Pathological LSD1 mutations cause HDAC-mediated aberrant gene repression during early cell differentiation

Authors

Daria Bunina[1,2,5], Pierre-Luc Germain[3], Alejandro Lopez Tobon[3], Nadine Fernandez-Novel Marx[1], Christian Arnold[2], Andrew Ó hEachteirn[2], Annique Claringbould[2], Mang Ching Lai[2], Sampath Rangasamy[4], Vinodh Narayanan[4], Giuseppe Testa[3], Judith Zaugg[2,5], Kyung Min Noh[1,5]

[1] - Genome Biology Unit, European Molecular Biology Laboratory, EMBL, Heidelberg, Germany

[2] - Structural and Computational Biology Unit, European Molecular Biology Laboratory, EMBL, Heidelberg, Germany

[3] - Department of Experimental Oncology, European Institute of Oncology, IRCCS, Milan, Italy

[4] - Center for Rare Childhood Disorders and Neurogenomics Division Translational Genomics Research Institute, Phoenix, USA

[5] - corresponding author

Abstract

Lysine-specific demethylase 1 (LSD1/KDM1A) removes methylation of histone and non-histone substrates, and recruits a repressive chromatin complex. De novo LSD1 mutations impairing protein function lead to a rare neurodevelopmental disorder, but the molecular mechanisms of the pathology are unclear. Using patient-derived fibroblasts, reprogrammed pluripotent stem cells, and differentiated cells, we found over 4000 differentially expressed genes and 68 transcription factors (TFs) whose motif accessibilities changed upon LSD1 mutation. An enhancer-mediated gene regulatory network approach identified impaired transcriptional repressor activity in fibroblast and stem cells, leading to erroneous activation of their target genes. Furthermore, our analysis revealed overall decreases in TF target genes specifically during early lineage differentiation of LSD1 mutant stem cells, likely caused by increased activity of repressive co-factors of LSD1 - histone deacetylases (HDACs). Consistently, HDAC inhibitor restored changes in gene expression including downregulation phenotype. Our findings provide insights into pathogenesis of LSD1 mutations and targets for further therapeutic studies.

Introduction

Epigenetic modifications on chromatin are essential for regulating cell type specific gene expression and, as a consequence, normal development. Specifically, proper positioning and dynamics of histone lysine methylation at enhancers and promoters is critical for multiple cellular processes, such as transcription[1], DNA repair[2], cell cycle[3] and differentiation capacity[4], thus highlighting the importance of lysine methyltransferases and demethylases. LSD1 (encoded by KDM1A gene) is a lysine demethylase, which specifically demethylates histone H3 lysine 4 mono- and di-methylated (H3K4me1/2)[5], a mark that is found at active and poised enhancers and promoters. In addition to modifying histone, LSD1 recruits repressive protein complexes such as CoREST[6] and NuRD[7] to the chromatin. LSD1 has been implicated in the control of cell differentiation[8] and early embryogenesis[9] possibly by silencing active enhancers in stem cells when they undergo differentiation. Functional alterations of LSD1 were found in pathological conditions, such as hematopoiesis impairment[10] and cancer[11].

More recently, dominant missense point mutations in LSD1 have been identified to cause a novel neurodevelopmental disorder[12] called Cleft Palate, Psychomotor Retardation and Distinctive Facial Features (CPRF), resembling the Kabuki syndrome, characterised by distinctive facial features, muscle and skeletal abnormalities and cognitive impairment. Using structural analysis and in vitro and cellular approaches, a previous work[13] has found that three of these pathological mutations (Glu403Lys, Asp580Gly, Tyr785His) affect the active site of LSD1, impair its catalytic activity, stability and, to varying extents, LSD1 binding to the histone tail[13]. Nevertheless, the molecular mechanisms by which the LSD1 mutations cause pathological manifestation have not been elucidated in the patient cells.

Here, we used patient cells and investigated the effects of two of these pathological mutations (Glu403Lys and Tyr785His) on chromatin and gene expression in different cell types and across differentiation. We observed widespread mis-regulation of gene regulatory networks and increased transcriptional repression during differentiation, likely caused by increased activity of LSD1 co-factors such as histone deacetylases (HDACs). Treating the cells with an inhibitor of HDAC - an FDA approved drug, rescued the molecular phenotype of the cells. Our results underscore how genomic data analysis from the limited numbers of patient samples identifies an unexpected target pathway that may help to treat a rare human disease.

Results

LSD1 mutation impairs chromatin accessibility at regulatory elements

To investigate the mechanism through which de novo heterozygous mutations in LSD1 cause a rare CPRF syndrome, we obtained primary skin fibroblasts from two male pediatric patients (family A - Glu403Lys and B - Tyr785His), and their healthy fathers (Fig. 1A). LSD1 largely acts on chromatin at regulatory elements such as enhancers and promoters, thus we hypothesized that the mutant cells might exhibit a disruption of regulatory elements due to the existence of mutant LSD1 in addition to the wild-type LSD1.

As LSD1 is expressed in primary fibroblasts (Supp. Fig.1A), using ATAC-seq, we identified the changes in chromatin accessibility at regulatory elements associated with LSD1 mutation

($LSD1^{MUT}$) of the CPRF patients, compared to healthy fathers (two biological replicates for each condition). Principal component analysis (PCA) showed clear separation of the patient and healthy samples (Fig. 1B). Differential chromatin accessibility analysis revealed 16,425 peaks (from the total 96,467 peaks), evenly distributed between less and more accessible in $LSD1^{MUT}$ vs. $LSD1^{WT}$ (FDR < 0.05, Fig. 1C). More accessible regions in $LSD1^{MUT}$ are enriched for active and bivalent/poised transcription start sites (+/- 3kb from TSS, 24% of peaks in $LSD1^{MUT}$ vs. 6.5% of peaks in $LSD1^{WT}$) as well as GC-rich promoter motifs (HOMER motif enrichment tool). In contrast, less accessible regions in $LSD1^{MUT}$ are enriched for active and bivalent enhancers but depleted at the TSS compared to the background peaks (SuppFig 1B).

