gary D Stormo, and Scot A Wolfe. An optimized two-finger archive for zfn-mediated gene targeting. *Nat Methods*, 9(6):588–90, Jun 2012.
- [146] Ehsan Habibi, Arie B Brinkman, Julia Arand, Leonie I Kroeze, Hindrik H D Kerstens, Filomena Matarese, Konstantin Lepikhov, Marta Gut, Isabelle Brun-Heath, Nina C Hubner, Rosaria Benedetti, Lucia Altucci, Joop H Jansen, Jörn Walter, Ivo G Gut, Hendrik Marks, and Hendrik G Stunnenberg. Whole-genome bisulfite sequencing of two distinct interconvertible dna methylomes of mouse embryonic stem cells. *Cell Stem Cell*, 13(3):360–9, Sep 2013.
- [147] Jamie A Hackett and M Azim Surani. Regulatory principles of pluripotency: from the ground state up. *Cell Stem Cell*, 15(4):416–30, Oct 2014.
- [148] John Hall, Ge Guo, Jason Wray, Isobel Eyres, Jennifer Nichols, Lars Grotewold, Sofia Morfopoulou, Peter Humphreys, William Mansfield, Rachael Walker, Simon Tomlinson, and Austin Smith. Oct4 and lif/stat3 additively induce krüppel factors to sustain embryonic stem cell self-renewal. *Cell Stem Cell*, 5(6):597–609, Dec 2009.
- [149] S M Hammond, E Bernstein, D Beach, and G J Hannon. An rna-directed nuclease mediates post-transcriptional gene silencing in drosophila cells. *Nature*, 404(6775):293–6, Mar 2000.
- [150] Kenta Hara, Yoshiko Maruki, Xiaomeng Long, Ken-ichi Yoshino, Noriko Oshiro, Sujuti Hidayat, Chiharu Tokunaga, Joseph Avruch, and Kazuyoshi Yonezawa. Raptor, a binding partner of target of rapamycin (tor), mediates tor action. *Cell*, 110(2):177–89, Jul 2002.

- [151] Serge Hardy, Vincent Legagneux, Yann Audic, and Luc Paillard. Reverse genetics in eukaryotes. *Biol Cell*, 102(10):561–80, Oct 2010.
- [152] Laura S Harrington, Greg M Findlay, and Richard F Lamb. Restraining pi3k: mtor signalling goes back to the membrane. *Trends Biochem Sci*, 30(1):35–42, Jan 2005.
- [153] Traver Hart, Megha Chandrashekhar, Michael Aregger, Zachary Steinhart, Kevin R Brown, Graham MacLeod, Monika Mis, Michal Zimmermann, Amelie Fradet-Turcotte, Song Sun, Patricia Mero, Peter Dirks, Sachdev Sidhu, Frederick P Roth, Olivia S Rissland, Daniel Durocher, Stephane Angers, and Jason Moffat. High-resolution crispr screens reveal fitness genes and genotype-specific cancer liabilities. *Cell*, 163(6):1515–26, Dec 2015.
- [154] Katsuhiko Hayashi, Susana M Chuva de Sousa Lopes, Fuchou Tang, Kaiqin Lao, and M Azim Surani. Dynamic equilibrium and heterogeneity of mouse pluripotent stem cells with distinct functional and epigenetic states. *Cell Stem Cell*, 3(4):391–401, Oct 2008.
- [155] Isaac B Hilton, Anthony M D’Ippolito, Christopher M Vockley, Pratiksha I Thakore, Gregory E Crawford, Timothy E Reddy, and Charles A Gersbach. Epigenome editing by a crispr-cas9-based acetyltransferase activates genes from promoters and enhancers. *Nat Biotechnol*, 33(5):510–7, May 2015.
- [156] S Hitotsumachi, D A Carpenter, and W L Russell. Dose-repetition increases the mutagenic effectiveness of n-ethyl-n-nitrosourea in mouse spermatogonia. *Proc Natl Acad Sci U S A*, 82(19):6619–21, Oct 1985.
- [157] Dirk Hockemeyer, Haoyi Wang, Samira Kiani, Christine S Lai, Qing Gao, John P Cassady, Gregory J Cost, Lei Zhang, Yolanda Santiago, Jeffrey C Miller, Bryan Zeitler, Jennifer M Cherone, Xiangdong Meng, Sarah J Hinkley, Edward J Rebar, Philip D Gregory, Fyodor D Urnov, and Rudolf Jaenisch. Genetic engineering of human pluripotent cells using tale nucleases. *Nat Biotechnol*, 29(8):731–4, Jul 2011.
- [158] Alex Hodgkins, Anna Farne, Sajith Perera, Tiago Grego, David J Parry-Smith, William C Skarnes, and Vivek Iyer. Wge: a crispr database for genome engineering. *Bioinformatics*, 31(18):3078–80, Sep 2015.
- [159] Marina K Holz, Bryan A Ballif, Steven P Gygi, and John Blenis. mtor and s6k1 mediate assembly of the translation preinitiation complex through dynamic protein interchange and ordered phosphorylation events. *Cell*, 123(4):569–80, Nov 2005.

- [160] Patrick D Hsu, Eric S Lander, and Feng Zhang. Development and applications of crispr-cas9 for genome engineering. *Cell*, 157(6):1262–78, Jun 2014.
- [161] Patrick D Hsu, David A Scott, Joshua A Weinstein, F Ann Ran, Silvana Konermann, Vineeta Agarwala, Yinqing Li, Eli J Fine, Xuebing Wu, Ophir Shalem, Thomas J Cradick, Luciano A Marraffini, Gang Bao, and Feng Zhang. Dna targeting specificity of rna-guided cas9 nucleases. *Nat Biotechnol*, 31(9):827–32, Sep 2013.
- [162] Guang Hu, Jonghwan Kim, Qikai Xu, Yumei Leng, Stuart H Orkin, and Stephen J Elledge. A genome-wide rna screen identifies a new transcriptional module required for self-renewal. *Genes Dev*, 23(7):837–48, Apr 2009.
- [163] Jingxiang Huang, Christian C Dibble, Mika Matsuzaki, and Brendan D Manning. The tsc1-tsc2 complex is required for proper activation of mtor complex 2. *Mol Cell Biol*, 28(12):4104–15, Jun 2008.
- [164] Woong Y Hwang, Yanfang Fu, Deepak Reyon, Morgan L Maeder, Shengdar Q Tsai, Jeffrey D Sander, Randall T Peterson, J-R Joanna Yeh, and J Keith Joung. Efficient genome editing in zebrafish using a crispr-cas system. *Nat Biotechnol*, 31(3):227–9, Mar 2013.
- [165] Ken Inoki, Yong Li, Tianquan Zhu, Jun Wu, and Kun-Liang Guan. Tsc2 is phosphorylated and inhibited by akt and suppresses mtor signalling. *Nat Cell Biol*, 4(9):648–57, Sep 2002.
- [166] Ken Inoki, Tianqing Zhu, and Kun-Liang Guan. Tsc2 mediates cellular energy response to control cell growth and survival. *Cell*, 115(5):577–90, Nov 2003.
- [167] M Isalan, Y Choo, and A Klug. Synergy between adjacent zinc fingers in sequence-specific dna recognition. *Proc Natl Acad Sci U S A*, 94(11):5617–21, May 1997.
- [168] Y Ishino, H Shinagawa, K Makino, M Amemura, and A Nakata. Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in escherichia coli, and identification of the gene product. *J Bacteriol*, 169(12):5429–33, Dec 1987.
- [169] Shinsuke Ito, Ana C D’Alessio, Olena V Taranova, Kwonho Hong, Lawrence C Sowers, and Yi Zhang. Role of tet proteins in 5mc to 5hmc conversion, es-cell self-renewal and inner cell mass specification. *Nature*, 466(7310):1129–33, Aug 2010.

- [170] Natalia Ivanova, Radu Dobrin, Rong Lu, Iulia Kotenko, John Levorse, Christina DeCoste, Xenia Schafer, Yi Lun, and Ihor R Lemischka. Dissecting self-renewal in stem cells with rna interference. *Nature*, 442(7102):533–8, Aug 2006.
- [171] Z Ivics, P B Hackett, R H Plasterk, and Z Izsvák. Molecular reconstruction of sleeping beauty, a tc1-like transposon from fish, and its transposition in human cells. *Cell*, 91(4):501–10, Nov 1997.
- [172] Ruud Jansen, Jan D A van Embden, Wim Gaastra, and Leo M Schouls. Identification of genes that are associated with dna repeats in prokaryotes. *Mol Microbiol*, 43(6):1565–75, Mar 2002.
- [173] Xinjun Ji, Jian Kong, and Stephen A Liebhaber. An rna-protein complex links enhanced nuclear 3' processing with cytoplasmic mrna stabilization. *EMBO J*, 30(13):2622–33, May 2011.
- [174] Fuguo Jiang and Jennifer A Doudna. Crispr-cas9 structures and mechanisms. *Annu Rev Biophys*, 46:505–529, May 2017.
- [175] N Jinawath, C Vasoontara, K-L Yap, M M Thiaville, K Nakayama, T-L Wang, and I-M Shih. Nac-1, a potential stem cell pluripotency factor, contributes to paclitaxel resistance in ovarian cancer through inactivating gadd45 pathway. *Oncogene*, 28(18):1941–8, May 2009.
- [176] Martin Jinek, Krzysztof Chylinski, Ines Fonfara, Michael Hauer, Jennifer A Doudna, and Emmanuelle Charpentier. A programmable dual-rna-guided dna endonuclease in adaptive bacterial immunity. *Science*, 337(6096):816–21, Aug 2012.
- [177] Martin Jinek, Fuguo Jiang, David W Taylor, Samuel H Sternberg, Emine Kaya, Enbo Ma, Carolin Anders, Michael Hauer, Kaihong Zhou, Steven Lin, Matias Kaplan, Anthony T Iavarone, Emmanuelle Charpentier, Eva Nogales, and Jennifer A Doudna. Structures of cas9 endonucleases reveal rna-mediated conformational activation. *Science*, 343(6176):1247997, Mar 2014.
- [178] Erik M Jorgensen and Susan E Mango. The art and design of genetic screens: caenorhabditis elegans. *Nat Rev Genet*, 3(5):356–69, May 2002.
- [179] J Keith Joung and Jeffry D Sander. Talens: a widely applicable technology for targeted genome editing. *Nat Rev Mol Cell Biol*, 14(1):49–55, Jan 2013.

- [180] Jennifer Jung, Heide Marika Genau, and Christian Behrends. Amino acid-dependent mtorc1 regulation by the lysosomal membrane protein slc38a9. *Mol Cell Biol*, 35(14):2479–94, Jul 2015.
- [181] M J Justice, J K Noveroske, J S Weber, B Zheng, and A Bradley. Mouse enu mutagenesis. *Hum Mol Genet*, 8(10):1955–63, 1999.
- [182] Michael H Kagey, Jamie J Newman, Steve Bilodeau, Ye Zhan, David A Orlando, Nynke L van Berkum, Christopher C Ebmeier, Jesse Goossens, Peter B Rahl, Stuart S Levine, Dylan J Taatjes, Job Dekker, and Richard A Young. Mediator and cohesin connect gene expression and chromatin architecture. *Nature*, 467(7314):430–5, Sep 2010.
- [183] Keisuke Kaji, Isabel Martín Caballero, Ruth MacLeod, Jennifer Nichols, Valerie A Wilson, and Brian Hendrich. The nurd component mbd3 is required for pluripotency of embryonic stem cells. *Nat Cell Biol*, 8(3):285–92, Mar 2006.
- [184] Tüzér Kalkan, Nelly Olova, Mila Roode, Carla Mulas, Heather J Lee, Isabelle Nett, Hendrik Marks, Rachael Walker, Hendrik G Stunnenberg, Kathryn S Lilley, Jennifer Nichols, Wolf Reik, Paul Bertone, and Austin Smith. Tracking the embryonic stem cell transition from ground state pluripotency. *Development*, 144(7):1221–1234, Apr 2017.
- [185] Tüzér Kalkan and Austin Smith. Mapping the route from naive pluripotency to lineage specification. *Philos Trans R Soc Lond B Biol Sci*, 369(1657), Dec 2014.
- [186] Tibor Kalmar, Chea Lim, Penelope Hayward, Silvia Muñoz-Descalzo, Jennifer Nichols, Jordi Garcia-Ojalvo, and Alfonso Martinez Arias. Regulated fluctuations in nanog expression mediate cell fate decisions in embryonic stem cells. *PLoS Biol*, 7(7):e1000149, Jul 2009.
- [187] Chryssa Kanellopoulou, Stefan A Muljo, Andrew L Kung, Shridar Ganesan, Ronny Drapkin, Thomas Jenuwein, David M Livingston, and Klaus Rajewsky. Dicer-deficient mouse embryonic stem cells are defective in differentiation and centromeric silencing. *Genes Dev*, 19(4):489–501, Feb 2005.
- [188] Violetta Karwacki-Neisius, Jonathan Göke, Rodrigo Osorno, Florian Halbritter, Jia Hui Ng, Andrea Y Weiße, Frederick C K Wong, Alessia Gagliardi, Nicholas P Mullin, Nicola Festuccia, Douglas Colby, Simon R Tomlinson, Huck-Hui Ng, and Ian Chambers. Reduced oct4 expression directs a robust pluripotent state with distinct

- signaling activity and increased enhancer occupancy by oct4 and nanog. *Cell Stem Cell*, 12(5):531–45, May 2013.
- [189] Nicola A Kearns, Hannah Pham, Barbara Tabak, Ryan M Genga, Noah J Silverstein, Manuel Garber, and René Maehr. Functional annotation of native enhancers with a cas9-histone demethylase fusion. *Nat Methods*, 12(5):401–3, May 2015.
- [190] James Kehler, Elena Tolkunova, Birgit Koschorz, Maurizio Pesce, Luca Gentile, Michele Boiani, Hilda Lomelí, Andras Nagy, K John McLaughlin, Hans R Schöler, and Alexey Tomilin. Oct4 is required for primordial germ cell survival. *EMBO Rep*, 5(11):1078–83, Nov 2004.
- [191] Maria Keramari, Janet Razavi, Karen A Ingman, Christoph Patsch, Frank Edenhofer, Christopher M Ward, and Susan J Kimber. Sox2 is essential for formation of trophectoderm in the preimplantation embryo. *PLoS One*, 5(11):e13952, Nov 2010.
- [192] Menno F Kielman, Maaret Rindapää, Claudia Gaspar, Nicole van Poppel, Cor Breukel, Sandra van Leeuwen, Makoto Mark Taketo, Scott Roberts, Ron Smits, and Riccardo Fodde. Apc modulates embryonic stem-cell differentiation by controlling the dosage of beta-catenin signaling. *Nat Genet*, 32(4):594–605, Dec 2002.
- [193] Benjamin T Kile and Douglas J Hilton. The art and design of genetic screens: mouse. *Nat Rev Genet*, 6(7):557–67, Jul 2005.
- [194] Eunjung Kim, Pankuri Goraksha-Hicks, Li Li, Thomas P Neufeld, and Kun-Liang Guan. Regulation of torc1 by rag gtpases in nutrient response. *Nat Cell Biol*, 10(8):935–45, Aug 2008.
- [195] Hyongbum Kim and Jin-Soo Kim. A guide to genome engineering with programmable nucleases. *Nat Rev Genet*, 15(5):321–34, May 2014.
- [196] Joungmok Kim, Mondira Kundu, Benoit Viollet, and Kun-Liang Guan. Ampk and mtor regulate autophagy through direct phosphorylation of ulk1. *Nat Cell Biol*, 13(2):132–41, Feb 2011.
- [197] Y G Kim, J Cha, and S Chandrasegaran. Hybrid restriction enzymes: zinc finger fusions to fok i cleavage domain. *Proc Natl Acad Sci U S A*, 93(3):1156–60, Feb 1996.
- [198] Tyler S Klann, Joshua B Black, Malathi Chellappan, Alexias Safi, Lingyun Song, Isaac B Hilton, Gregory E Crawford, Timothy E Reddy, and Charles A Gersbach.

