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 knockdown 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	*piggyBac*; *Blm*-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	*piggyBac*; 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	Snai1 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 1×10^4 per cm^2 without LIF or 2i for 48 hours before being lifted for cell sorting. These condition 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 pertubation. 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.

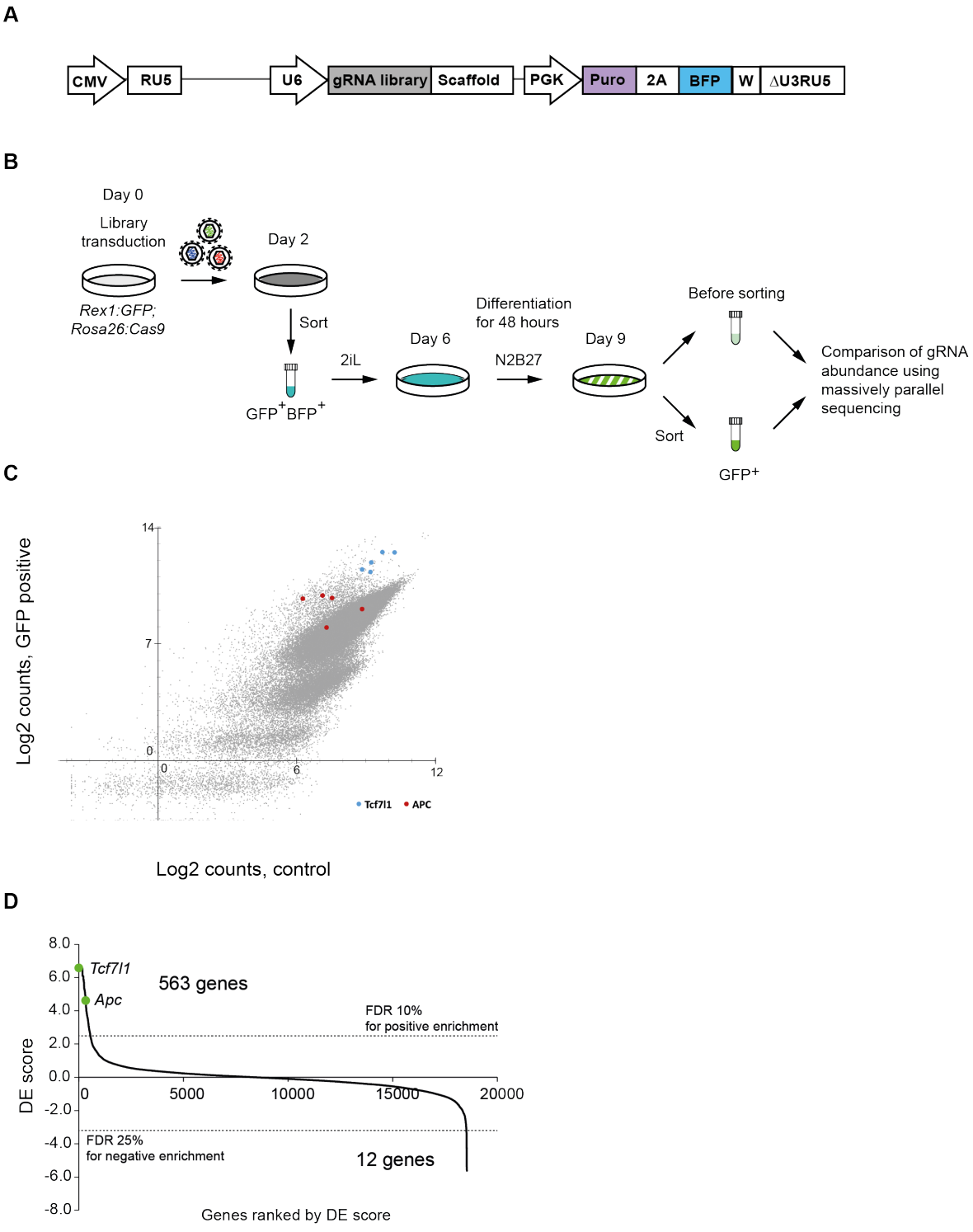

Figure 4.1: Screening Strategy and Result Analysis. (A) gRNA library expression vector. CMV: CMV promoter; RU5: 5′ long terminal repeat; U6: U6 promoter; gRNA library: gRNA sequence from the mouse V2 library; scaffold: gRNA scaffold; PGK: mouse phosphoglycerate kinase 1 promoter; Puro: puromycin resistant gene; BFP: blue fluorescent protein; 2A, Thosea asigna virus 2A peptides; W: Woodchuck Hepatitis Virus posttranscriptional regulatory element; ΔU3RU5: 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 was 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 nonefunctional 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} (Depletion P value) + [- log_{10} (Enrichment P 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.