To examine the involvement of specific transcription factor (TF) binding associated with changes in chromatin accessibility between $LSD1^{MUT}$ and $LSD1^{WT}$, we applied diffTF[14]. This quantifies genome-wide differential accessibility of TF motifs and refers to this measure as TF activity, where increased and decreased TF motifs accessibility corresponds to higher and lower TF activity, respectively. Comparing the activity of TFs in $LSD1^{MUT}$ and $LSD1^{WT}$, followed by excluding age-related TFs to minimize the effect of the donors' age (see Methods) resulted in 68 significant TFs (Fig. 1D). The 44 TFs with increased activity in $LSD1^{MUT}$ include several developmental factors (Tbx3, Mesp1, Hoxb6/a11) and epithelial-to-mesenchymal transition (EMT) TFs Zeb1 and Snai1, whereas several TFs with decreased activity in $LSD1^{MUT}$ belong to the AP-1 complex, which controls response to various stimulations. Zeb1 and Snai1 are repressive TFs and their expression was affected by $LSD1^{MUT}$ in addition to changes in their activity, indicating the overall impairment of the repressive function of these TFs. Protein-protein interactions data (STRING database) showed more interconnection between differentially active TFs (average node degree is 3.85 compared to 1.5 in a network of non-significant TFs and protein-protein interaction enrichment p-value is < 1e-16), suggesting that genome-wide chromatin accessibility differences upon $LSD1^{MUT}$ are associated with a group of functionally related TFs.

Relationship between changes in transcription and chromatin accessibility

Having identified TFs that are mis-regulated in $LSD1^{MUT}$, we sought to determine transcriptional changes in fibroblasts. RNA-seq analysis revealed distinct separations of LSD1 mutants and controls (Fig. 1E). We found more than 2500 differentially expressed genes, of which 1488 were upregulated and 1178 downregulated in $LSD1^{MUT}$ (DESeq2, FDR < 0.1, Fig. 1F). Up-regulated genes were enriched for cell cycle transition and DNA replication GO terms (Fig. 1G), however upregulation of the cell cycle genes occurred in only one (E376K) of the two $LSD1^{MUT}$ samples, undermining its wide relevance in CPRF syndrome. The E376K mutant showed a strong impairment in the histone H3 binding[13] that likely leads to the decreased recruitment of repressive HDAC co-factors and upregulation of the target genes. Down-regulated genes in $LSD1^{MUT}$ were enriched for the extracellular matrix, inflammatory response and angiogenesis GO terms, consistent with the role of extracellular matrix in EMT and impairment of the repressive EMT TFs Zeb1 and Snai1 identified above in the diffTF analysis.

To estimate the relationships between transcriptional and chromatin accessibility changes in $LSD1^{MUT}$ fibroblasts, we linked each ATAC-seq peak to the most proximal gene and analyzed these peak-gene links (Fig. 1H). While there was no overall correlation between gene expression and peak accessibility changes in $LSD1^{MUT}$ (pearson's correlation coefficient 0.07), significantly

differential peaks and genes are positively correlated (pearson's correlation coefficient 0.29). However, although significantly increased peaks in LSD1^{MUT} are more associated with promoter regions of the genes than decreased peaks (SuppFig.1B), we did not observe stronger upregulation of gene expression in the proximal genes (Fig. 1H). This suggests that the proximity of a differential peak to gene promoter does not directly translate to the functional outcome on gene expression level and a more robust approach is required to disentangle functional interactions between changes in chromatin accessibility and gene expression.

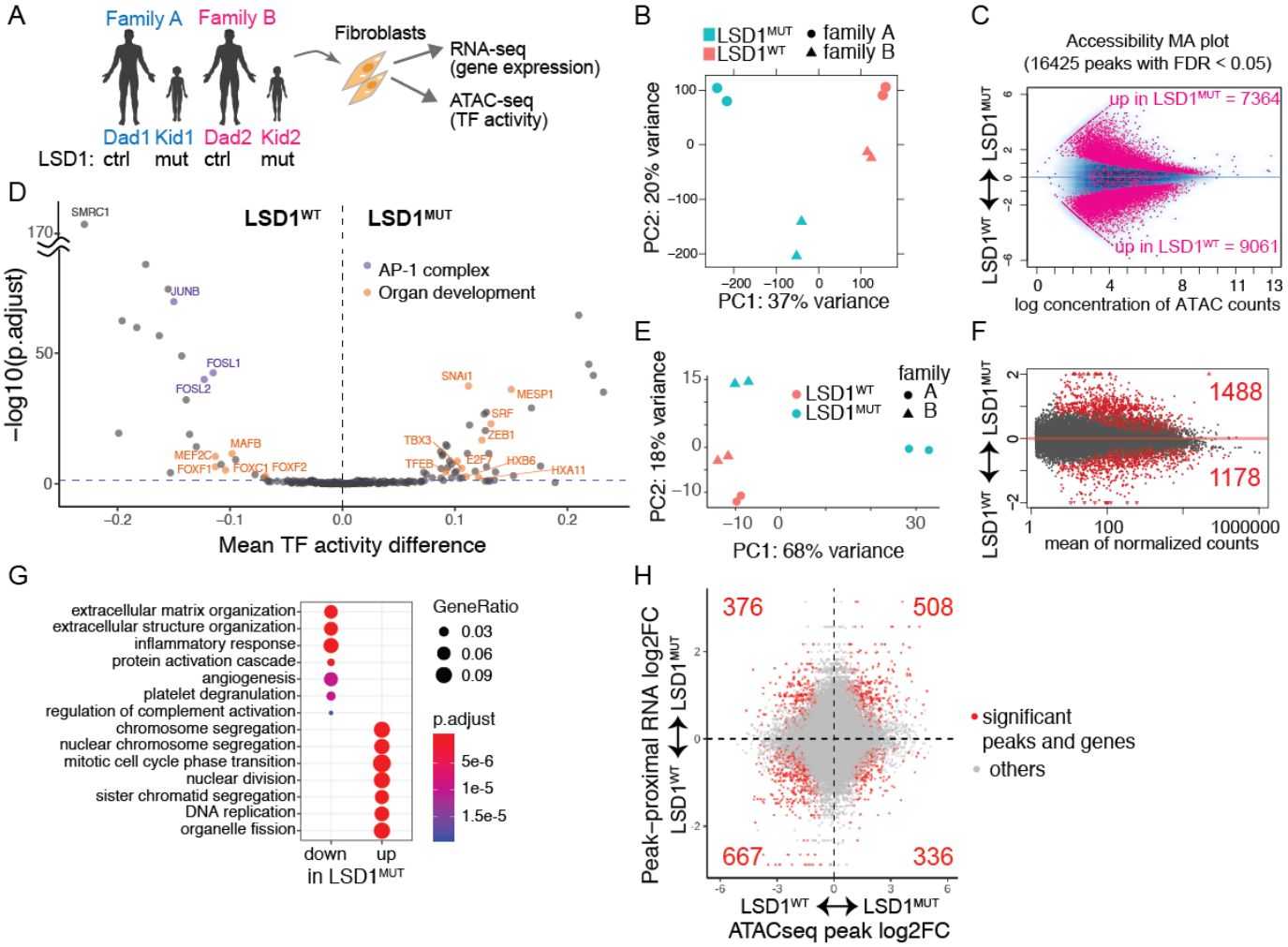
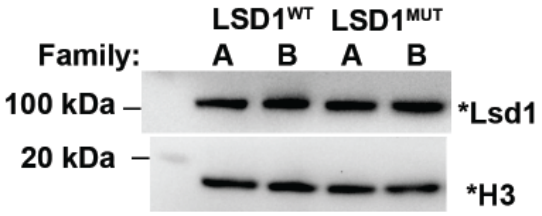
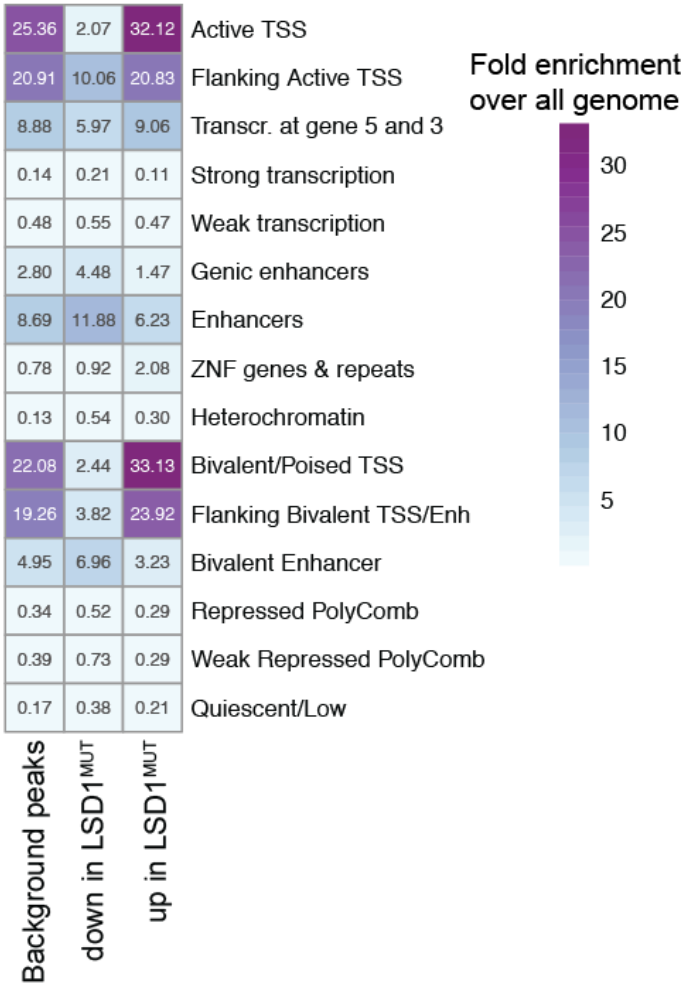