- Crispr-cas9 epigenome editing enables high-throughput screening for functional regulatory elements in the human genome. *Nat Biotechnol*, 35(6):561–568, Jun 2017.
- [199] Matthew T Knuesel, Krista D Meyer, Carrie Bernecky, and Dylan J Taatjes. The human cdk8 subcomplex is a molecular switch that controls mediator coactivator function. *Genes Dev*, 23(4):439–51, Feb 2009.
- [200] A Koga, M Suzuki, H Inagaki, Y Bessho, and H Hori. Transposable element in fish. *Nature*, 383(6595):30, Sep 1996.
- [201] Kian Peng Koh, Akiko Yabuuchi, Sridhar Rao, Yun Huang, Kerriane Cunniff, Julie Nardone, Asta Laiho, Mamta Tahiliani, Cesar A Sommer, Gustavo Mostoslavsky, Riitta Lahesmaa, Stuart H Orkin, Scott J Rodig, George Q Daley, and Anjana Rao. Tet1 and tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells. *Cell Stem Cell*, 8(2):200–13, Feb 2011.
- [202] Hiroko Koike-Yusa, Yilong Li, E-Pien Tan, Martin Del Castillo Velasco-Herrera, and Kosuke Yusa. Genome-wide recessive genetic screening in mammalian cells with a lentiviral crispr-guide rna library. *Nat Biotechnol*, 32(3):267–73, Mar 2014.
- [203] Aleksandra A Kolodziejczyk, Jong Kyoung Kim, Jason C H Tsang, Tomislav Ilicic, Johan Henriksson, Kedar N Natarajan, Alex C Tuck, Xuefei Gao, Marc Bühler, Pentao Liu, John C Marioni, and Sarah A Teichmann. Single cell rna-sequencing of pluripotent states unlocks modular transcriptional variation. *Cell Stem Cell*, 17(4):471–85, Oct 2015.
- [204] Silvana Konermann, Mark D Brigham, Alexandro E Trevino, Patrick D Hsu, Matthias Heidenreich, Le Cong, Randall J Platt, David A Scott, George M Church, and Feng Zhang. Optical control of mammalian endogenous transcription and epigenetic states. *Nature*, 500(7463):472–6, Aug 2013.
- [205] Silvana Konermann, Mark D Brigham, Alexandro E Trevino, Julia Joung, Omar O Abudayyeh, Clea Barcena, Patrick D Hsu, Naomi Habib, Jonathan S Gootenberg, Hiroshi Nishimasu, Osamu Nureki, and Feng Zhang. Genome-scale transcriptional activation by an engineered crispr-cas9 complex. *Nature*, 517(7536):583–8, Jan 2015.
- [206] Renate König, Yingyao Zhou, Daniel Elleder, Tracy L Diamond, Ghislain M C Bonamy, Jeffrey T Irelan, Chih-Yuan Chiang, Buu P Tu, Paul D De Jesus, Caroline E Lilley, Shannon Seidel, Amanda M Opaluch, Jeremy S Caldwell, Matthew D Weitzman, Kelli L Kuhen, Sourav Bandyopadhyay, Trey Ideker, Anthony P Orth,

- Loren J Miraglia, Frederic D Bushman, John A Young, and Sumit K Chanda. Global analysis of host-pathogen interactions that regulate early-stage hiv-1 replication. *Cell*, 135(1):49–60, Oct 2008.
- [207] Gozde Korkmaz, Rui Lopes, Alejandro P Ugalde, Ekaterina Nevedomskaya, Ruiqi Han, Ksenia Myacheva, Wilbert Zwart, Ran Elkon, and Reuven Agami. Functional genetic screens for enhancer elements in the human genome using crispr-cas9. *Nat Biotechnol*, 34(2):192–8, Feb 2016.
- [208] M Kotecki, P S Reddy, and B H Cochran. Isolation and characterization of a near-haploid human cell line. *Exp Cell Res*, 252(2):273–80, Nov 1999.
- [209] Tilo Kunath, Danielle Arnaud, Gary D Uy, Ikuhiro Okamoto, Corinne Chureau, Yojiro Yamanaka, Edith Heard, Richard L Gardner, Philip Avner, and Janet Rossant. Imprinted x-inactivation in extra-embryonic endoderm cell lines from mouse blastocysts. *Development*, 132(7):1649–61, Apr 2005.
- [210] Tilo Kunath, Marc K Saba-El-Leil, Marwa Almousailleakh, Jason Wray, Sylvain Meloche, and Austin Smith. Fgf stimulation of the erk1/2 signalling cascade triggers transition of pluripotent embryonic stem cells from self-renewal to lineage commitment. *Development*, 134(16):2895–902, Aug 2007.
- [211] Cem Kuscu, Sevki Arslan, Ritambhara Singh, Jeremy Thorpe, and Mazhar Adli. Genome-wide analysis reveals characteristics of off-target sites bound by the cas9 endonuclease. *Nat Biotechnol*, 32(7):677–83, Jul 2014.
- [212] Dudley W Lamming, Lan Ye, Pekka Katajisto, Marcus D Goncalves, Maki Saitoh, Deanna M Stevens, James G Davis, Adam B Salmon, Arlan Richardson, Rexford S Ahima, David A Guertin, David M Sabatini, and Joseph A Baur. Rapamycin-induced insulin resistance is mediated by mtorc2 loss and uncoupled from longevity. *Science*, 335(6076):1638–43, Mar 2012.
- [213] C P Landel, S Z Chen, and G A Evans. Reverse genetics using transgenic mice. *Annu Rev Physiol*, 52:841–51, 1990.
- [214] Fredrik Lanner and Janet Rossant. The role of fgf/erk signaling in pluripotent cells. *Development*, 137(20):3351–60, Oct 2010.
- [215] Mathieu Laplante and David M Sabatini. mtor signaling at a glance. *J Cell Sci*, 122(Pt 20):3589–94, Oct 2009.



- [216] Bum-Kyu Lee, Wenwen Shen, Jiwoon Lee, Catherine Rhee, Haewon Chung, Kun-Yong Kim, In-Hyun Park, and Jonghwan Kim. Tgif1 counterbalances the activity of core pluripotency factors in mouse embryonic stem cells. *Cell Rep*, 13(1):52–60, Oct 2015.
- [217] Martin Leeb, Sabine Dietmann, Maike Paramor, Hitoshi Niwa, and Austin Smith. Genetic exploration of the exit from self-renewal using haploid embryonic stem cells. *Cell Stem Cell*, 14(3):385–93, Mar 2014.
- [218] Martin Leeb, Diego Pasini, Maria Novatchkova, Markus Jaritz, Kristian Helin, and Anton Wutz. Polycomb complexes act redundantly to repress genomic repeats and genes. *Genes Dev*, 24(3):265–76, Feb 2010.
- [219] Martin Leeb and Anton Wutz. Derivation of haploid embryonic stem cells from mouse embryos. *Nature*, 479(7371):131–4, Sep 2011.
- [220] M J Lenardo, L Staudt, P Robbins, A Kuang, R C Mulligan, and D Baltimore. Repression of the igh enhancer in teratocarcinoma cells associated with a novel octamer factor. *Science*, 243(4890):544–6, Jan 1989.
- [221] Chang Li, Ximmeng Guan, Tao Du, Wei Jin, Biao Wu, Yalan Liu, Ping Wang, Bodan Hu, George E Griffin, Robin J Shattock, and Qinxue Hu. Inhibition of hiv-1 infection of primary cd4+ t-cells by gene editing of ccr5 using adenovirus-delivered crispr/cas9. *J Gen Virol*, 96(8):2381–93, Aug 2015.
- [222] Hongmei Lisa Li, Naoko Fujimoto, Noriko Sasakawa, Saya Shirai, Tokiko Ohkame, Tetsushi Sakuma, Michihiro Tanaka, Naoki Amano, Akira Watanabe, Hidetoshi Sakurai, Takashi Yamamoto, Shinya Yamanaka, and Akitsu Hotta. Precise correction of the dystrophin gene in duchenne muscular dystrophy patient induced pluripotent stem cells by talen and crispr-cas9. *Stem Cell Reports*, 4(1):143–54, Jan 2015.
- [223] Mo Li, Guang-Hui Liu, and Juan Carlos Izpisua Belmonte. Navigating the epigenetic landscape of pluripotent stem cells. *Nat Rev Mol Cell Biol*, 13(8):524–35, Jul 2012.
- [224] Tangliang Li, Yue Shi, Pei Wang, Luis Miguel Guachalla, Baofa Sun, Tjard Joerss, Yu-Sheng Chen, Marco Groth, Anja Krueger, Matthias Platzer, Yun-Gui Yang, Karl Lenhard Rudolph, and Zhao-Qi Wang. Smg6/est1 licenses embryonic stem cell differentiation via nonsense-mediated mrna decay. *EMBO J*, 34(12):1630–47, Jun 2015.

- [225] Wei Li, Han Xu, Tengfei Xiao, Le Cong, Michael I Love, Feng Zhang, Rafael A Irizarry, Jun S Liu, Myles Brown, and X Shirley Liu. Mageck enables robust identification of essential genes from genome-scale crispr/cas9 knockout screens. *Genome Biol*, 15(12):554, 2014.
- [226] Jiancong Liang, Ma Wan, Yi Zhang, Peili Gu, Huawei Xin, Sung Yun Jung, Jun Qin, Jiemin Wong, Austin J Cooney, Dan Liu, and Zhou Songyang. Nanog and oct4 associate with unique transcriptional repression complexes in embryonic stem cells. *Nat Cell Biol*, 10(6):731–9, Jun 2008.
- [227] X Lin, E M Buff, N Perrimon, and A M Michelson. Heparan sulfate proteoglycans are essential for fgf receptor signaling during drosophila embryonic development. *Development*, 126(17):3715–23, Sep 1999.
- [228] S John Liu, Max A Horlbeck, Seung Woo Cho, Harjus S Birk, Martina Malatesta, Daniel He, Frank J Attenello, Jacqueline E Villalta, Min Y Cho, Yuwen Chen, Mohammad A Mandegar, Michael P Olvera, Luke A Gilbert, Bruce R Conklin, Howard Y Chang, Jonathan S Weissman, and Daniel A Lim. Crispri-based genome-scale identification of functional long noncoding rna loci in human cells. *Science*, 355(6320), 01 2017.
- [229] X Shawn Liu, Hao Wu, Xiong Ji, Yonatan Stelzer, Xuebing Wu, Szymon Czauderna, Jian Shu, Daniel Dadon, Richard A Young, and Rudolf Jaenisch. Editing dna methylation in the mammalian genome. *Cell*, 167(1):233–247.e17, Sep 2016.
- [230] Alan Lloyd, Christopher L Plaisier, Dana Carroll, and Gary N Drews. Targeted mutagenesis using zinc-finger nucleases in arabidopsis. *Proc Natl Acad Sci U S A*, 102(6):2232–7, Feb 2005.
- [231] David A F Loebel, Catherine M Watson, R Andrea De Young, and Patrick P L Tam. Lineage choice and differentiation in mouse embryos and embryonic stem cells. *Dev Biol*, 264(1):1–14, Dec 2003.
- [232] Yui-Han Loh, Qiang Wu, Joon-Lin Chew, Vinsensius B Vega, Weiwei Zhang, Xi Chen, Guillaume Bourque, Joshy George, Bernard Leong, Jun Liu, Kee-Yew Wong, Ken W Sung, Charlie W H Lee, Xiao-Dong Zhao, Kuo-Ping Chiu, Leonard Lipovich, Vladimir A Kuznetsov, Paul Robson, Lawrence W Stanton, Chia-Lin Wei, Yijun Ruan, Bing Lim, and Huck-Hui Ng. The oct4 and nanog transcription network