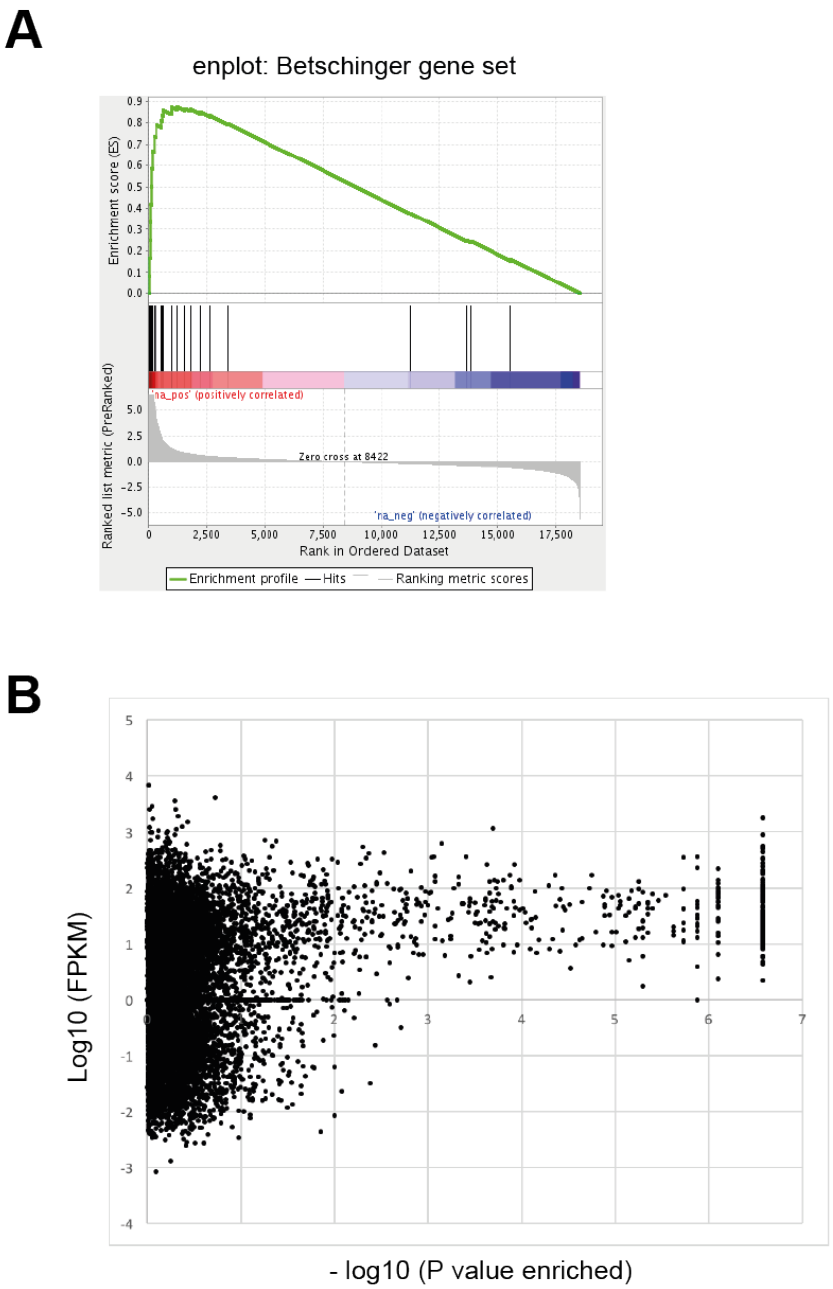

Figure 4.2: Screening Result Validation. (A) Gene Set Enrichment Analysis (GSEA) comparing the screening result with a validated gene list [19]. The validated gene set includes *Apc, Hgs, Mapksp1, Tcf7l1, Flcn, Tsc2, Kras, Zfp281, L3mbtl3, Trp53, Rnf2, Josd1, Hand1, Ctbp2, Sox2, Relt, Smg1, Ewsr1, Klhl23, E2f4, Upf1, Csde1, Trrap, Raf1, Taf6I, 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 Ffgr 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.