Figure 1. Effects of LSD1 mutation on chromatin accessibility and gene expression in fibroblasts. A) Schematic of the experimental set-up B) Principal component analysis (PCA) plot of ATAC-seq peak accessibility data in primary fibroblasts. Principal components 1 and 2 are shown together with the variance they explain. C) Differential accessibility analysis with MA plot showing the concentration of accessible fragments in the ATAC-seq peaks against log2 fold change of peak accessibility in LSD1^{MUT} vs. LSD1^{WT}. Significantly differential peaks (FDR < 0.05) are marked in pink. D) Differential TF activity analysis in primary fibroblasts comparing healthy controls (LSD1^{WT}) vs patients with LSD1^{MUT}. n=2 families, each 4 replicates. TFs colored by "AP-1 complex" (purple) and GO term "Embryonic organ development" (orange). E) PCA plot of RNA-seq data in fibroblasts F) Differential expression analysis in fibroblasts with MA plot showing average normalised gene expression values against log2 fold change in gene expression. Significantly differential genes (FDR < 0.1) are shown in red. G) gene ontology enrichments of significantly differential genes in fibroblasts H) Scatter plot showing correlations between ATAC-seq peak accessibility and gene expression changes in LSD1^{MUT}, where peaks were linked to the nearest gene. Significantly differentially accessible peaks and differentially expressed proximal genes are shown in red.

A



B



Supp.Fig.1. A) Western blot of whole cell extracts from primary fibroblasts showing Lsd1 protein expression (92 kDa size) across samples and replicates. B) Enrichments of non-differential ATAC-seq peaks and significantly up- or down-regulated peaks in $LSD1^{MUT}$ in 15 chromatin states identified in the Roadmap epigenome project for foreskin fibroblasts (E055). Fold enrichments relative to the whole genome background were calculated with ChromHMM software http://compbio.mit.edu/ChromHMM/.

GRN identified TF-gene modules affected by $LSD1^{MUT}$

Changes in chromatin accessibility and gene expression are elaborately intertwined. To dissect this relationship that might appear in both mutations and relate the changes in TF activity with the downstream effects on target gene expression, we constructed an enhancer-mediated gene regulatory network (GRN) based on co-variation across samples with paired RNA and ATAC-seq data[15]. Specifically, to connect the genes to their putative regulatory elements and TFs, we calculate correlations across samples between TF expression and ATAC-seq signal at their binding motifs (TF-to-peak correlations), and between ATAC-seq peaks and RNA expression of

nearby genes (peak-to-gene correlations, Fig. 2A and Methods). To obtain the TF-enhancer-gene links in the cell types relevant for the LSD1^{MUT} and CPRF syndrome phenotype, we used published datasets containing paired ATAC- and RNA-seq data of healthy brain samples and of a differentiation time course from pluripotent stem cells via mesoderm to cardiomyocytes from 2 independent studies, in addition to our data in fibroblasts (Table 1). After filtering for significant TF-to-peak and peak-to-gene connections, we obtained a network of 393 TFs linked to 15,128 unique genes through 506,564 regulatory elements (Fig. 2A and Methods). This network represents the global set of all TF-target gene connections in healthy, unperturbed cell types. Notably, we found only 13% of peaks linked to the nearest gene in terms of distance to TSS (Fig. 2B), consistent with the previous study reporting <20% of peaks [15], highlighting the importance of cell type specific peak-gene connections.

Source	# samples	Type	Link
This study	8	Fibroblasts	
Snyder/Murry lab	40	iPSC, mesoderm, cardiac mesoderm, cardiomyocyte	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE106690; https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE85332
PsychENCODE	11	Brain	RNAseq: https://www.synapse.org/#!Synapse:syn7062404 ATAC-seq: https://www.synapse.org/#!Synapse:syn7349497

Table 1.