- regulates pluripotency in mouse embryonic stem cells. *Nat Genet*, 38(4):431–40, Apr 2006.
- [233] Yui-Han Loh, Weiwei Zhang, Xi Chen, Joshy George, and Huck-Hui Ng. Jmjd1a and jmjd2c histone h3 lys 9 demethylases regulate self-renewal in embryonic stem cells. *Genes Dev*, 21(20):2545–57, Oct 2007.
- [234] Angelo Lombardo, Pietro Genovese, Christian M Beausejour, Silvia Colleoni, Ya-Li Lee, Kenneth A Kim, Dale Ando, Fyodor D Urnov, Cesare Galli, Philip D Gregory, Michael C Holmes, and Luigi Naldini. Gene editing in human stem cells using zinc finger nucleases and integrase-defective lentiviral vector delivery. *Nat Biotechnol*, 25(11):1298–306, Nov 2007.
- [235] Xiaomeng Long, Yenshou Lin, Sara Ortiz-Vega, Kazuyoshi Yonezawa, and Joseph Avruch. Rheb binds and regulates the mtor kinase. *Curr Biol*, 15(8):702–13, Apr 2005.
- [236] Chanchao Lorthongpanich, Daniel M Messerschmidt, Siew Wee Chan, Wanjin Hong, Barbara B Knowles, and Davor Solter. Temporal reduction of lats kinases in the early preimplantation embryo prevents icm lineage differentiation. *Genes Dev*, 27(13):1441–6, Jul 2013.
- [237] Xinyi Lu, Jonathan Göke, Friedrich Sachs, Pierre-Étienne Jacques, Hongqing Liang, Bo Feng, Guillaume Bourque, Paula A Bubulya, and Huck-Hui Ng. Son connects the splicing-regulatory network with pluripotency in human embryonic stem cells. *Nat Cell Biol*, 15(10):1141–1152, Oct 2013.
- [238] Yu Lu, Yui-Han Loh, Hu Li, Marcella Cesana, Scott B Ficarro, Jignesh R Parikh, Nathan Salomonis, Cheng-Xu Delon Toh, Stelios T Andreadis, C John Luckey, James J Collins, George Q Daley, and Jarrod A Marto. Alternative splicing of mbd2 supports self-renewal in human pluripotent stem cells. *Cell Stem Cell*, 15(1):92–101, Jul 2014.
- [239] Biao Luo, Hiu Wing Cheung, Aravind Subramanian, Tanaz Sharifnia, Michael Okamoto, Xiaoping Yang, Greg Hinkle, Jesse S Boehm, Rameen Beroukhim, Barbara A Weir, Craig Mermel, David A Barbie, Tarif Awad, Xiaochuan Zhou, Tuyen Nguyen, Bruno Piqani, Cheng Li, Todd R Golub, Matthew Meyerson, Nir Hacohen, William C Hahn, Eric S Lander, David M Sabatini, and David E Root. Highly

- parallel identification of essential genes in cancer cells. *Proc Natl Acad Sci U S A*, 105(51):20380–5, Dec 2008.
- [240] G Luo, Z Ivics, Z Izsvák, and A Bradley. Chromosomal transposition of a *tc1*/mariner-like element in mouse embryonic stem cells. *Proc Natl Acad Sci U S A*, 95(18):10769–73, Sep 1998.
- [241] G Luo, I M Santoro, L D McDaniel, I Nishijima, M Mills, H Youssoufian, H Vogel, R A Schultz, and A Bradley. Cancer predisposition caused by elevated mitotic recombination in bloom mice. *Nat Genet*, 26(4):424–9, Dec 2000.
- [242] J Luo, R Sladek, J A Bader, A Matthyssen, J Rossant, and V Giguère. Placental abnormalities in mouse embryos lacking the orphan nuclear receptor *err-beta*. *Nature*, 388(6644):778–82, Aug 1997.
- [243] Morgan L Maeder, Samantha J Linder, Vincent M Cascio, Yanfang Fu, Quan H Ho, and J Keith Joung. Crispr rna-guided activation of endogenous human genes. *Nat Methods*, 10(10):977–9, Oct 2013.
- [244] Morgan L Maeder, Stacey Thibodeau-Beganny, Anna Osiak, David A Wright, Reshma M Anthony, Magdalena Eichinger, Tao Jiang, Jonathan E Foley, Ronnie J Winfrey, Jeffrey A Townsend, Erica Unger-Wallace, Jeffrey D Sander, Felix Müller-Lerch, Fengli Fu, Joseph Pearlberg, Carl Göbel, Justin P Dassie, Shondra M Pruett-Miller, Matthew H Porteus, Dennis C Sgroi, A John Iafrate, Drena Dobbs, Paul B McCray, Jr, Toni Cathomen, Daniel F Voytas, and J Keith Joung. Rapid “open-source” engineering of customized zinc-finger nucleases for highly efficient gene modification. *Mol Cell*, 31(2):294–301, Jul 2008.
- [245] Amanda Nga-Sze Mak, Philip Bradley, Raul A Cernadas, Adam J Bogdanove, and Barry L Stoddard. The crystal structure of tal effector *pthxo1* bound to its dna target. *Science*, 335(6069):716–9, Feb 2012.
- [246] Baldwin C Mak, Ken-Ichi Takemaru, Heidi L Kenerson, Randall T Moon, and Raymond S Yeung. The tuberin-hamartin complex negatively regulates beta-catenin signaling activity. *J Biol Chem*, 278(8):5947–51, Feb 2003.
- [247] Kira S Makarova, Nick V Grishin, Svetlana A Shabalina, Yuri I Wolf, and Eugene V Koonin. A putative rna-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic rnai, and hypothetical mechanisms of action. *Biol Direct*, 1:7, Mar 2006.

- [248] Kira S Makarova, Yuri I Wolf, Omer S Alkhnbashi, Fabrizio Costa, Shiraz A Shah, Sita J Saunders, Rodolphe Barrangou, Stan J J Brouns, Emmanuelle Charpentier, Daniel H Haft, Philippe Horvath, Sylvain Moineau, Francisco J M Mojica, Rebecca M Terns, Michael P Terns, Malcolm F White, Alexander F Yakunin, Roger A Garrett, John van der Oost, Rolf Backofen, and Eugene V Koonin. An updated evolutionary classification of crispr-cas systems. *Nat Rev Microbiol*, 13(11):722–36, 11 2015.
- [249] Prashant Mali, John Aach, P Benjamin Stranges, Kevin M Esvelt, Mark Moosburner, Sriram Kosuri, Luhan Yang, and George M Church. Cas9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat Biotechnol*, 31(9):833–8, Sep 2013.
- [250] Prashant Mali, Luhan Yang, Kevin M Esvelt, John Aach, Marc Guell, James E DiCarlo, Julie E Norville, and George M Church. Rna-guided human genome engineering via cas9. *Science*, 339(6121):823–6, Feb 2013.
- [251] Robert T Manguso, Hans W Pope, Margaret D Zimmer, Flavian D Brown, Kathleen B Yates, Brian C Miller, Natalie B Collins, Kevin Bi, Martin W LaFleur, Vikram R Juneja, Sarah A Weiss, Jennifer Lo, David E Fisher, Diana Miao, Eliezer Van Allen, David E Root, Arlene H Sharpe, John G Doench, and W Nicholas Haining. In vivo crispr screening identifies ptpn2 as a cancer immunotherapy target. *Nature*, 547(7664):413–418, Jul 2017.
- [252] Brendan D Manning. Balancing akt with s6k: implications for both metabolic diseases and tumorigenesis. *J Cell Biol*, 167(3):399–403, Nov 2004.
- [253] Brendan D Manning, Andrew R Tee, M Nicole Logsdon, John Blenis, and Lewis C Cantley. Identification of the tuberous sclerosis complex-2 tumor suppressor gene product tuberlin as a target of the phosphoinositide 3-kinase/akt pathway. *Mol Cell*, 10(1):151–62, Jul 2002.
- [254] S L Mansour, K R Thomas, and M R Capecchi. Disruption of the proto-oncogene int-2 in mouse embryo-derived stem cells: a general strategy for targeting mutations to non-selectable genes. *Nature*, 336(6197):348–52, Nov 1988.
- [255] Luciano A Marraffini and Erik J Sontheimer. Crispr interference limits horizontal gene transfer in staphylococci by targeting dna. *Science*, 322(5909):1843–5, Dec 2008.

- [256] Graziano Martello. Let's sp(1)ice up pluripotency! *EMBO J*, 32(22):2903–4, Nov 2013.
- [257] Graziano Martello, Paul Bertone, and Austin Smith. Identification of the missing pluripotency mediator downstream of leukaemia inhibitory factor. *EMBO J*, 32(19):2561–74, Oct 2013.
- [258] Graziano Martello and Austin Smith. The nature of embryonic stem cells. *Annu Rev Cell Dev Biol*, 30:647–75, 2014.
- [259] Graziano Martello, Toshimi Sugimoto, Evangelia Diamanti, Anagha Joshi, Rebecca Hannah, Satoshi Ohtsuka, Berthold Göttgens, Hitoshi Niwa, and Austin Smith. Esrrb is a pivotal target of the gsk3/tcf3 axis regulating embryonic stem cell self-renewal. *Cell Stem Cell*, 11(4):491–504, Oct 2012.
- [260] G R Martin. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc Natl Acad Sci U S A*, 78(12):7634–8, Dec 1981.
- [261] J Massagué. Tgf-beta signal transduction. *Annu Rev Biochem*, 67:753–91, 1998.
- [262] Shinji Masui, Yuhki Nakatake, Yayoi Toyooka, Daisuke Shimosato, Rika Yagi, Kazue Takahashi, Hitoshi Okochi, Akihiko Okuda, Ryo Matoba, Alexei A Sharov, Minoru S H Ko, and Hitoshi Niwa. Pluripotency governed by sox2 via regulation of oct3/4 expression in mouse embryonic stem cells. *Nat Cell Biol*, 9(6):625–35, Jun 2007.
- [263] Shinji Masui, Satoshi Ohtsuka, Rika Yagi, Kadue Takahashi, Minoru S H Ko, and Hitoshi Niwa. Rex1/zfp42 is dispensable for pluripotency in mouse es cells. *BMC Dev Biol*, 8:45, Apr 2008.
- [264] T Matsuda, T Nakamura, K Nakao, T Arai, M Katsuki, T Heike, and T Yokota. Stat3 activation is sufficient to maintain an undifferentiated state of mouse embryonic stem cells. *EMBO J*, 18(15):4261–9, Aug 1999.
- [265] Y Matsui, K Zsebo, and B L Hogan. Derivation of pluripotential embryonic stem cells from murine primordial germ cells in culture. *Cell*, 70(5):841–7, Sep 1992.
- [266] Alexander Meissner. Epigenetic modifications in pluripotent and differentiated cells. *Nat Biotechnol*, 28(10):1079–88, Oct 2010.
- [267] Alexander Meissner, Tarjei S Mikkelsen, Hongcang Gu, Marius Wernig, Jacob Hanna, Andrey Sivachenko, Xiaolan Zhang, Bradley E Bernstein, Chad Nusbaum,

- David B Jaffe, Andreas Gnirke, Rudolf Jaenisch, and Eric S Lander. Genome-scale dna methylation maps of pluripotent and differentiated cells. *Nature*, 454(7205):766–70, Aug 2008.
- [268] Xiangdong Meng, Marcus B Noyes, Lihua J Zhu, Nathan D Lawson, and Scot A Wolfe. Targeted gene inactivation in zebrafish using engineered zinc-finger nucleases. *Nat Biotechnol*, 26(6):695–701, Jun 2008.
- [269] B J Merrill, U Gat, R DasGupta, and E Fuchs. Tcf3 and lef1 regulate lineage differentiation of multipotent stem cells in skin. *Genes Dev*, 15(13):1688–705, Jul 2001.
- [270] Bradley J Merrill, H Amalia Pasolli, Lisa Polak, Michael Rendl, Maria J García-García, Kathryn V Anderson, and Elaine Fuchs. Tcf3: a transcriptional regulator of axis induction in the early embryo. *Development*, 131(2):263–74, Jan 2004.
- [271] Eran Meshorer, Dhananjay Yellajoshula, Eric George, Peter J Scambler, David T Brown, and Tom Misteli. Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells. *Dev Cell*, 10(1):105–16, Jan 2006.
- [272] Emmanouil Metzakopian, Alex Strong, Vivek Iyer, Alex Hodgkins, Konstantinos Tzelepis, Liliana Antunes, Mathias J Friedrich, Qiaohua Kang, Teresa Davidson, Jacob Lamberth, Christina Hoffmann, Gregory D Davis, George S Vassiliou, William C Skarnes, and Allan Bradley. Enhancing the genome editing toolbox: genome wide crispr arrayed libraries. *Sci Rep*, 7(1):2244, May 2017.
- [273] J Miller, A D McLachlan, and A Klug. Repetitive zinc-binding domains in the protein transcription factor iiiA from xenopus oocytes. *EMBO J*, 4(6):1609–14, Jun 1985.
- [274] J C Miller and C O Pabo. Rearrangement of side-chains in a zif268 mutant highlights the complexities of zinc finger-dna recognition. *J Mol Biol*, 313(2):309–15, Oct 2001.
- [275] Jeffrey C Miller, Michael C Holmes, Jianbin Wang, Dmitry Y Guschin, Ya-Li Lee, Igor Rupniewski, Christian M Beausejour, Adam J Waite, Nathaniel S Wang, Kenneth A Kim, Philip D Gregory, Carl O Pabo, and Edward J Rebar. An improved zinc-finger nuclease architecture for highly specific genome editing. *Nat Biotechnol*, 25(7):778–85, Jul 2007.