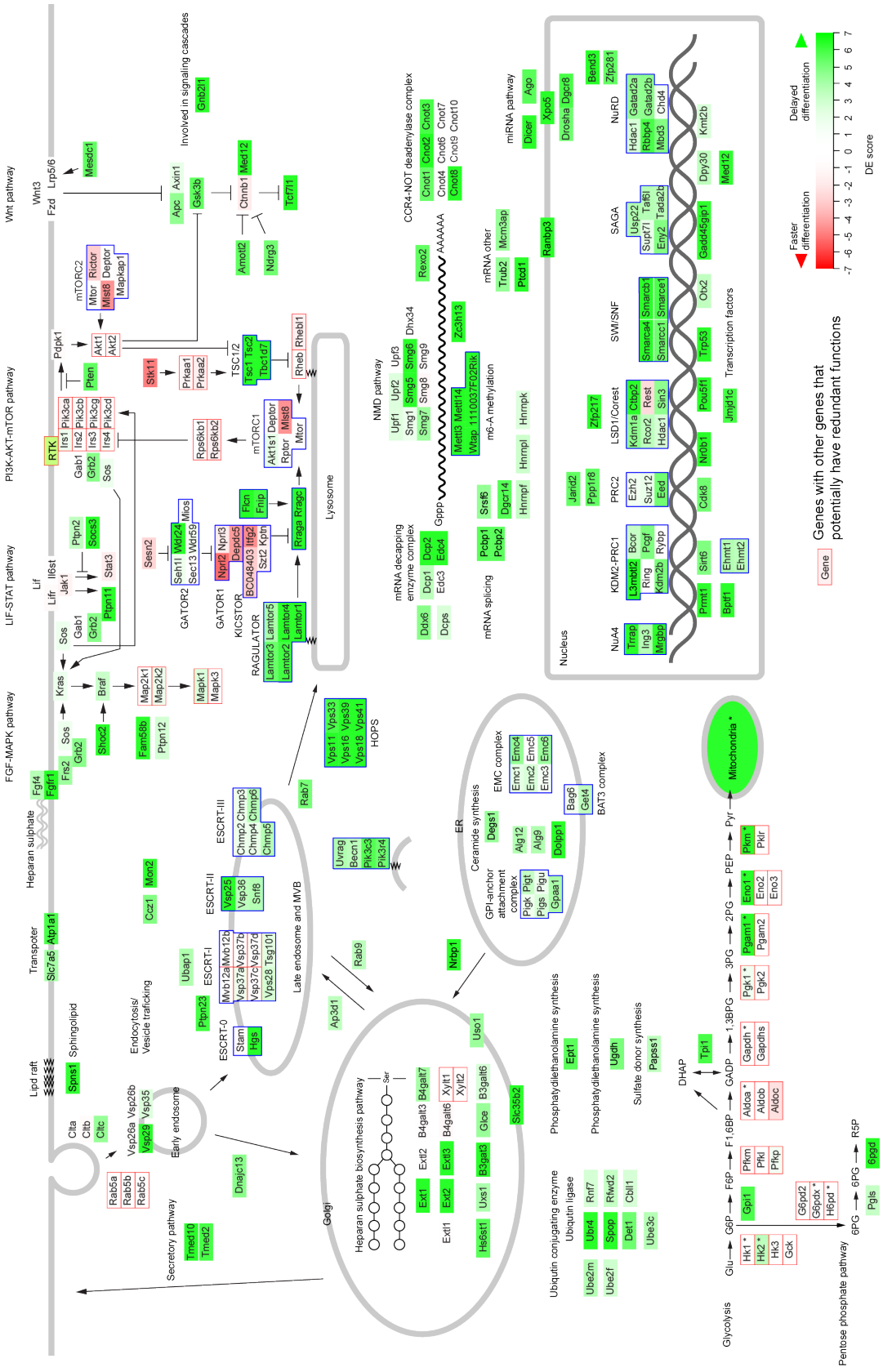

Figure 4.3: Cellular pathways underpinning the initiation of differentiation. Pathways were manually curated according to statistically significant screening hits. Green-coloured label: Knockouts that exhibited delayed differentiation phenotype; Red-coloured label: knockouts that differentiate in an 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^6A 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^6A-methyladenosine (m^6A) 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^6A 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 methytransferase 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 committment [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 deacelylates 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 decommisioning 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].

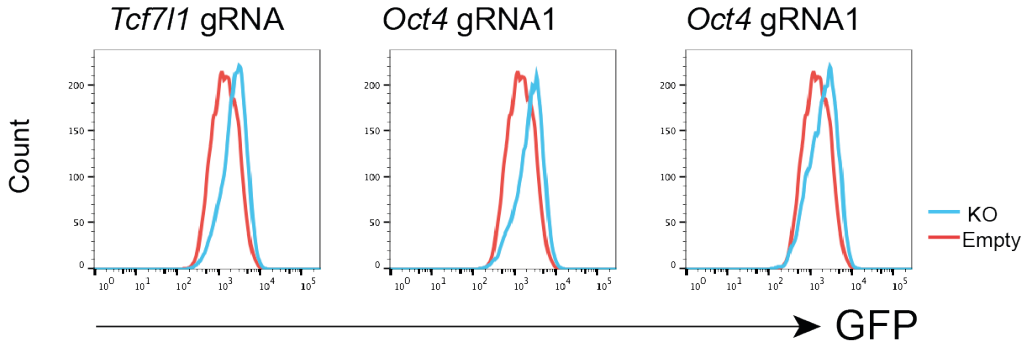

Figure 4.4: Rex1GFP differentiation profile of Oct4 knockout mESCs. ESCs were transduced with lentivirus expressing gRNAs targeting Oct4. Differentiation was induced by 2i/LIF removal. Rex1GFP was measured at 29 hours differentiation. Blue - Rex1:GFP cells transduced with gRNAs targeting indicated genes; 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/β-catenin pathway that drives colon tumorigenesis and transformation[115] [199]. However, β-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 β-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 approprate 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.