Given the genes connected to each TF in the GRN, we defined these as TF regulons (TF-gene modules). For each TF, we assessed whether its regulon activity, i.e. the distribution of aggregated expression changes of all TF target genes compared in LSD1^{MUT} vs. LSD1^{WT} fibroblasts, was significantly different from the expression changes in all other genes (see Methods and SuppFig. 2A). Globally the distribution of gene expression of all 15,128 genes retained in the GRN was not affected by LSD1^{MUT} (mean log2 fold changes are centered around 0, Supp.Fig. 2B), however we found >100 TFs, whose regulons' expression (log2 fold change in LSD1^{MUT} vs. LSD1^{WT}) in fibroblasts deviated significantly from the background (Fig. 2C). Most of these regulons were upregulated in LSD1 mutated fibroblasts, consistent with an overall expected de-repression of genes due to the reduced functional effect in LSD1^{MUT} cells. Furthermore, de-repression of regulons of ZEB1 and GLIS1/2/3, which often function as repressive TFs, indicates perturbed repressor TF activity in LSD1^{MUT} fibroblasts. The differential activity of this set of TFs based on ATAC-seq data correlates well with the activity of their regulons from the GRN (pearson correlation: 0.5, Supp.Fig.2C), confirming the direct effects of LSD1^{MUT}-mediated TF motif accessibility disruption on TF target gene expression. Our GRN

therefore identified a specific set of TFs likely driving the changes in gene expression in LSD1^{MUT} fibroblasts, with most of these changes being upregulation of genes.

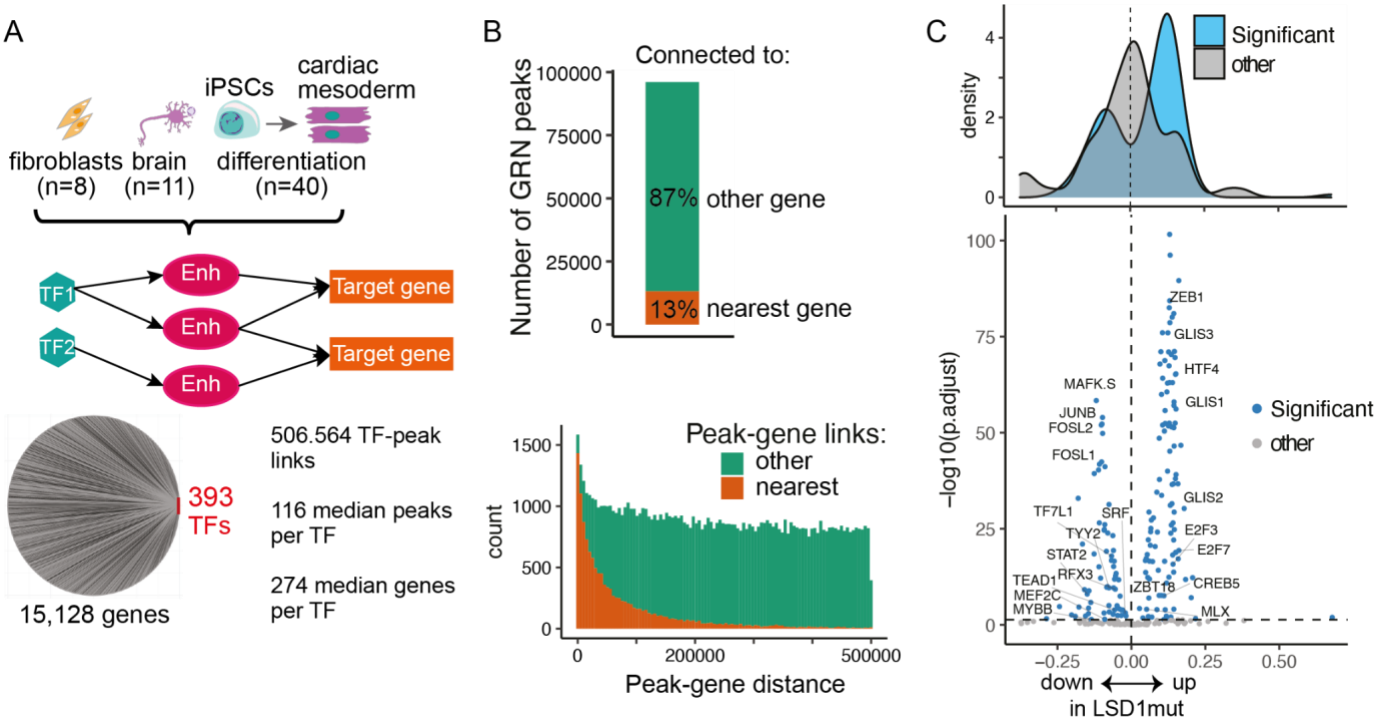
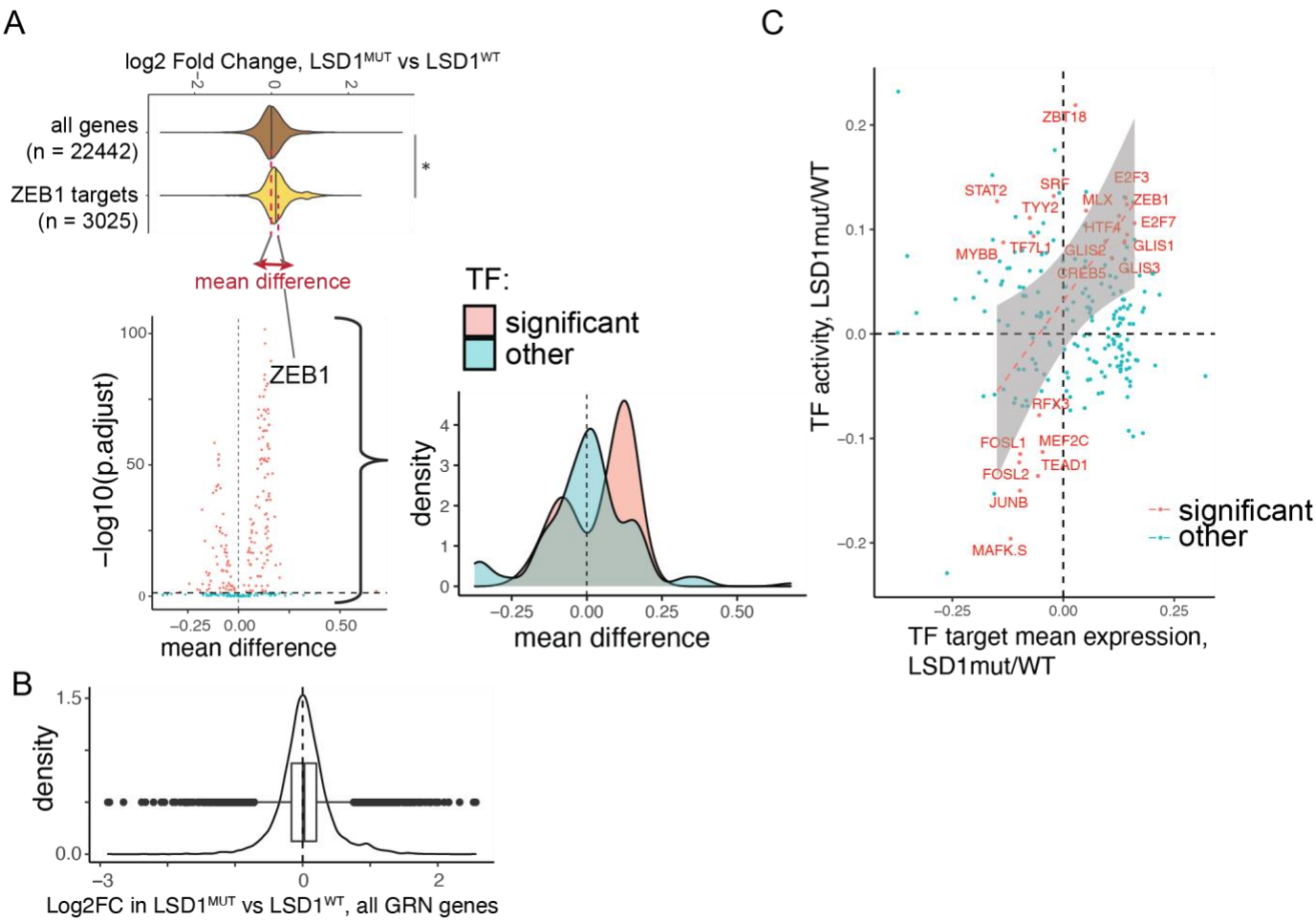