- [276] Rick S Mitchell, Brett F Beitzel, Astrid R W Schroder, Paul Shinn, Huaming Chen, Charles C Berry, Joseph R Ecker, and Frederic D Bushman. Retroviral dna integration: Aslv, hiv, and mlv show distinct target site preferences. *PLoS Biol*, 2(8):E234, Aug 2004.
- [277] Kaoru Mitsui, Yoshimi Tokuzawa, Hiroaki Itoh, Kohichi Segawa, Mirei Murakami, Kazutoshi Takahashi, Masayoshi Maruyama, Mitsuyo Maeda, and Shinya Yamanaka. The homeoprotein nanog is required for maintenance of pluripotency in mouse epiblast and es cells. *Cell*, 113(5):631–42, May 2003.
- [278] F J Mojica, C Díez-Villaseñor, E Soria, and G Juez. Biological significance of a family of regularly spaced repeats in the genomes of archaea, bacteria and mitochondria. *Mol Microbiol*, 36(1):244–6, Apr 2000.
- [279] Francisco J M Mojica, César Díez-Villaseñor, Jesús García-Martínez, and Elena Soria. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J Mol Evol*, 60(2):174–82, Feb 2005.
- [280] M K Montgomery, S Xu, and A Fire. Rna as a target of double-stranded rna-mediated genetic interference in *caenorhabditis elegans*. *Proc Natl Acad Sci U S A*, 95(26):15502–7, Dec 1998.
- [281] Victoria Moreno-Manzano, Francisco J Rodríguez-Jiménez, Jose L Aceña-Bonilla, Santos Fustero-Lardies, Slaven Erceg, Joaquin Dopazo, David Montaner, Miodrag Stojkovic, and Jose M Sánchez-Puelles. Fm19g11, a new hypoxia-inducible factor (hif) modulator, affects stem cell differentiation status. *J Biol Chem*, 285(2):1333–42, Jan 2010.
- [282] Eva Marie Y Moresco, Xiaohong Li, and Bruce Beutler. Going forward with genetics: recent technological advances and forward genetics in mice. *Am J Pathol*, 182(5):1462–73, May 2013.
- [283] Jason Morton, M Wayne Davis, Erik M Jorgensen, and Dana Carroll. Induction and repair of zinc-finger nuclease-targeted double-strand breaks in *caenorhabditis elegans* somatic cells. *Proc Natl Acad Sci U S A*, 103(44):16370–5, Oct 2006.
- [284] U Müller. Ten years of gene targeting: targeted mouse mutants, from vector design to phenotype analysis. *Mech Dev*, 82(1-2):3–21, Apr 1999.



- [285] Diana M Munoz, Pamela J Cassiani, Li Li, Eric Billy, Joshua M Korn, Michael D Jones, Javad Golji, David A Ruddy, Kristine Yu, Gregory McAllister, Antoine DeWeck, Dorothee Abramowski, Jessica Wan, Matthew D Shirley, Sarah Y Neshat, Daniel Rakiec, Rosalie de Beaumont, Odile Weber, Audrey Kauffmann, E Robert McDonald, 3rd, Nicholas Keen, Francesco Hofmann, William R Sellers, Tobias Schmelzle, Frank Stegmeier, and Michael R Schlabach. Crispr screens provide a comprehensive assessment of cancer vulnerabilities but generate false-positive hits for highly amplified genomic regions. *Cancer Discov*, 6(8):900–13, Aug 2016.
- [286] N Nakayama, J Lee, and L Chiu. Vascular endothelial growth factor synergistically enhances bone morphogenetic protein-4-dependent lymphohematopoietic cell generation from embryonic stem cells in vitro. *Blood*, 95(7):2275–83, Apr 2000.
- [287] James C Neil and Ewan R Cameron. Retroviral insertion sites and cancer: fountain of all knowledge? *Cancer Cell*, 2(4):253–5, Oct 2002.
- [288] Ray Kit Ng, Wendy Dean, Claire Dawson, Diana Lucifero, Zofia Madeja, Wolf Reik, and Myriam Hemberger. Epigenetic restriction of embryonic cell lineage fate by methylation of elf5. *Nat Cell Biol*, 10(11):1280–90, Nov 2008.
- [289] J Nichols, B Zevnik, K Anastassiadis, H Niwa, D Klewe-Nebenius, I Chambers, H Schöler, and A Smith. Formation of pluripotent stem cells in the mammalian embryo depends on the pou transcription factor oct4. *Cell*, 95(3):379–91, Oct 1998.
- [290] Jennifer Nichols, Kenneth Jones, Jenny M Phillips, Stephen A Newland, Mila Roode, William Mansfield, Austin Smith, and Anne Cooke. Validated germline-competent embryonic stem cell lines from nonobese diabetic mice. *Nat Med*, 15(7):814–8, Jul 2009.
- [291] Jennifer Nichols, Jose Silva, Mila Roode, and Austin Smith. Suppression of erk signalling promotes ground state pluripotency in the mouse embryo. *Development*, 136(19):3215–22, Oct 2009.
- [292] Nicos A Nicola and Jeffrey J Babon. Leukemia inhibitory factor (lif). *Cytokine Growth Factor Rev*, 26(5):533–44, Oct 2015.
- [293] Hiroshi Nishimasu, F Ann Ran, Patrick D Hsu, Silvana Konermann, Soraya I Shehata, Naoshi Dohmae, Ryuichiro Ishitani, Feng Zhang, and Osamu Nureki. Crystal structure of cas9 in complex with guide rna and target dna. *Cell*, 156(5):935–49, Feb 2014.

- [294] Akira Nishiyama, Alexei A Sharov, Yulan Piao, Misa Amano, Tomokazu Amano, Hien G Hoang, Bernard Y Binder, Richard Tapnio, Uwem Bassey, Justin N Malinonou, Lina S Correa-Cerro, Hong Yu, Li Xin, Emily Meyers, Michal Zalzman, Yuhki Nakatake, Carole Stagg, Lioudmila Sharova, Yong Qian, Dawood Dudekula, Sarah Sheer, Jean S Cadet, Tetsuya Hirata, Hsih-Te Yang, Ilya Goldberg, Michele K Evans, Dan L Longo, David Schlessinger, and Minoru S H Ko. Systematic repression of transcription factors reveals limited patterns of gene expression changes in es cells. *Sci Rep*, 3:1390, 2013.
- [295] H Niwa, T Burdon, I Chambers, and A Smith. Self-renewal of pluripotent embryonic stem cells is mediated via activation of stat3. *Genes Dev*, 12(13):2048–60, Jul 1998.
- [296] H Niwa, J Miyazaki, and A G Smith. Quantitative expression of oct-3/4 defines differentiation, dedifferentiation or self-renewal of es cells. *Nat Genet*, 24(4):372–6, Apr 2000.
- [297] Hitoshi Niwa, Kazuya Ogawa, Daisuke Shimosato, and Kenjiro Adachi. A parallel circuit of lif signalling pathways maintains pluripotency of mouse es cells. *Nature*, 460(7251):118–22, Jul 2009.
- [298] Hitoshi Niwa, Yayoi Toyooka, Daisuke Shimosato, Dan Strumpf, Kadue Takahashi, Rika Yagi, and Janet Rossant. Interaction between oct3/4 and cdx2 determines trophoctoderm differentiation. *Cell*, 123(5):917–29, Dec 2005.
- [299] Takahiro Nobukuni, Manel Joaquin, Marta Roccio, Stephen G Dann, So Young Kim, Pawan Gulati, Maya P Byfield, Jonathan M Backer, Francois Natt, Johannes L Bos, Fried J T Zwartkruis, and George Thomas. Amino acids mediate mtor/raptor signaling through activation of class 3 phosphatidylinositol 3oh-kinase. *Proc Natl Acad Sci U S A*, 102(40):14238–43, Oct 2005.
- [300] Hiroki Nojima, Chiharu Tokunaga, Satoshi Eguchi, Noriko Oshiro, Sujuti Hidayat, Ken-ichi Yoshino, Kenta Hara, Noriaki Tanaka, Joseph Avruch, and Kazuyoshi Yonezawa. The mammalian target of rapamycin (mtor) partner, raptor, binds the mtor substrates p70 s6 kinase and 4e-bp1 through their tor signaling (tos) motif. *J Biol Chem*, 278(18):15461–4, May 2003.
- [301] Kazuya Ogawa, Ryuichi Nishinakamura, Yuko Iwamatsu, Daisuke Shimosato, and Hitoshi Niwa. Synergistic action of wnt and lif in maintaining pluripotency of mouse es cells. *Biochem Biophys Res Commun*, 343(1):159–66, Apr 2006.

- [302] H Oh, Y Fujio, K Kunisada, H Hirota, H Matsui, T Kishimoto, and K Yamauchi-Takahara. Activation of phosphatidylinositol 3-kinase through glycoprotein 130 induces protein kinase b and p70 s6 kinase phosphorylation in cardiac myocytes. *J Biol Chem*, 273(16):9703–10, Apr 1998.
- [303] K Okamoto, H Okazawa, A Okuda, M Sakai, M Muramatsu, and H Hamada. A novel octamer binding transcription factor is differentially expressed in mouse embryonic cells. *Cell*, 60(3):461–72, Feb 1990.
- [304] M Okano, D W Bell, D A Haber, and E Li. Dna methyltransferases dnmt3a and dnmt3b are essential for de novo methylation and mammalian development. *Cell*, 99(3):247–57, Oct 1999.
- [305] Swee Hoe Ong, Yilong Li, Hiroko Koike-Yusa, and Kosuke Yusa. Optimised metrics for crispr-ko screens with second-generation grna libraries. *Sci Rep*, 7(1):7384, Aug 2017.
- [306] T L Orr-Weaver, J W Szostak, and R J Rothstein. Yeast transformation: a model system for the study of recombination. *Proc Natl Acad Sci U S A*, 78(10):6354–8, Oct 1981.
- [307] 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. *Dev Biol*, 166(1):259–67, Nov 1994.
- [308] Guangjin Pan, Shulan Tian, Jeff Nie, Chuhu Yang, Victor Ruotti, Hairong Wei, Gudrun A Jonsdottir, Ron Stewart, and James A Thomson. Whole-genome analysis of histone h3 lysine 4 and lysine 27 methylation in human embryonic stem cells. *Cell Stem Cell*, 1(3):299–312, Sep 2007.
- [309] Mercedes Pardo, Benjamin Lang, Lu Yu, Haydn Prosser, Allan Bradley, M Madan Babu, and Jyoti Choudhary. An expanded oct4 interaction network: implications for stem cell biology, development, and disease. *Cell Stem Cell*, 6(4):382–95, Apr 2010.
- [310] Oren Parnas, Marko Jovanovic, Thomas M Eisenhaure, Rebecca H Herbst, Atray Dixit, Chun Jimmie Ye, Dariusz Przybylski, Randall J Platt, Itay Tirosh, Neville E Sanjana, Ophir Shalem, Rahul Satija, Raktima Raychowdhury, Philipp Mertins, Steven A Carr, Feng Zhang, Nir Hacohen, and Aviv Regev. A genome-wide crispr

- screen in primary immune cells to dissect regulatory networks. *Cell*, 162(3):675–86, Jul 2015.
- [311] Vikram Pattanayak, Steven Lin, John P Guilinger, Enbo Ma, Jennifer A Doudna, and David R Liu. High-throughput profiling of off-target dna cleavage reveals rna-programmed cas9 nuclease specificity. *Nat Biotechnol*, 31(9):839–43, Sep 2013.
- [312] N P Pavletich and C O Pabo. Zinc finger-dna recognition: crystal structure of a zif268-dna complex at 2.1 a. *Science*, 252(5007):809–17, May 1991.
- [313] Lucas Pelkmans, Eugenio Fava, Hannes Grabner, Michael Hannus, Bianca Habermann, Eberhard Krausz, and Marino Zerial. Genome-wide analysis of human kinases in clathrin- and caveolae/raft-mediated endocytosis. *Nature*, 436(7047):78–86, Jul 2005.
- [314] Min Peng, Na Yin, and Ming O Li. Szt2 dictates gator control of mtorc1 signalling. *Nature*, 543(7645):433–437, 03 2017.
- [315] Martin F Pera, Jessica Andrade, Souheir Houssami, Benjamin Reubinoff, Alan Trounson, Edouard G Stanley, Dorien Ward-van Oostwaard, and Christine Mummery. Regulation of human embryonic stem cell differentiation by bmp-2 and its antagonist noggin. *J Cell Sci*, 117(Pt 7):1269–80, Mar 2004.
- [316] Laura Pereira, Fei Yi, and Bradley J Merrill. Repression of nanog gene transcription by tcf3 limits embryonic stem cell self-renewal. *Mol Cell Biol*, 26(20):7479–91, Oct 2006.
- [317] Pablo Perez-Pinera, D Dewran Kocak, Christopher M Vockley, Andrew F Adler, Ami M Kabadi, Lauren R Polstein, Pratiksha I Thakore, Katherine A Glass, David G Ousterout, Kam W Leong, Farshid Guilak, Gregory E Crawford, Timothy E Reddy, and Charles A Gersbach. Rna-guided gene activation by crispr-cas9-based transcription factors. *Nat Methods*, 10(10):973–6, Oct 2013.
- [318] M Pesce and H R Schöler. Oct-4: gatekeeper in the beginnings of mammalian development. *Stem Cells*, 19(4):271–8, 2001.
- [319] Timothy R Peterson, Mathieu Laplante, Carson C Thoreen, Yasemin Sancak, Seong A Kang, W Michael Kuehl, Nathanael S Gray, and David M Sabatini. Deptor is an mtor inhibitor frequently overexpressed in multiple myeloma cells and required for their survival. *Cell*, 137(5):873–86, May 2009.