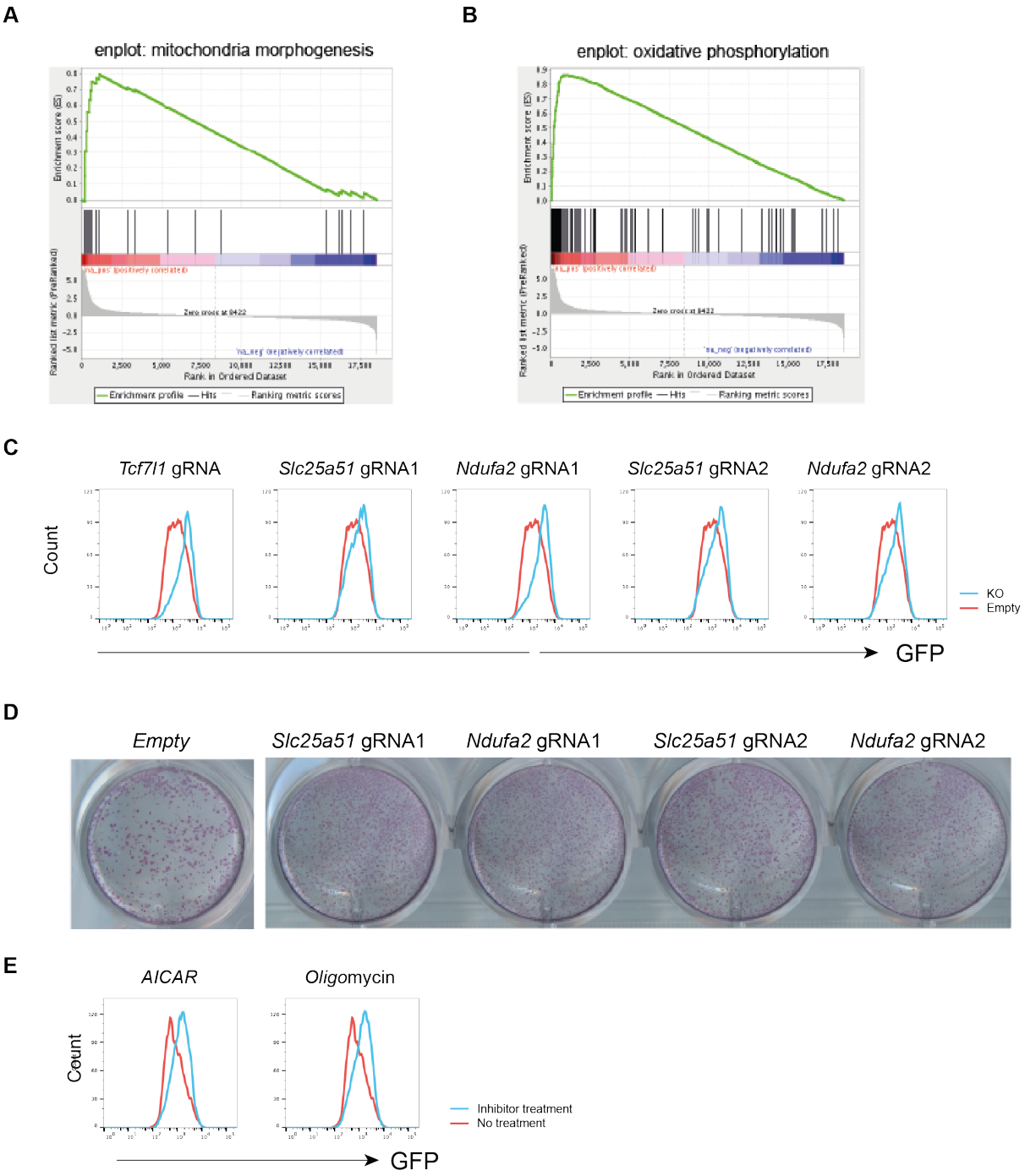

Figure 4.5: Validation of genes in mitochondria-related pathways. (A) and (B) Gene Set Enrichment Analysis using predefined datasets: mitochondria morphogenesis gene set (A) and oxidative phosphorylation gene set (B). Screening hits were pre-ranked according to the Depletion/Enrichment (DE) score, which 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 been confirmed and investigated in detail by two independent studies conducted by Karwacki-Neisius et al. and Radzisheuskaya 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.