Figure 2. A) schematic of how the gene regulatory network (GRN) connects TFs and their target genes through enhancers. GRN statistics of the network resulting from healthy samples across 6 cell types and differentiation states (Table 1). B) Top panel: percentage of GRN peaks linked to the nearest gene or to the other gene, from the total of 96,050 peaks. Bottom panel: the distribution of distances between peak and gene in the GRN, colored by peaks linked to the nearest gene or other gene. C) For each TF, the mean differential expression in fibroblasts (LSD1^{MUT} vs LSD1^{WT}) of its target genes (=TF regulons) is shown as a volcano plot, and a density plot stratified for significant (blue) and non-significant (grey) TF regulons.



Supp.Fig.2. A) Violin plot showing that expression of ZEB1 regulons (its target genes) is on average increased compared to all expressed genes in fibroblasts and the difference is calculated as mean(ZEB1 regulon) - mean(all genes). Mean difference for each TF regulon is shown below as volcano plot against adjusted p-value. The distributions of the significant and other TFs are visualized by density plot. B) Distribution of gene expression of all unique genes in the GRN in fibroblasts (LSD1^{MUT} vs LSD1^{WT}). C) Scatter plot showing correlation of TF activity (defined by diffTF) and TF regulon expression (defined in panel A).

LSD1^{MUT} alters gene expression during early lineage differentiation

The role of LSD1 in decommissioning enhancer regions via removal of H3K4me1/2 marks is particularly important during cellular differentiation, when the enhancer landscape has to be re-established [16]. To study the differentiation of patient-derived cells, we reprogrammed patient- and control-derived primary fibroblasts of the two families described above into human induced pluripotent stem cells (hiPSCs). All hiPSCs clonal lines (n= 5, LSD1^{WT}; n = 4, LSD1^{MUT}) show normal chromosome numbers, intense staining of pluripotency markers (e.g., alkaline phosphatase and SSEA-4), and high expression levels of pluripotency marker genes (Supp.Fig.3A). We differentiated hiPSC lines into three main embryonic lineages: endoderm, mesoderm and neuronal lineage, and analyzed the effects of LSD1^{MUT} on gene expression (Fig. 3A). In the pluripotent state, we observed very limited changes in gene expression between LSD1^{MUT} and LSD1^{WT} (58 differentially expressed genes, Fig. 3B) partially due to the high variability between the hiPSC lines (SuppFig. 3B), a feature observed in reprogrammed hiPSC studies[17]. However, during differentiation into three cell lineages, LSD1^{MUT} affected a large

number of gene expression. We found over 4700 differential genes in endoderm, over 1600 in neuronal lineage and over 700 in mesoderm (FDR < 0.05, Fig. 3B), consistent with the phenotype of the affected kids who show symptoms related to all lineages (e.g., developmental delay, muscle hypotonia, distinctive facial features). Most of the differential genes were mis-regulated in only one cell lineage, with only 87 genes in common in all three cell lineages, but a larger set of genes (596) were commonly mis-regulated in endoderm and neuronal lineages. Down-regulated genes in endoderm and neuronal lineages were enriched in cell cycle related GO terms (DNA replication, sister chromatid segregation, SuppFig. 3D). Notably, unlike increased cell cycle genes found in one of the two LSD1^{MUT} individuals in primary fibroblasts, the downregulation of cell cycle related genes across differentiating lineages occurred in both LSD1^{MUT} samples implying more wide relevance in CPRF syndrome. Differentiation efficiency and the overall cell morphology in each lineage were not impaired by LSD1^{MUT}, consistent with comparable expression levels of lineage-specific marker genes (SuppFig. 3C), suggesting that cell cycle gene decreases are not caused by developmental defects.

Downregulation of genes in LSD1^{MUT} observed in differentiated endoderm, mesoderm and neuronal cells can result from less activation of up-regulated genes during differentiation or stronger repression of down-regulated genes. To distinguish these two scenarios, we compared gene expression changes in LSD1^{MUT} and LSD1^{WT} endoderm split into up- and down-regulated genes relative to expression levels in hiPSCs (Supp.Fig.3E). This analysis shows that genes, whose expression decreased in endoderm compared to hiPSCs, were even stronger downregulated in LSD1^{MUT} endoderm. Functional enrichment analysis among TF regulons in our subnetwork revealed significant overrepresentation of cell proliferation related categories such as DNA replication, translation and mitotic cell cycle (Supp. Fig.3F). As expected from reduced cell proliferation capacity when transitioning from hiPSCs to defined cell lineages, genes in these categories decreased in expression, but the decrease was even stronger in LSD1^{MUT} (Supp. Fig.3D).