- [320] Timothy R Peterson, Shomit S Sengupta, Thurl E Harris, Anne E Carmack, Seong A Kang, Eric Balderas, David A Guertin, Katherine L Madden, Anne E Carpenter, Brian N Finck, and David M Sabatini. mtor complex 1 regulates lipin 1 localization to control the srebp pathway. *Cell*, 146(3):408–20, Aug 2011.
- [321] Constance S Petit, Agnes Roczniak-Ferguson, and Shawn M Ferguson. Recruitment of folliculin to lysosomes supports the amino acid-dependent activation of rag gtpases. *J Cell Biol*, 202(7):1107–22, Sep 2013.
- [322] Randall J Platt, Sidi Chen, Yang Zhou, Michael J Yim, Lukasz Swiech, Hannah R Kempton, James E Dahlman, Oren Parnas, Thomas M Eisenhaure, Marko Jovanovic, Daniel B Graham, Siddharth Jhunjhunwala, Matthias Heidenreich, Ramnik J Xavier, Robert Langer, Daniel G Anderson, Nir Hacohen, Aviv Regev, Guoping Feng, Phillip A Sharp, and Feng Zhang. Crispr-cas9 knockin mice for genome editing and cancer modeling. *Cell*, 159(2):440–55, Oct 2014.
- [323] Christian Popp, Wendy Dean, Suhua Feng, Shawn J Cokus, Simon Andrews, Matteo Pellegrini, Steven E Jacobsen, and Wolf Reik. Genome-wide erasure of dna methylation in mouse primordial germ cells is affected by aid deficiency. *Nature*, 463(7284):1101–5, Feb 2010.
- [324] C Pourcel, G Salvignol, and G Vergnaud. Crispr elements in yersinia pestis acquire new repeats by preferential uptake of bacteriophage dna, and provide additional tools for evolutionary studies. *Microbiology*, 151(Pt 3):653–63, Mar 2005.
- [325] D A Pye and J T Gallagher. Monomer complexes of basic fibroblast growth factor and heparan sulfate oligosaccharides are the minimal functional unit for cell activation. *J Biol Chem*, 274(19):13456–61, May 1999.
- [326] Lei S Qi, Matthew H Larson, Luke A Gilbert, Jennifer A Doudna, Jonathan S Weissman, Adam P Arkin, and Wendell A Lim. Repurposing crispr as an rna-guided platform for sequence-specific control of gene expression. *Cell*, 152(5):1173–83, Feb 2013.
- [327] Roland Rad, Lena Rad, Wei Wang, Juan Cadinanos, George Vassiliou, Stephen Rice, Lia S Campos, Kosuke Yusa, Ruby Banerjee, Meng Amy Li, Jorge de la Rosa, Alexander Strong, Dong Lu, Peter Ellis, Nathalie Conte, Fang Tang Yang, Pentao Liu, and Allan Bradley. Piggybac transposon mutagenesis: a tool for cancer gene discovery in mice. *Science*, 330(6007):1104–7, Nov 2010.

- [328] Aliaksandra Radzisheuskaya, Gloryn Le Bin Chia, Rodrigo L dos Santos, Thorold W Theunissen, L Filipe C Castro, Jennifer Nichols, and José C R Silva. A defined oct4 level governs cell state transitions of pluripotency entry and differentiation into all embryonic lineages. *Nat Cell Biol*, 15(6):579–90, Jun 2013.
- [329] F Ann Ran, Patrick D Hsu, Chie-Yu Lin, Jonathan S Gootenberg, Silvana Konermann, Alexandro E Trevino, David A Scott, Azusa Inoue, Shogo Matoba, Yi Zhang, and Feng Zhang. Double nicking by rna-guided crispr cas9 for enhanced genome editing specificity. *Cell*, 154(6):1380–9, Sep 2013.
- [330] Manuele Rebsamen, Lorena Pochini, Taras Stasyk, Mariana E G de Araújo, Michele Galluccio, Richard K Kandasamy, Berend Snijder, Astrid Fauster, Elena L Rudashevskaya, Manuela Bruckner, Stefania Scorzoni, Przemyslaw A Filipek, Kilian V M Huber, Johannes W Bigenzahn, Leonhard X Heinz, Claudine Kraft, Keiryn L Bennett, Cesare Indiveri, Lukas A Huber, and Giulio Superti-Furga. Slc38a9 is a component of the lysosomal amino acid sensing machinery that controls mtorc1. *Nature*, 519(7544):477–81, Mar 2015.
- [331] J L Resnick, L S Bixler, L Cheng, and P J Donovan. Long-term proliferation of mouse primordial germ cells in culture. *Nature*, 359(6395):550–1, Oct 1992.
- [332] Nicola Reynolds, Paulina Latos, Antony Hynes-Allen, Remco Loos, Donna Leaford, Aoife O’Shaughnessy, Olukunbi Mosaku, Jason Signolet, Philip Brennecke, Tüzer Kalkan, Ita Costello, Peter Humphreys, William Mansfield, Kentaro Nakagawa, John Strouboulis, Axel Behrens, Paul Bertone, and Brian Hendrich. Nurd suppresses pluripotency gene expression to promote transcriptional heterogeneity and lineage commitment. *Cell Stem Cell*, 10(5):583–94, May 2012.
- [333] Deepak Reyon, Shengdar Q Tsai, Cyd Khayter, Jennifer A Foden, Jeffrey D Sander, and J Keith Joung. Flash assembly of talens for high-throughput genome editing. *Nat Biotechnol*, 30(5):460–5, May 2012.
- [334] E M Rinchik. Chemical mutagenesis and fine-structure functional analysis of the mouse genome. *Trends Genet*, 7(1):15–21, Jan 1991.
- [335] E Robertson, A Bradley, M Kuehn, and M Evans. Germ-line transmission of genes introduced into cultured pluripotential cells by retroviral vector. *Nature*, 323(6087):445–8, 1986.

- [336] Patrick Römer, Simone Hahn, Tina Jordan, Tina Strauss, Ulla Bonas, and Thomas Lahaye. Plant pathogen recognition mediated by promoter activation of the pepper bs3 resistance gene. *Science*, 318(5850):645–8, Oct 2007.
- [337] P Rouet, F Smih, and M Jasin. Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. *Mol Cell Biol*, 14(12):8096–106, Dec 1994.
- [338] Adrien Rousseau and Anne Bertolotti. An evolutionarily conserved pathway controls proteasome homeostasis. *Nature*, 536(7615):184–9, 08 2016.
- [339] L B Russell, P R Hunsicker, N L Cacheiro, J W Bangham, W L Russell, and M D Shelby. Chlorambucil effectively induces deletion mutations in mouse germ cells. *Proc Natl Acad Sci U S A*, 86(10):3704–8, May 1989.
- [340] W L Russell, E M Kelly, P R Hunsicker, J W Bangham, S C Maddux, and E L Phipps. Specific-locus test shows ethylnitrosourea to be the most potent mutagen in the mouse. *Proc Natl Acad Sci U S A*, 76(11):5818–9, Nov 1979.
- [341] D M Sabatini, H Erdjument-Bromage, M Lui, P Tempst, and S H Snyder. Raft1: a mammalian protein that binds to fkbp12 in a rapamycin-dependent fashion and is homologous to yeast tors. *Cell*, 78(1):35–43, Jul 1994.
- [342] Timothy J Sadlon, Ian D Lewis, and Richard J D’Andrea. Bmp4: its role in development of the hematopoietic system and potential as a hematopoietic growth factor. *Stem Cells*, 22(4):457–74, 2004.
- [343] Masayo Sakaki-Yumoto, Yoko Katsuno, and Rik Derynck. Tgf- $\beta$  family signaling in stem cells. *Biochim Biophys Acta*, 1830(2):2280–96, Feb 2013.
- [344] Nehmé Saksouk, Teresa K Barth, Celine Ziegler-Birling, Nelly Olova, Agnieszka Nowak, Elodie Rey, Julio Mateos-Langerak, Serge Urbach, Wolf Reik, Maria-Elena Torres-Padilla, Axel Imhof, Jérôme Déjardin, and Elisabeth Simboeck. Redundant mechanisms to form silent chromatin at pericentromeric regions rely on bend3 and dna methylation. *Mol Cell*, 56(4):580–94, Nov 2014.
- [345] Yasemin Sancak, Liron Bar-Peled, Roberto Zoncu, Andrew L Markhard, Shigeyuki Nada, and David M Sabatini. Ragulator-rag complex targets mtorc1 to the lysosomal surface and is necessary for its activation by amino acids. *Cell*, 141(2):290–303, Apr 2010.

- [346] Yasemin Sancak, Timothy R Peterson, Yoav D Shaul, Robert A Lindquist, Carson C Thoreen, Liron Bar-Peled, and David M Sabatini. The rag gtpases bind raptor and mediate amino acid signaling to mtorc1. *Science*, 320(5882):1496–501, Jun 2008.
- [347] Yasemin Sancak, Carson C Thoreen, Timothy R Peterson, Robert A Lindquist, Seong A Kang, Eric Spooner, Steven A Carr, and David M Sabatini. Pras40 is an insulin-regulated inhibitor of the mtorc1 protein kinase. *Mol Cell*, 25(6):903–15, Mar 2007.
- [348] Jeffrey D Sander, Lindsay Cade, Cyd Khayter, Deepak Reyon, Randall T Peterson, J Keith Joung, and Jing-Ruey J Yeh. Targeted gene disruption in somatic zebrafish cells using engineered talens. *Nat Biotechnol*, 29(8):697–8, Aug 2011.
- [349] Jeffrey D Sander, Elizabeth J Dahlborg, Mathew J Goodwin, Lindsay Cade, Feng Zhang, Daniel Cifuentes, Shaun J Curtin, Jessica S Blackburn, Stacey Thibodeau-Beganny, Yiping Qi, Christopher J Pierick, Ellen Hoffman, Morgan L Maeder, Cyd Khayter, Deepak Reyon, Drena Dobbs, David M Langenau, Robert M Stupar, Antonio J Giraldez, Daniel F Voytas, Randall T Peterson, Jing-Ruey J Yeh, and J Keith Joung. Selection-free zinc-finger-nuclease engineering by context-dependent assembly (coda). *Nat Methods*, 8(1):67–9, Jan 2011.
- [350] Neville E Sanjana, Jason Wright, Kaijie Zheng, Ophir Shalem, Pierre Fontanillas, Julia Joung, Christine Cheng, Aviv Regev, and Feng Zhang. High-resolution interrogation of functional elements in the noncoding genome. *Science*, 353(6307):1545–1549, 09 2016.
- [351] Dos D Sarbassov, Siraj M Ali, Shomit Sengupta, Joon-Ho Sheen, Peggy P Hsu, Alex F Bagley, Andrew L Markhard, and David M Sabatini. Prolonged rapamycin treatment inhibits mtorc2 assembly and akt/pkb. *Mol Cell*, 22(2):159–68, Apr 2006.
- [352] Noboru Sato, Laurent Meijer, Leandros Skaltsounis, Paul Greengard, and Ali H Brivanlou. Maintenance of pluripotency in human and mouse embryonic stem cells through activation of wnt signaling by a pharmacological gsk-3-specific inhibitor. *Nat Med*, 10(1):55–63, Jan 2004.
- [353] Noriko Sato, Mitsumasa Kondo, and Ken-ichi Arai. The orphan nuclear receptor gcnf recruits dna methyltransferase for oct-3/4 silencing. *Biochem Biophys Res Commun*, 344(3):845–51, Jun 2006.



- [354] T Sato, A Nakashima, L Guo, K Coffman, and F Tamanoi. Single amino-acid changes that confer constitutive activation of mtor are discovered in human cancer. *Oncogene*, 29(18):2746–52, May 2010.
- [355] Leslie J Saucedo, Xinsheng Gao, Dominic A Chiarelli, Ling Li, Duoija Pan, and Bruce A Edgar. Rheb promotes cell growth as a component of the insulin/tor signalling network. *Nat Cell Biol*, 5(6):566–71, Jun 2003.
- [356] Carla Sawan, Hector Hernandez-Vargas, Rabih Murr, Fabrice Lopez, Thomas Vaissière, Akram Y Ghantous, Cyrille Cuenin, Jean Imbert, Zhao-Qi Wang, Bing Ren, and Zdenko Herceg. Histone acetyltransferase cofactor trrap maintains self-renewal and restricts differentiation of embryonic stem cells. *Stem Cells*, 31(5):979–91, May 2013.
- [357] Sandeep Saxena, Zophonías O Jónsson, and Anindya Dutta. Small rnas with imperfect match to endogenous mrna repress translation. implications for off-target activity of small inhibitory rna in mammalian cells. *J Biol Chem*, 278(45):44312–9, Nov 2003.
- [358] Robert A Saxton, Kevin E Knockenhauer, Rachel L Wolfson, Lynne Chantranupong, Michael E Pacold, Tim Wang, Thomas U Schwartz, and David M Sabatini. Structural basis for leucine sensing by the sestrin2-mtorc1 pathway. *Science*, 351(6268):53–8, Jan 2016.
- [359] Stefanie S Schalm, Diane C Fingar, David M Sabatini, and John Blenis. Tos motif-mediated raptor binding regulates 4e-bp1 multisite phosphorylation and function. *Curr Biol*, 13(10):797–806, May 2003.
- [360] Christoph Schaniel, Yen-Sin Ang, Kajan Ratnakumar, Catherine Cormier, Taneisha James, Emily Bernstein, Ihor R Lemischka, and Patrick J Paddison. Smarcc1/baf155 couples self-renewal gene repression with changes in chromatin structure in mouse embryonic stem cells. *Stem Cells*, 27(12):2979–91, Dec 2009.
- [361] W P Schiemann, J L Bartoe, and N M Nathanson. Box 3-independent signaling mechanisms are involved in leukemia inhibitory factor receptor alpha- and gp130-mediated stimulation of mitogen-activated protein kinase. evidence for participation of multiple signaling pathways which converge at ras. *J Biol Chem*, 272(26):16631–6, Jun 1997.