GRN reveals TF regulons downregulated in LSD1^{MUT} at early lineage differentiation

Our GRN identified a specific set of TFs, whose regulons were mis-regulated in LSD1^{MUT} fibroblasts. To determine the effects of LSD1^{MUT} on TF-gene pairs in the other cell types, we performed the TF regulon analysis in LSD1^{MUT} vs. LSD1^{WT} hiPSCs and hiPSC-derived cell lineages. Similar to the fibroblast data, most TF regulons were upregulated in LSD1^{MUT} hiPSCs (Fig. 3C), consistent with the repressive function of LSD1. However, during early cell differentiation (from hiPSCs to endoderm, mesoderm and neuronal lineages) we observed the opposite: for the majority of TFs the regulons were downregulated in LSD1^{MUT} (Fig. 3C). To study more, we focused on TFs that were not differentially expressed across differentiations (FDR > 0.1 and absolute log2FC < 0.58) but whose regulon activity was significantly altered in LSD1^{MUT}. This revealed 144 TFs, which were then grouped into three classes based on unsupervised clustering of their regulon activity across cell types (Fig. 3D). Cluster 1 contains TFs with upregulated regulon activity in fibroblasts, unaltered in iPSCs, and downregulated in endoderm, mesoderm and neuronal lineage in LSD1^{MUT} vs LSD1^{WT}. Conversely, TF regulons of cluster 2 are downregulated in both fibroblasts and differentiation lineages, yet upregulated in hiPSCs. Cluster 3 contains TFs whose regulons are not altered in fibroblasts but downregulated in lineages. Taken together, while TF regulons are predominantly upregulated by LSD1^{MUT} in

fibroblasts and hiPSCs (Fig. 2C and Fig. 3C), these are separated with two distinct sets (clusters 1 and 2), suggesting that LSD1^{MUT} affects TF activity in a cell type-specific manner. In contrast, all TF regulons are repressed by LSD1^{MUT} in all three differentiating lineages. The effect size of the repression is the strongest in cluster 2 and overall more substantial in the endoderm (Fig. 3D).

To examine specific TFs and their regulons in full, we selected the top 10 TFs whose regulons were significantly altered in endoderm by LSD1^{MUT} (Fig. 3F). Four of these TFs are known chromatin interactors, for instance, autoimmune regulator AIRE selectively binds unmethylated H3K4 [18]; transcriptional coactivator ARID5B is involved in histone H3K9me2 de-methylation [19]; BPTF is a histone-binding component of a chromatin remodelling complex NURF, specifically recognizing H3K4me3 [20]; BCL6 is a transcriptional repressor which recruits histone deacetylases [21]. SOX11 and DLX1 are developmental regulators and HMGA2 is involved in cell cycle regulation. These 10 TFs are linked to 158 genes which are differentially expressed in LSD1 mutated endoderm and the majority of these genes (133) is regulated by only one of the 10 TFs (Fig. 3F). These data indicate that the TFs, whose regulons were most affected by LSD1^{MUT} during differentiation, might be directly relevant for LSD1^{MUT} phenotype due to their interaction with histones or histone-binding proteins and their role in regulating developmental processes.

Additionally, differences between distinct TF regulon clusters could be associated with specific genetic risk factors. To investigate if regulatory regions underlying TF regulon clusters 1-3 are enriched in common genetic traits, we performed linkage disequilibrium score regression (LDSC) analysis [22] using 442 GWAS catalogue traits (filtered by the number of SNPs and sample size of GWAS, Fig. 3E and Methods). TF regulons of cluster 1 are enriched for immune and blood-related GWAS traits (osteoarthritis, asthma, Crohn’s disease, haemoglobin and basophil counts), suggesting that the opposite TF regulon expression patterns in fibroblasts and in early differentiation are associated with immune system phenotypes. TF regulons of cluster 2, down-regulated in all cell types except iPSCs, are enriched in cardio-vascular and brain related traits (small vessel stroke, pulse pressure, conscientiousness measurement), pointing towards association of these traits with the TF regulons conserved across multiple cell types. TF regulons of cluster 3, which are down-regulated specifically in early differentiation, are enriched in neurological GWAS traits (social interaction measurement, educational attainment, handedness), confirming the neurodevelopmental phenotypes of the CPRF patients and indicating the relevance of disease-associated traits to early cell differentiation of patient-derived cell lines.

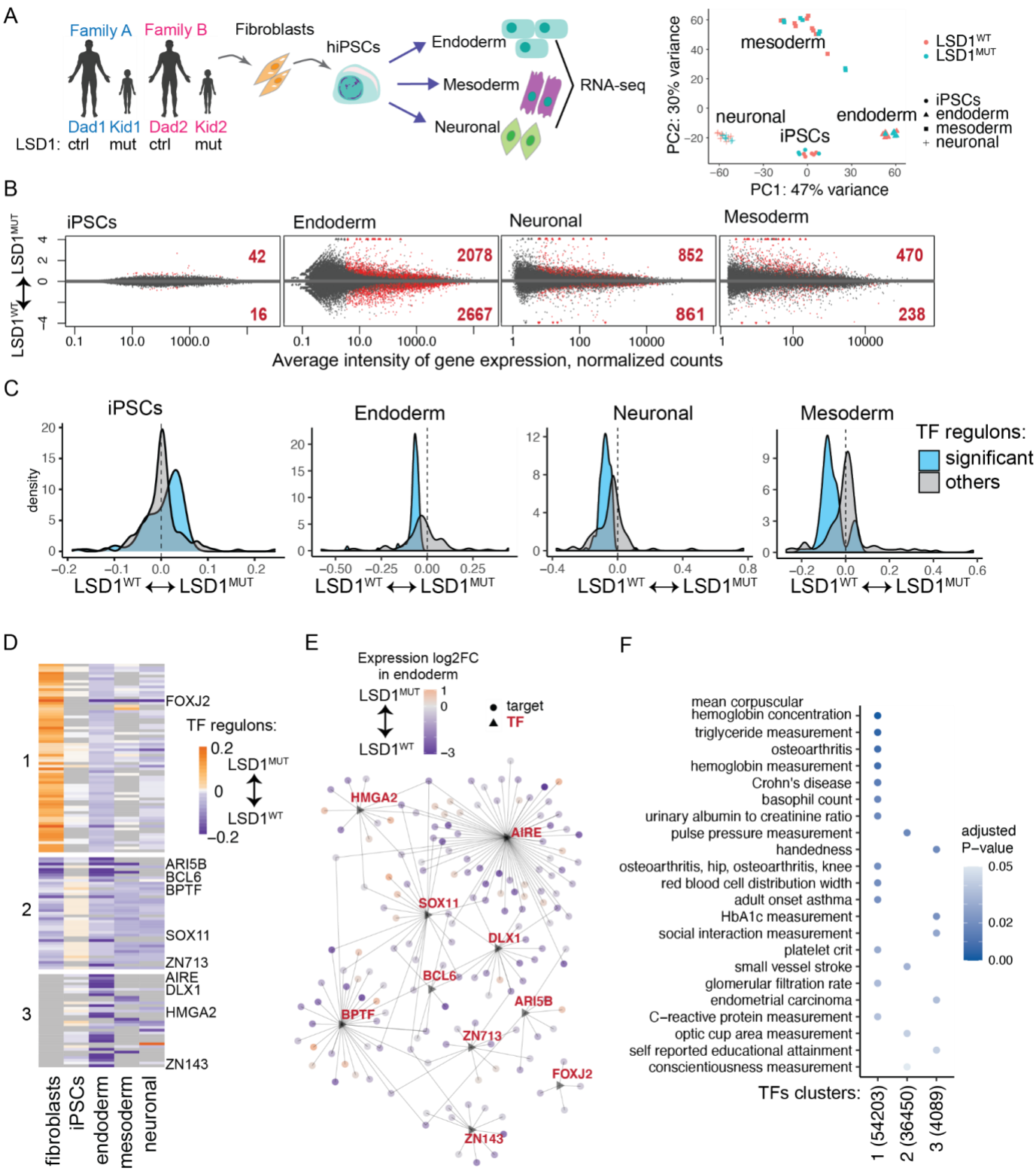
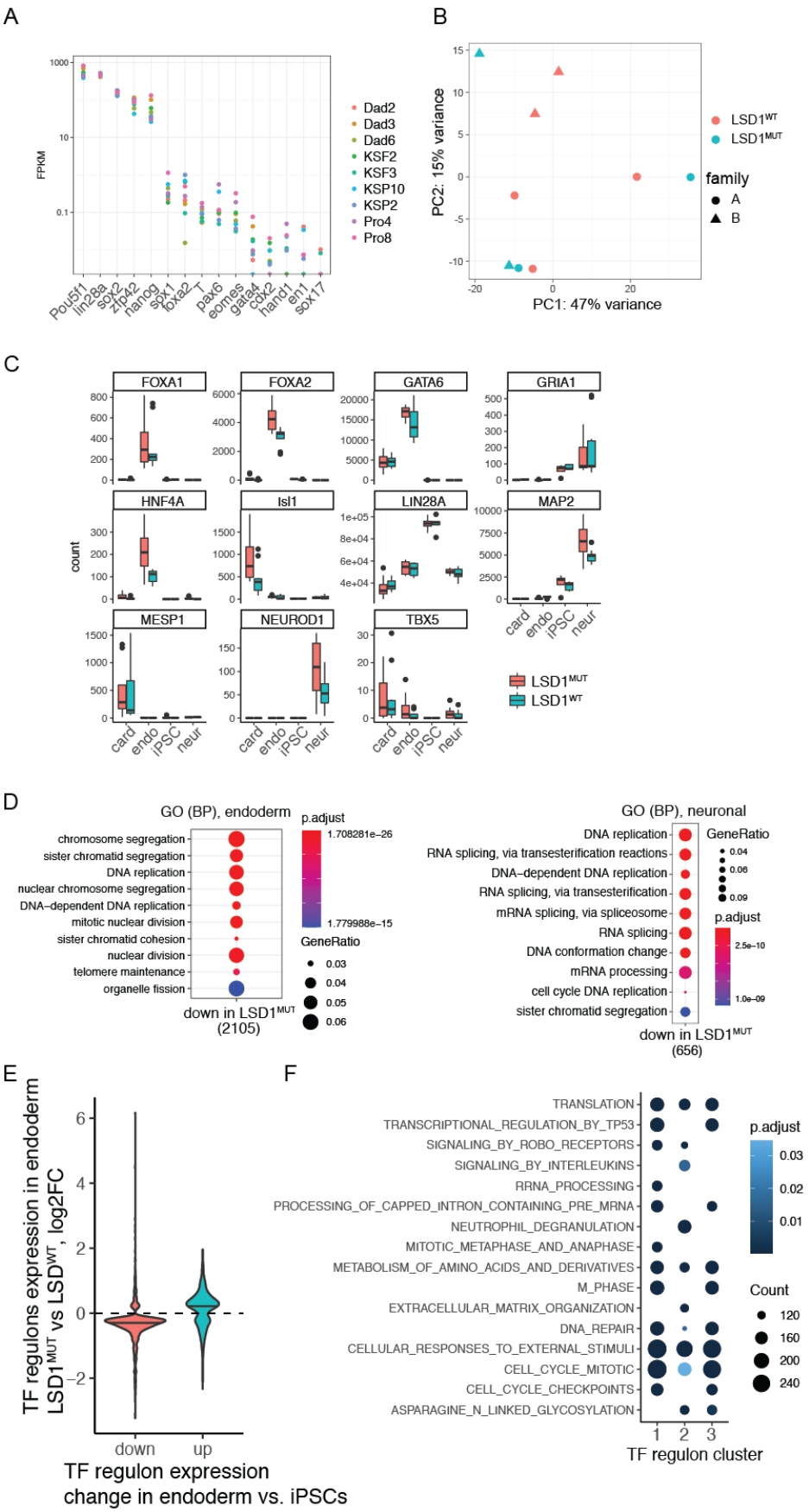


Figure 3. A) reprogramming of primary fibroblasts into iPSCs and subsequent differentiation into three main embryonic lineages, B) differential expression analysis of LSD1^{MUT} effects on early differentiation in three embryonic lineages, C) density plots of TF targets expression in LSD1^{MUT} iPSCs and derived from them endoderm, mesoderm and neuronal lineages. Significant TF targets are upregulated in iPSCs but downregulated in lineages, D) heatmap of a GRN subnetwork showing average expression changes of TF regulons in LSD1^{MUT} across cell types. This subnetwork includes 144 TFs that are differentially active and whose targets' expression is significantly changed in LSD1^{MUT} cell lineages. E) LDSC analysis results showing significant enrichments of GWAS traits in regulatory regions associated with TF regulons in clusters 1-3, see panel D. Adjusted p-values were calculated within each trait while including all GRN peaks as background. F) Graph representation of a GRN subset of the 10 TFs, whose regulons were most affected in endoderm differentiation. Only significantly differential TF targets (circles) are shown, colored by their expression log2 fold change in LSD1^{MUT} vs. LSD1^{WT}.