- [362] Jonathan L Schmid-Burgk, Tobias Schmidt, Vera Kaiser, Klara Höning, and Veit Hornung. A ligation-independent cloning technique for high-throughput assembly of transcription activator-like effector genes. *Nat Biotechnol*, 31(1):76–81, Jan 2013.
- [363] H R Schöler, R Balling, A K Hatzopoulos, N Suzuki, and P Gruss. Octamer binding proteins confer transcriptional activity in early mouse embryogenesis. *EMBO J*, 8(9):2551–7, Sep 1989.
- [364] Gerald Schwank, Bon-Kyoung Koo, Valentina Sasselli, Johanna F Dekkers, Inha Heo, Turan Demircan, Nobuo Sasaki, Sander Boymans, Edwin Cuppen, Cornelis K van der Ent, Edward E S Nieuwenhuis, Jeffrey M Beekman, and Hans Clevers. Functional repair of cftr by crispr/cas9 in intestinal stem cell organoids of cystic fibrosis patients. *Cell Stem Cell*, 13(6):653–8, Dec 2013.
- [365] P L Schwartzberg, S P Goff, and E J Robertson. Germ-line transmission of a c-abl mutation produced by targeted gene disruption in es cells. *Science*, 246(4931):799–803, Nov 1989.
- [366] Ophir Shalem, Neville E Sanjana, Ella Hartenian, Xi Shi, David A Scott, Tarjei Mikkelsen, Dirk Heckl, Benjamin L Ebert, David E Root, John G Doench, and Feng Zhang. Genome-scale crispr-cas9 knockout screening in human cells. *Science*, 343(6166):84–87, Jan 2014.
- [367] Ophir Shalem, Neville E Sanjana, and Feng Zhang. High-throughput functional genomics using crispr-cas9. *Nat Rev Genet*, 16(5):299–311, May 2015.
- [368] R P Sharma and V L Chopra. Effect of the wingless (wg1) mutation on wing and haltere development in drosophila melanogaster. *Dev Biol*, 48(2):461–5, Feb 1976.
- [369] Bin Shen, Wensheng Zhang, Jun Zhang, Jiankui Zhou, Jianying Wang, Li Chen, Lu Wang, Alex Hodgkins, Vivek Iyer, Xingxu Huang, and William C Skarnes. Efficient genome modification by crispr-cas9 nickase with minimal off-target effects. *Nat Methods*, 11(4):399–402, Apr 2014.
- [370] Junwei Shi, Eric Wang, Joseph P Milazzo, Zihua Wang, Justin B Kinney, and Christopher R Vakoc. Discovery of cancer drug targets by crispr-cas9 screening of protein domains. *Nat Biotechnol*, 33(6):661–7, Jun 2015.
- [371] Yanhong Shi, Haruhisa Inoue, Joseph C Wu, and Shinya Yamanaka. Induced pluripotent stem cell technology: a decade of progress. *Nat Rev Drug Discov*, 16(2):115–130,

- 02 2017.
- [372] Koji Shimosaki, Kinichi Nakashima, Hitoshi Niwa, and Tetsuya Taga. Involvement of oct3/4 in the enhancement of neuronal differentiation of es cells in neurogenesis-inducing cultures. *Development*, 130(11):2505–12, Jun 2003.
- [373] Sergey Shmakov, Omar O Abudayyeh, Kira S Makarova, Yuri I Wolf, Jonathan S Gootenberg, Ekaterina Semenova, Leonid Minakhin, Julia Joung, Silvana Konermann, Konstantin Severinov, Feng Zhang, and Eugene V Koonin. Discovery and functional characterization of diverse class 2 crispr-cas systems. *Mol Cell*, 60(3):385–97, Nov 2015.
- [374] Brian R Shy, Chun-I Wu, Galina F Khramtsova, Jenny Y Zhang, Olufunmilayo I Olopade, Kathleen H Goss, and Bradley J Merrill. Regulation of tcf7l1 dna binding and protein stability as principal mechanisms of wnt/ $\beta$ -catenin signaling. *Cell Rep*, 4(1):1–9, Jul 2013.
- [375] Jose Silva, Jennifer Nichols, Thorold W Theunissen, Ge Guo, Anouk L van Oosten, Ornella Barrandon, Jason Wray, Shinya Yamanaka, Ian Chambers, and Austin Smith. Nanog is the gateway to the pluripotent ground state. *Cell*, 138(4):722–37, Aug 2009.
- [376] Jose M Silva, Mamie Z Li, Ken Chang, Wei Ge, Michael C Golding, Richard J Rickles, Despina Siolas, Guang Hu, Patrick J Paddison, Michael R Schlabach, Nihar Sheth, Jeff Bradshaw, Julia Burchard, Amit Kulkarni, Guy Cavet, Ravi Sachidanandan, W Richard McCombie, Michele A Cleary, Stephen J Elledge, and Gregory J Hannon. Second-generation shrna libraries covering the mouse and human genomes. *Nat Genet*, 37(11):1281–8, Nov 2005.
- [377] Amar M Singh, Takashi Hamazaki, Katherine E Hankowski, and Naohiro Terada. A heterogeneous expression pattern for nanog in embryonic stem cells. *Stem Cells*, 25(10):2534–42, Oct 2007.
- [378] Lasse Sinkkonen, Tabea Hugenschmidt, Philipp Berninger, Dimos Gaidatzis, Fabio Mohn, Caroline G Artus-Revel, Mihaela Zavolan, Petr Svoboda, and Witold Filipowicz. Micrnas control de novo dna methylation through regulation of transcriptional repressors in mouse embryonic stem cells. *Nat Struct Mol Biol*, 15(3):259–67, Mar 2008.

- [379] Carol A Sledz, Michelle Holko, Michael J de Veer, Robert H Silverman, and Bryan R G Williams. Activation of the interferon system by short-interfering rnas. *Nat Cell Biol*, 5(9):834–9, Sep 2003.
- [380] Sébastien A Smallwood, Shin-Ichi Tomizawa, Felix Krueger, Nico Ruf, Natasha Carli, Anne Segonds-Pichon, Shun Sato, Kenichiro Hata, Simon R Andrews, and Gavin Kelsey. Dynamic cpg island methylation landscape in oocytes and preimplantation embryos. *Nat Genet*, 43(8):811–4, Jun 2011.
- [381] 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*, 336(6200):688–90, Dec 1988.
- [382] A G Smith and M L Hooper. Buffalo rat liver cells produce a diffusible activity which inhibits the differentiation of murine embryonal carcinoma and embryonic stem cells. *Dev Biol*, 121(1):1–9, May 1987.
- [383] G E Smith, M D Summers, and M J Fraser. Production of human beta interferon in insect cells infected with a baculovirus expression vector. *Mol Cell Biol*, 3(12):2156–65, Dec 1983.
- [384] Zachary D Smith, Michelle M Chan, Tarjei S Mikkelsen, Hongcang Gu, Andreas Gnirke, Aviv Regev, and Alexander Meissner. A unique regulatory phase of dna methylation in the early mammalian embryo. *Nature*, 484(7394):339–44, Mar 2012.
- [385] Zachary D Smith and Alexander Meissner. Dna methylation: roles in mammalian development. *Nat Rev Genet*, 14(3):204–20, Mar 2013.
- [386] O Smithies, R G Gregg, S S Boggs, M A Koralewski, and R S Kucherlapati. Insertion of dna sequences into the human chromosomal beta-globin locus by homologous recombination. *Nature*, 317(6034):230–4, 1985.
- [387] Daniel St Johnston. The art and design of genetic screens: *Drosophila melanogaster*. *Nat Rev Genet*, 3(3):176–88, Mar 2002.
- [388] N Stahl, T G Boulton, T Farruggella, N Y Ip, S Davis, B A Witthuhn, F W Quelle, O Silvennoinen, G Barbieri, and S Pellegrini. Association and activation of jak-tyk kinases by cntf-lif-osm-il-6 beta receptor components. *Science*, 263(5143):92–5, Jan 1994.

- [389] G R Stark, I M Kerr, B R Williams, R H Silverman, and R D Schreiber. How cells respond to interferons. *Annu Rev Biochem*, 67:227–64, 1998.
- [390] Marios P Stavridis, J Simon Lunn, Barry J Collins, and Kate G Storey. A discrete period of fgf-induced erk1/2 signalling is required for vertebrate neural specification. *Development*, 134(16):2889–94, Aug 2007.
- [391] Samuel H Sternberg, Sy Redding, Martin Jinek, Eric C Greene, and Jennifer A Doudna. Dna interrogation by the crispr rna-guided endonuclease cas9. *Nature*, 507(7490):62–7, Mar 2014.
- [392] C L Stewart, P Kaspar, L J Brunet, H Bhatt, I Gadi, F Köntgen, and S J Abbondanzo. Blastocyst implantation depends on maternal expression of leukaemia inhibitory factor. *Nature*, 359(6390):76–9, Sep 1992.
- [393] Hugo Stocker, Thomas Radimerski, Benno Schindelholz, Franz Wittwer, Priyanka Belawat, Pierre Daram, Sebastian Breuer, George Thomas, and Ernst Hafen. Rheb is an essential regulator of s6k in controlling cell growth in drosophila. *Nat Cell Biol*, 5(6):559–65, Jun 2003.
- [394] Dan Strumpf, Chai-An Mao, Yojiro Yamanaka, Amy Ralston, Kallayane Chawengsaksophak, Felix Beck, and Janet Rossant. Cdx2 is required for correct cell fate specification and differentiation of trophectoderm in the mouse blastocyst. *Development*, 132(9):2093–102, May 2005.
- [395] Aravind Subramanian, Pablo Tamayo, Vamsi K Mootha, Sayan Mukherjee, Benjamin L Ebert, Michael A Gillette, Amanda Paulovich, Scott L Pomeroy, Todd R Golub, Eric S Lander, and Jill P Mesirov. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*, 102(43):15545–50, Oct 2005.
- [396] Chuanhai Sun, Yuhki Nakatake, Tadayuki Akagi, Hiroki Ura, Takahiko Matsuda, Akira Nishiyama, Hiroshi Koide, Minoru S H Ko, Hitoshi Niwa, and Takashi Yokota. Dax1 binds to oct3/4 and inhibits its transcriptional activity in embryonic stem cells. *Mol Cell Biol*, 29(16):4574–83, Aug 2009.
- [397] Robyn T Sussman, Timothy J Stanek, Paul Estes, John D Gearhart, Karen E Knudsen, and Steven B McMahon. The epigenetic modifier ubiquitin-specific protease 22 (usp22) regulates embryonic stem cell differentiation via transcriptional

- repression of sex-determining region y-box 2 (sox2). *J Biol Chem*, 288(33):24234–46, Aug 2013.
- [398] Michal Szczepek, Vincent Brondani, Janine Büchel, Luis Serrano, David J Segal, and Toni Cathomen. Structure-based redesign of the dimerization interface reduces the toxicity of zinc-finger nucleases. *Nat Biotechnol*, 25(7):786–93, Jul 2007.
- [399] Chih-I Tai and Qi-Long Ying. Gbx2, a lif/stat3 target, promotes reprogramming to and retention of the pluripotent ground state. *J Cell Sci*, 126(Pt 5):1093–8, Mar 2013.
- [400] Kazutoshi Takahashi and Shinya Yamanaka. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*, 126(4):663–76, Aug 2006.
- [401] Yutaka Takahashi, Michiko Takahashi, Nick Carpino, Shiann-Tarng Jou, Jyh-Rong Chao, Satoshi Tanaka, Yasufumi Shigeyoshi, Evan Parganas, and James N Ihle. Leukemia inhibitory factor regulates trophoblast giant cell differentiation via janus kinase 1-signal transducer and activator of transcription 3-suppressor of cytokine signaling 3 pathway. *Mol Endocrinol*, 22(7):1673–81, Jul 2008.
- [402] Yasuhiro Takashima, Ge Guo, Remco Loos, Jennifer Nichols, Gabriella Ficz, Felix Krueger, David Oxley, Fatima Santos, James Clarke, William Mansfield, Wolf Reik, Paul Bertone, and Austin Smith. Resetting transcription factor control circuitry toward ground-state pluripotency in human. *Cell*, 158(6):1254–69, Sep 2014.
- [403] Yasuhiro Takashima, Ge Guo, Remco Loos, Jennifer Nichols, Gabriella Ficz, Felix Krueger, David Oxley, Fatima Santos, James Clarke, William Mansfield, Wolf Reik, Paul Bertone, and Austin Smith. Resetting transcription factor control circuitry toward ground-state pluripotency in human. *Cell*, 158(6):1254–1269, Sep 2014.
- [404] Wai-Leong Tam, Chin Yan Lim, Jianyong Han, Jinqiu Zhang, Yen-Sin Ang, Huck-Hui Ng, Henry Yang, and Bing Lim. T-cell factor 3 regulates embryonic stem cell pluripotency and self-renewal by the transcriptional control of multiple lineage pathways. *Stem Cells*, 26(8):2019–31, Aug 2008.
- [405] S Tanaka, T Kunath, A K Hadjantonakis, A Nagy, and J Rossant. Promotion of trophoblast stem cell proliferation by fgf4. *Science*, 282(5396):2072–5, Dec 1998.

- [406] Marvin E Tanenbaum, Luke A Gilbert, Lei S Qi, Jonathan S Weissman, and Ronald D Vale. A protein-tagging system for signal amplification in gene expression and fluorescence imaging. *Cell*, 159(3):635–46, Oct 2014.
- [407] Pablo Tebas, David Stein, Winson W Tang, Ian Frank, Shelley Q Wang, Gary Lee, S Kaye Spratt, Richard T Surosky, Martin A Giedlin, Geoff Nichol, Michael C Holmes, Philip D Gregory, Dale G Ando, Michael Kalos, Ronald G Collman, Gwendolyn Binder-Scholl, Gabriela Plesa, Wei-Ting Hwang, Bruce L Levine, and Carl H June. Gene editing of *ccr5* in autologous cd4 t cells of persons infected with hiv. *N Engl J Med*, 370(10):901–10, Mar 2014.
- [408] Andrew R Tee, Brendan D Manning, Philippe P Roux, Lewis C Cantley, and John Blenis. Tuberous sclerosis complex gene products, tuberin and hamartin, control mtor signaling by acting as a gtpase-activating protein complex toward rheb. *Curr Biol*, 13(15):1259–68, Aug 2003.
- [409] Wee-Wei Tee, Steven S Shen, Ozgur Oksuz, Varun Narendra, and Danny Reinberg. Erk1/2 activity promotes chromatin features and rnapii phosphorylation at developmental promoters in mouse escs. *Cell*, 156(4):678–90, Feb 2014.
- [410] Paul J Tesar, Josh G Chenoweth, Frances A Brook, Timothy J Davies, Edward P Evans, David L Mack, Richard L Gardner, and Ronald D G McKay. New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature*, 448(7150):196–9, Jul 2007.
- [411] Pratiksha I Thakore, Anthony M D’Ippolito, Lingyun Song, Alexias Safi, Nishkala K Shivakumar, Ami M Kabadi, Timothy E Reddy, Gregory E Crawford, and Charles A Gersbach. Highly specific epigenome editing by crispr-cas9 repressors for silencing of distal regulatory elements. *Nat Methods*, 12(12):1143–9, Dec 2015.
- [412] Thorold W Theunissen, Benjamin E Powell, Haoyi Wang, Maya Mitalipova, Dina A Faddah, Jessica Reddy, Zi Peng Fan, Dorothea Maetzel, Kibibi Ganz, Linyu Shi, Tenzin Lungjangwa, Sumeth Imsoonthornruksa, Yonatan Stelzer, Sudharshan Rangarajan, Ana D’Alessio, Jianming Zhang, Qing Gao, Meelad M Dawlaty, Richard A Young, Nathanael S Gray, and Rudolf Jaenisch. Systematic identification of culture conditions for induction and maintenance of naive human pluripotency. *Cell Stem Cell*, 15(4):471–487, Oct 2014.

- [413] B Thoma, T A Bird, D J Friend, D P Gearing, and S K Dower. Oncostatin m and leukemia inhibitory factor trigger overlapping and different signals through partially shared receptor complexes. *J Biol Chem*, 269(8):6215–22, Feb 1994.
- [414] K R Thomas and M R Capecchi. Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. *Cell*, 51(3):503–12, Nov 1987.
- [415] K R Thomas, K R Folger, and M R Capecchi. High frequency targeting of genes to specific sites in the mammalian genome. *Cell*, 44(3):419–28, Feb 1986.
- [416] L H Thompson, K W Brookman, N J Jones, S A Allen, and A V Carrano. Molecular cloning of the human *xrcc1* gene, which corrects defective dna strand break repair and sister chromatid exchange. *Mol Cell Biol*, 10(12):6160–71, Dec 1990.
- [417] Yayoi Toyooka, Daisuke Shimosato, Kazuhiro Murakami, Kadue Takahashi, and Hitoshi Niwa. Identification and characterization of subpopulations in undifferentiated es cell culture. *Development*, 135(5):909–18, Mar 2008.
- [418] Shengdar Q Tsai, Zongli Zheng, Nhu T Nguyen, Matthew Liebers, Ved V Topkar, Vishal Thapar, Nicolas Wyvekens, Cyd Khayter, A John Iafrate, Long P Le, Martin J Aryee, and J Keith Joung. Guide-seq enables genome-wide profiling of off-target cleavage by crispr-cas nucleases. *Nat Biotechnol*, 33(2):187–197, Feb 2015.
- [419] Akiko Tsumura, Tomohiro Hayakawa, Yuichi Kumaki, Shin-ichiro Takebayashi, Morito Sakaue, Chisa Matsuoka, Kunitada Shimotohno, Fuyuki Ishikawa, En Li, Hiroki R Ueda, Jun-ichi Nakayama, and Masaki Okano. Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of dna methyltransferases *dnmt1*, *dnmt3a* and *dnmt3b*. *Genes Cells*, 11(7):805–14, Jul 2006.
- [420] Zhi-Yang Tsun, Liron Bar-Peled, Lynne Chantranupong, Roberto Zoncu, Tim Wang, Choah Kim, Eric Spooner, and David M Sabatini. The folliculin tumor suppressor is a gap for the *ragc/d* gtpases that signal amino acid levels to *mtorc1*. *Mol Cell*, 52(4):495–505, Nov 2013.
- [421] Konstantinos Tzelepis, Hiroko Koike-Yusa, Etienne De Braekeleer, Yilong Li, Emmanouil Metzakopian, Oliver M Dovey, Annalisa Mupo, Vera Grinkevich, Meng Li, Milena Mazan, Malgorzata Gozdecka, Shuhei Ohnishi, Jonathan Cooper, Miten Patel, Thomas McKerrell, Bin Chen, Ana Filipa Domingues, Paolo Gallipoli, Sarah Teichmann, Hannes Ponstingl, Ultan McDermott, Julio Saez-Rodriguez, Brian J P Huntly, Francesco Iorio, Cristina Pina, George S Vassiliou, and Kosuke Yusa. A



- crispr dropout screen identifies genetic vulnerabilities and therapeutic targets in acute myeloid leukemia. *Cell Rep*, 17(4):1193–1205, Oct 2016.
- [422] Konstantinos Tzelepis, Hiroko Koike-Yusa, Etienne De Braekeleer, Yilong Li, Emmanouil Metzakopian, Oliver M Dovey, Annalisa Mupo, Vera Grinkevich, Meng Li, Milena Mazan, Malgorzata Gozdecka, Shuhei Ohnishi, Jonathan Cooper, Miten Patel, Thomas McKerrell, Bin Chen, Ana Filipa Domingues, Paolo Gallipoli, Sarah Teichmann, Hannes Ponstingl, Ultan McDermott, Julio Saez-Rodriguez, Brian J P Huntly, Francesco Iorio, Cristina Pina, George S Vassiliou, and Kosuke Yusa. A crispr dropout screen identifies genetic vulnerabilities and therapeutic targets in acute myeloid leukemia. *Cell Rep*, 17(4):1193–1205, 10 2016.
- [423] Ludovic Vallier, Sasha Mendjan, Stephanie Brown, Zhenzhi Chng, Adrian Teo, Lucy E Smithers, Matthew W B Trotter, Candy H-H Cho, Amelie Martinez, Peter Rugg-Gunn, Gabrielle Brons, and Roger A Pedersen. Activin/nodal signalling maintains pluripotency by controlling nanog expression. *Development*, 136(8):1339–49, Apr 2009.
- [424] Ludovic Vallier, Thomas Touboul, Zhenzhi Chng, Minodora Brimpari, Nicholas Hannan, Enrique Millan, Lucy E Smithers, Matthew Trotter, Peter Rugg-Gunn, Anne Weber, and Roger A Pedersen. Early cell fate decisions of human embryonic stem cells and mouse epiblast stem cells are controlled by the same signalling pathways. *PLoS One*, 4(6):e6082, Jun 2009.
- [425] Debbie L C van den Berg, Tim Snoek, Nick P Mullin, Adam Yates, Karel Bezstarosti, Jeroen Demmers, Ian Chambers, and Raymond A Poot. An oct4-centered protein interaction network in embryonic stem cells. *Cell Stem Cell*, 6(4):369–81, Apr 2010.
- [426] Debbie L C van den Berg, Wensheng Zhang, Adam Yates, Erik Engelen, Katalin Takacs, Karel Bezstarosti, Jeroen Demmers, Ian Chambers, and Raymond A Poot. Estrogen-related receptor beta interacts with oct4 to positively regulate nanog gene expression. *Mol Cell Biol*, 28(19):5986–95, Oct 2008.
- [427] Tandis Vazin and William J Freed. Human embryonic stem cells: derivation, culture, and differentiation: a review. *Restor Neurol Neurosci*, 28(4):589–603, 2010.
- [428] Thomas J Vigdal, Christopher D Kaufman, Zsuzsanna Izsvák, Daniel F Voytas, and Zoltán Ivics. Common physical properties of dna affecting target site selec-

- tion of sleeping beauty and other *tc1*/mariner transposable elements. *J Mol Biol*, 323(3):441–52, Oct 2002.
- [429] Philipp Voigt, Wee-Wei Tee, and Danny Reinberg. A double take on bivalent promoters. *Genes Dev*, 27(12):1318–38, Jun 2013.
- [430] Haibin Wang and Sudhansu K Dey. Roadmap to embryo implantation: clues from mouse models. *Nat Rev Genet*, 7(3):185–99, Mar 2006.
- [431] Haoyi Wang, Hui Yang, Chikdu S Shivalila, Meelad M Dawlaty, Albert W Cheng, Feng Zhang, and Rudolf Jaenisch. One-step generation of mice carrying mutations in multiple genes by crispr/cas-mediated genome engineering. *Cell*, 153(4):910–8, May 2013.
- [432] Shuyu Wang, Zhi-Yang Tsun, Rachel L Wolfson, Kuang Shen, Gregory A Wyant, Molly E Plovanich, Elizabeth D Yuan, Tony D Jones, Lynne Chantranupong, William Comb, Tim Wang, Liron Bar-Peled, Roberto Zoncu, Christoph Straub, Choah Kim, Jiwon Park, Bernardo L Sabatini, and David M Sabatini. Metabolism. lysosomal amino acid transporter *slc38a9* signals arginine sufficiency to *mtorc1*. *Science*, 347(6218):188–94, Jan 2015.
- [433] Tim Wang, Kivanç Birsoy, Nicholas W Hughes, Kevin M Krupczak, Yorick Post, Jenny J Wei, Eric S Lander, and David M Sabatini. Identification and characterization of essential genes in the human genome. *Science*, 350(6264):1096–101, Nov 2015.
- [434] Tim Wang, Jenny J Wei, David M Sabatini, and Eric S Lander. Genetic screens in human cells using the crispr-cas9 system. *Science*, 343(6166):80–4, Jan 2014.
- [435] Tim Wang, Haiyan Yu, Nicholas W Hughes, Bingxu Liu, Arek Kendirli, Klara Klein, Walter W Chen, Eric S Lander, and David M Sabatini. Gene essentiality profiling reveals gene networks and synthetic lethal interactions with oncogenic *ras*. *Cell*, 168(5):890–903.e15, Feb 2017.
- [436] Zheng Wang, Efrat Oron, Brynna Nelson, Spiro Razis, and Natalia Ivanova. Distinct lineage specification roles for *nanog*, *oct4*, and *sox2* in human embryonic stem cells. *Cell Stem Cell*, 10(4):440–54, Apr 2012.
- [437] Carol B Ware, Angelique M Nelson, Brigham Mecham, Jennifer Hesson, Wenyu Zhou, Erica C Jonlin, Antonio J Jimenez-Caliani, Xinxian Deng, Christopher Ca-

- vanaugh, Savannah Cook, Paul J Tesar, Jeffrey Okada, Lilyana Margaretha, Henrik Sperber, Michael Choi, C Anthony Blau, Piper M Treuting, R David Hawkins, Vincenzo Cirulli, and Hannele Ruohola-Baker. Derivation of naive human embryonic stem cells. *Proc Natl Acad Sci U S A*, 111(12):4484–9, Mar 2014.
- [438] Akira Watanabe, Yasuhiro Yamada, and Shinya Yamanaka. Epigenetic regulation in pluripotent stem cells: a key to breaking the epigenetic barrier. *Philos Trans R Soc Lond B Biol Sci*, 368(1609):20120292, Jan 2013.
- [439] S Watanabe, H Umehara, K Murayama, M Okabe, T Kimura, and T Nakano. Activation of akt signaling is sufficient to maintain pluripotency in mouse and primate embryonic stem cells. *Oncogene*, 25(19):2697–707, May 2006.
- [440] Kathryn E Wellen, Georgia Hatzivassiliou, Uma M Sachdeva, Thi V Bui, Justin R Cross, and Craig B Thompson. Atp-citrate lyase links cellular metabolism to histone acetylation. *Science*, 324(5930):1076–80, May 2009.
- [441] Bart A Westerman, A Koen Braat, Nicole Taub, Marko Potman, Joseph H A Vissers, Marleen Blom, Els Verhoeven, Hans Stoop, Ad Gillis, Arno Velds, Wouter Nijkamp, Roderick Beijersbergen, Lukas A Huber, Leendert H J Looijenga, and Maarten van Lohuizen. A genome-wide rnai screen in mouse embryonic stem cells identifies mp1 as a key mediator of differentiation. *J Exp Med*, 208(13):2675–89, Dec 2011.
- [442] Warren A Whyte, Steve Bilodeau, David A Orlando, Heather A Hoke, Garrett M Frampton, Charles T Foster, Shaun M Cowley, and Richard A Young. Enhancer decommissioning by lsd1 during embryonic stem cell differentiation. *Nature*, 482(7384):221–5, Feb 2012.
- [443] P J Wilder, D Kelly, K Brigman, C L Peterson, T Nowling, Q S Gao, R D McComb, M R Capecchi, and A Rizzino. Inactivation of the fgf-4 gene in embryonic stem cells alters the growth and/or the survival of their early differentiated progeny. *Dev Biol*, 192(2):614–29, Dec 1997.
- [444] A F Wilks. Two putative protein-tyrosine kinases identified by application of the polymerase chain reaction. *Proc Natl Acad Sci U S A*, 86(5):1603–7, Mar 1989.
- [445] 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*, 336(6200):684–7, Dec 1988.

- [446] Maciej Wiznerowicz and Didier Trono. Conditional suppression of cellular genes: lentivirus vector-mediated drug-inducible rna interference. *J Virol*, 77(16):8957–61, Aug 2003.
- [447] S A Wolfe, H A Greisman, E I Ramm, and C O Pabo. Analysis of zinc fingers optimized via phage display: evaluating the utility of a recognition code. *J Mol Biol*, 285(5):1917–34, Feb 1999.
- [448] Rachel L Wolfson, Lynne Chantranupong, Gregory A Wyant, Xin Gu, Jose M Orozco, Kuang Shen, Kendall J Condon, Sabrina Petri, Jibril Kedir, Sonia M Scaria, Monther Abu-Remaileh, Wayne N Frankel, and David M Sabatini. Kicstor recruits gator1 to the lysosome and is necessary for nutrients to regulate mtorc1. *Nature*, 543(7645):438–442, 03 2017.
- [449] Knut Woltjen, Iacovos P Michael, Paria Mohseni, Ridham Desai, Maria Mileikovsky, Riikka Hämäläinen, Rebecca Cowling, Wei Wang, Pentao Liu, Marina Gertsenstein, Keisuke Kaji, Hoon-Ki Sung, and Andras Nagy. piggybac transposition reprograms fibroblasts to induced pluripotent stem cells. *Nature*, 458(7239):766–70, Apr 2009.
- [450] Jason Wray, Tüzer Kalkan, Sandra Gomez-Lopez, Dominik Eckardt, Andrew Cook, Rolf Kemler, and Austin Smith. Inhibition of glycogen synthase kinase-3 alleviates tcf3 repression of the pluripotency network and increases embryonic stem cell resistance to differentiation. *Nat Cell Biol*, 13(7):838–45, Jun 2011.
- [451] Jason Wray, Tuzer Kalkan, and Austin G Smith. The ground state of pluripotency. *Biochem Soc Trans*, 38(4):1027–32, Aug 2010.
- [452] Addison V Wright, James K Nuñez, and Jennifer A Doudna. Biology and applications of crispr systems: Harnessing nature’s toolbox for genome engineering. *Cell*, 164(1-2):29–44, Jan 2016.
- [453] Chun-I Wu, Jackson A Hoffman, Brian R Shy, Erin M Ford, Elaine Fuchs, Hoang Nguyen, and Bradley J Merrill. Function of wnt/ $\beta$ -catenin in counteracting tcf3 repression through the tcf3- $\beta$ -catenin interaction. *Development*, 139(12):2118–29, Jun 2012.
- [454] Dianqing Wu and Weijun Pan. Gsk3: a multifaceted kinase in wnt signaling. *Trends Biochem Sci*, 35(3):161–8, Mar 2010.

- [455] Xiaolin Wu, Yuan Li, Bruce Crise, and Shawn M Burgess. Transcription start regions in the human genome are favored targets for mlv integration. *Science*, 300(5626):1749–51, Jun 2003.
- [456] Xuebing Wu, David A Scott, Andrea J Kriz, Anthony C Chiu, Patrick D Hsu, Daniel B Dadon, Albert W Cheng, Alexandro E Trevino, Silvana Konermann, Sidi Chen, Rudolf Jaenisch, Feng Zhang, and Phillip A Sharp. Genome-wide binding of the crispr endonuclease cas9 in mammalian cells. *Nat Biotechnol*, 32(7):670–6, Jul 2014.
- [457] Yuxuan Wu, Hai Zhou, Xiaoying Fan, Ying Zhang, Man Zhang, Yinghua Wang, Zhenfei Xie, Meizhu Bai, Qi Yin, Dan Liang, Wei Tang, Jiaoyang Liao, Chikai Zhou, Wujuan Liu, Ping Zhu, Hongshan Guo, Hong Pan, Chunlian Wu, Huijuan Shi, Ligang Wu, Fuchou Tang, and Jinsong Li. Correction of a genetic disease by crispr-cas9-mediated gene editing in mouse spermatogonial stem cells. *Cell Res*, 25(1):67–79, Jan 2015.
- [458] Joanna Wysocka, Tomek Swigut, Thomas A Milne, Yali Dou, Xin Zhang, Alma L Burlingame, Robert G Roeder, Ali H Brivanlou, and C David Allis. Wdr5 associates with histone h3 methylated at k4 and is essential for h3 k4 methylation and vertebrate development. *Cell*, 121(6):859–72, Jun 2005.
- [459] Haibin Xia, Qinwen Mao, Henry L Paulson, and Beverly L Davidson. sirna-mediated gene silencing in vitro and in vivo. *Nat Biotechnol*, 20(10):1006–10, Oct 2002.
- [460] Ren-He Xu, Xin Chen, Dong S Li, Rui Li, Gregory C Addicks, Clay Glennon, Thomas P Zwaka, and James A Thomson. Bmp4 initiates human embryonic stem cell differentiation to trophoblast. *Nat Biotechnol*, 20(12):1261–4, Dec 2002.
- [461] Haijuan Yang, Derek G Rudge, Joseph D Koos, Bhamini Vaidialingam, Hyo J Yang, and Nikola P Pavletich. mtor kinase structure, mechanism and regulation. *Nature*, 497(7448):217–23, May 2013.
- [462] Hui Yang, Haoyi Wang, Chikdu S Shivalila, Albert W Cheng, Linyu Shi, and Rudolf Jaenisch. One-step generation of mice carrying reporter and conditional alleles by crispr/cas-mediated genome engineering. *Cell*, 154(6):1370–9, Sep 2013.
- [463] Shen-Hsi Yang, Tüzer Kalkan, Claire Morissroe, Hendrik Marks, Hendrik Stunnenberg, Austin Smith, and Andrew D Sharrocks. Otx2 and oct4 drive early enhancer

- activation during embryonic stem cell transition from naive pluripotency. *Cell Rep*, 7(6):1968–81, Jun 2014.
- [464] Shen-Hsi Yang, Tuzer Kalkan, Claire Morrisroe, Austin Smith, and Andrew D Sharrocks. A genome-wide rnai screen reveals map kinase phosphatases as key erk pathway regulators during embryonic stem cell differentiation. *PLoS Genet*, 8(12):e1003112, 2012.
- [465] Lin Ye, Jiaming Wang, Ashley I Beyer, Fernando Teque, Thomas J Cradick, Zhongxia Qi, Judy C Chang, Gang Bao, Marcus O Muench, Jingwei Yu, Jay A Levy, and Yuet Wai Kan. Seamless modification of wild-type induced pluripotent stem cells to the natural *ccr5* $\Delta$ 32 mutation confers resistance to hiv infection. *Proc Natl Acad Sci U S A*, 111(26):9591–6, Jul 2014.
- [466] Shoudong Ye, Ping Li, Chang Tong, and Qi-Long Ying. Embryonic stem cell self-renewal pathways converge on the transcription factor *tfcp2l1*. *EMBO J*, 32(19):2548–60, Oct 2013.
- [467] Jia-Chi Yeo, Jianming Jiang, Zi-Ying Tan, Guo-Rong Yim, Jia-Hui Ng, Jonathan Göke, Petra Kraus, Hongqing Liang, Kevin Andrew Uy Gonzales, Han-Chung Chong, Cheng-Peow Tan, Yee-Siang Lim, Nguan-Soon Tan, Thomas Lufkin, and Huck-Hui Ng. *Klf2* is an essential factor that sustains ground state pluripotency. *Cell Stem Cell*, 14(6):864–72, Jun 2014.
- [468] Fei Yi, Laura Pereira, Jackson A Hoffman, Brian R Shy, Courtney M Yuen, David R Liu, and Bradley J Merrill. Opposing effects of *tcf3* and *tcf1* control wnt stimulation of embryonic stem cell self-renewal. *Nat Cell Biol*, 13(7):762–70, Jun 2011.
- [469] Qi Long Ying, Jennifer Nichols, Ian Chambers, and Austin Smith. Bmp induction of *id* proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with *stat3*. *Cell*, 115(3):281–92, Oct 2003.
- [470] Qi-Long Ying, Marios Stavridis, Dean Griffiths, Meng Li, and Austin Smith. Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture. *Nat Biotechnol*, 21(2):183–6, Feb 2003.
- [471] Qi-Long Ying, Jason Wray, Jennifer Nichols, Laura Batlle-Morera, Bradley Doble, James Woodgett, Philip Cohen, and Austin Smith. The ground state of embryonic stem cell self-renewal. *Nature*, 453(7194):519–23, May 2008.

- [472] Richard A Young. Control of the embryonic stem cell state. *Cell*, 144(6):940–54, Mar 2011.
- [473] 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 Dev*, 9(21):2635–45, Nov 1995.
- [474] Ping Yuan, Jianyong Han, Guoji Guo, Yuriy L Orlov, Mikael Huss, Yui-Han Loh, Lai-Ping Yaw, Paul Robson, Bing Lim, and Huck-Hui Ng. Eset partners with oct4 to restrict extraembryonic trophoblast lineage potential in embryonic stem cells. *Genes Dev*, 23(21):2507–20, Nov 2009.
- [475] Kosuke Yusa, Kyoji Horie, Gen Kondoh, Michiyoshi Kouno, Yusuke Maeda, Taroh Kinoshita, and Junji Takeda. Genome-wide phenotype analysis in es cells by regulated disruption of bloom’s syndrome gene. *Nature*, 429(6994):896–9, Jun 2004.
- [476] Kosuke Yusa, Roland Rad, Junji Takeda, and Allan Bradley. Generation of transgene-free induced pluripotent mouse stem cells by the piggybac transposon. *Nat Methods*, 6(5):363–9, May 2009.
- [477] Jesse G Zalatan, Michael E Lee, Ricardo Almeida, Luke A Gilbert, Evan H Whitehead, Marie La Russa, Jordan C Tsai, Jonathan S Weissman, John E Dueber, Lei S Qi, and Wendell A Lim. Engineering complex synthetic transcriptional programs with crispr rna scaffolds. *Cell*, 160(1-2):339–50, Jan 2015.
- [478] P D Zamore, T Tuschl, P A Sharp, and D P Bartel. Rnai: double-stranded rna directs the atp-dependent cleavage of mrna at 21 to 23 nucleotide intervals. *Cell*, 101(1):25–33, Mar 2000.
- [479] Dana Zeineddine, Aya Abou Hammoud, Mohamad Mortada, and H el ene Boeuf. The oct4 protein: more than a magic stemness marker. *Am J Stem Cells*, 3(2):74–82, 2014.
- [480] Dana Zeineddine, Evangelia Papadimou, Karim Chebli, Mathieu Gineste, Jun Liu, Corinne Grey, Sherry Thurig, Atta Behfar, Valerie A Wallace, Ilona S Skerjanc, and Michel Puc eat. Oct-3/4 dose dependently regulates specification of embryonic stem cells toward a cardiac lineage and early heart development. *Dev Cell*, 11(4):535–46, Oct 2006.

- [481] Hui H Zhang, Alex I Lipovsky, Christian C Dibble, Mustafa Sahin, and Brendan D Manning. S6k1 regulates gsk3 under conditions of mtor-dependent feedback inhibition of akt. *Mol Cell*, 24(2):185–97, Oct 2006.
- [482] Jin Zhang, Esther Nuebel, George Q Daley, Carla M Koehler, and Michael A Teitell. Metabolic regulation in pluripotent stem cells during reprogramming and self-renewal. *Cell Stem Cell*, 11(5):589–95, Nov 2012.
- [483] Jinqiu Zhang, Wai-Leong Tam, Guo Qing Tong, Qiang Wu, Hsiao-Yun Chan, Boon-Seng Soh, Yuefei Lou, Jianchang Yang, Yupu Ma, Li Chai, Huck-Hui Ng, Thomas Lufkin, Paul Robson, and Bing Lim. Sall4 modulates embryonic stem cell pluripotency and early embryonic development by the transcriptional regulation of pou5f1. *Nat Cell Biol*, 8(10):1114–23, Oct 2006.
- [484] Xiao-Yong Zhang, Maya Varthi, Stephen M Sykes, Charles Phillips, Claude Warzecha, Wenting Zhu, Anastasia Wyce, Alan W Thorne, Shelley L Berger, and Steven B McMahon. The putative cancer stem cell marker usp22 is a subunit of the human saga complex required for activated transcription and cell-cycle progression. *Mol Cell*, 29(1):102–11, Jan 2008.
- [485] Xiaofei Zhang, Juan Zhang, Tao Wang, Miguel A Esteban, and Duanqing Pei. Esrrb activates oct4 transcription and sustains self-renewal and pluripotency in embryonic stem cells. *J Biol Chem*, 283(51):35825–33, Dec 2008.
- [486] Y Zhang, R Proenca, M Maffei, M Barone, L Leopold, and J M Friedman. Positional cloning of the mouse obese gene and its human homologue. *Nature*, 372(6505):425–32, Dec 1994.
- [487] Suling Zhao, Jennifer Nichols, Austin G Smith, and Meng Li. Soxb transcription factors specify neuroectodermal lineage choice in es cells. *Mol Cell Neurosci*, 27(3):332–42, Nov 2004.
- [488] Honglin Zhou, Min Xu, Qian Huang, Adam T Gates, Xiaohua D Zhang, John C Castle, Erica Stec, Marc Ferrer, Berta Strulovici, Daria J Hazuda, and Amy S Espeseth. Genome-scale rnai screen for host factors required for hiv replication. *Cell Host Microbe*, 4(5):495–504, Nov 2008.
- [489] Wenyu Zhou, Michael Choi, Daciana Margineantu, Lilyana Margaretha, Jennifer Hesson, Christopher Cavanaugh, C Anthony Blau, Marshall S Horwitz, David Hockenbery, Carol Ware, and Hannele Ruohola-Baker. Hif1 $\alpha$  induced switch from bivalent



- to exclusively glycolytic metabolism during esc-to-episc/hesc transition. *EMBO J*, 31(9):2103–16, May 2012.
- [490] Yuexin Zhou, Shiyong Zhu, Changzu Cai, Pengfei Yuan, Chunmei Li, Yanyi Huang, and Wensheng Wei. High-throughput screening of a crispr/cas9 library for functional genomics in human cells. *Nature*, 509(7501):487–91, May 2014.
- [491] Shiyong Zhu, Wei Li, Jingze Liu, Chen-Hao Chen, Qi Liao, Ping Xu, Han Xu, Tengfei Xiao, Zhongzheng Cao, Jingyu Peng, Pengfei Yuan, Myles Brown, Xiaole Shirley Liu, and Wensheng Wei. Genome-scale deletion screening of human long non-coding rnas using a paired-guide rna crispr-cas9 library. *Nat Biotechnol*, 34(12):1279–1286, Dec 2016.
- [492] Roberto Zoncu, Liron Bar-Peled, Alejo Efeyan, Shuyong Wang, Yasemin Sancak, and David M Sabatini. mtorc1 senses lysosomal amino acids through an inside-out mechanism that requires the vacuolar h(+)-atpase. *Science*, 334(6056):678–83, Nov 2011.
- [493] Jizhong Zou, Morgan L Maeder, Prashant Mali, Shondra M Pruett-Miller, Stacey Thibodeau-Beganny, Bin-Kuan Chou, Guibin Chen, Zhaohui Ye, In-Hyun Park, George Q Daley, Matthew H Porteus, J Keith Joung, and Linzhao Cheng. Gene targeting of a disease-related gene in human induced pluripotent stem and embryonic stem cells. *Cell Stem Cell*, 5(1):97–110, Jul 2009.