Supp.Fig.3. A) Expression of pluripotency markers (Pou5f1, Linc28a, Sox2, Nanog) and lineage markers (Sox1, T, Eomes, Foxa2, Pax6) to confirm the successful reprogramming of primary fibroblasts into iPSCs: induction of pluripotency markers expression and repression of lineage genes. Expression values are shown as FPKMs from the RNA-seq data. Cell lines: LSD1^{WT} - “Dad” (individual from family 1) and “KSF” (individual from family 2), 2-3 clonal

lines per individual; LSD1^{MUT} - “Pro” and “KSP”, 2 clonal lines per individual. B) PCA plot of top 300 variable genes in iPSC lines derived from LSD1^{MUT} and LSD1^{WT} patient samples. C) Expression of endoderm, mesoderm and neuronal markers in iPSC-derived early differentiated lineages. D) GO enrichment of significantly differential genes in LSD1^{MUT} lineages, only gene sets with enrichments having adjusted p-value < 0.1 are shown (none for mesoderm lineage). E) expression changes of significantly differential genes in LSD1^{MUT} or control endoderm cells, split into two categories: genes up- or down-regulated (by log2FC) in endoderm relative to iPSCs. The genes are stronger downregulated in LSD1^{MUT} endoderm differentiation compared to control samples. F) gene ontology (GO) enrichment of TF regulons in the three clusters from Figure 3D.

LSD1^{MUT} increases HDAC activity and repression of target genes

We wondered about the mechanism of the repression of TF regulons in LSD1^{MUT} in early differentiation. To investigate it, we considered two possibilities: dosage effect and mutation-specific effect (Fig. 4A). Dosage effect was observed in fibroblasts and hiPSCs, where TF regulons were upregulated by LSD1^{MUT} due to the reduction in LSD1-mediated repression, whereas in early differentiation, stronger repression of TF regulons in LSD1^{MUT} can be caused indirectly, e.g. by overexpression of repressive TFs. Indeed, 14 repressor TFs are upregulated in endoderm, mesoderm or neuronal lineage but none of them significantly in all three cell types (Supp.Fig.4A), whereas the global downregulation of TF regulons is found in all three lineages, suggesting that overexpression of repressors does not fully explain the molecular phenotype. The possibility of mutation-specific effect is based on a previous biochemical study of the CPRF patients’ mutations [13], which has shown that LSD1^{MUT} isoform still able to interact with its repressive co-factors (HDACs and Co-REST complex) and, to some extent, histone substrates. Because most of the LSD1 co-factors have a repressive role, their increased recruitment and/or activity in LSD1^{MUT} cells could lead to the decrease in histone acetylation and observed over-repression of TF regulons (Fig.4A).

To further characterize the dosage and mutation-specific effects we analyzed RNA-seq of normal hiPSCs treated with a chemical inhibitor of LSD1 acutely (for 24h) or chronically (for 7 days) or with shRNA against LSD1 to induce its knock-down (see Methods). Comparing gene expression changes in any of these perturbations (reduced dosage of LSD1) with the LSD1^{MUT} data of reprogrammed hiPSCs (mutation-specific condition) allows us to identify dosage and mutation-specific effects (Fig. 4B). Genes that were differentially expressed in any of these conditions grouped into 6 main clusters, with most genes (cluster 1, n=133) being upregulated, as expected from the loss of repressive function of LSD1 in perturbations (Fig. 4B). Among the upregulated genes was a transcriptional repressor Zeb1, whose increased expression can cause stronger downregulation of its targets and contribute to the phenotype we observed during hiPSCs differentiation (Fig. 3C). Notably, the second cluster (n=39 genes) contained genes that were specifically downregulated only in LSD1^{MUT}, pointing towards a mutation-specific regulation. Using published data of HDAC substrate H3K27ac in human stem cells (hESCs, see Methods) as an indication of HDAC activity on chromatin, we found only cluster 2 genes were linked to the regions with less H3K27ac signal (Fig. 4C). These results suggest that increased HDAC activity appears specifically in the LSD1^{MUT} cells and not in the cells with reduced dosage of LSD1.

VPA treatment in LSD1^{MUT} reverses altered gene expression

Having identified a connection between decreased gene expression in LSD1^{MUT} and increased HDAC activity, we hypothesized that chemical inhibition of HDAC activity might rescue the over-repression of TF regulons during differentiation. To test this, we set out the differentiation of hiPSC lines (n= 5, LSD1^{WT}; n = 4, LSD1^{MUT}) into the endoderm, treated differentiating cells with HDAC inhibitor valproic acid (VPA) or vehicle control, and measured changes in gene expression (LSD1^{MUT} vs. LSD1^{WT} cells). VPA is a broad range HDAC inhibitor, inhibiting class I and class IIa HDACs, thus increasing the levels of acetylated histones (Supp. Fig.4B). VPA has existing human safety data due to its application for the treatment of epilepsy and mood disorders, approved by the FDA [23]. While in the untreated endoderm we found 1170 differentially expressed genes in LSD1^{MUT} samples, the VPA treatment rescued expression of nearly all of these genes, with only 21 genes remaining significantly differentially expressed (Fig. 4D). Most of the rescued genes that were downregulated in LSD1^{MUT} increased upon VPA treatment and became closer to the LSD1^{WT} level (Fig. 4E).

Using the TF regulons, we identified top TFs linked to the genes rescued by VPA (Supp. Fig. 4C), among which we found several TFs related to cell proliferation and metabolism (SP1, SP3, SP4, Oct4, Nanog, E2F6, ESR1) and EMT (Zeb1, Snai2, Twist1). Among all TFs linked to the VPA-rescued differential genes we found a significant enrichment of TFs, whose regulons were affected by LSD1^{MUT} in the previous network based analysis (Fig. 3D, Fisher's test odds ratio 1.35, p-value = 2e-16), indicating that VPA treatment might have rescued the observed TF regulons impairment by LSD1^{MUT}. To explicitly assess global effects of the VPA treatment on TF regulons, we compared the expression changes of TF regulons in LSD1^{MUT} in untreated and VPA-treated endoderm, which revealed significant upregulation of TF regulons and partial rescue of the LSD1^{MUT} effects (Supp.Fig.4D). TFs whose regulons' expression was most rescued by VPA, largely overlap with the most affected TFs from Fig. 3E, including TFs related to cell growth and differentiation (Erf1, Dlx1, HMGA2), as well as chromatin interacting TFs (AIRE, BCL6 and BPTF, Fig.3E and 4F).
Together these results suggest that the effects of LSD1^{MUT} on early cell differentiation are partially related to HDAC-mediated over-repression of cell proliferation genes, mediated by a specific group of TFs.

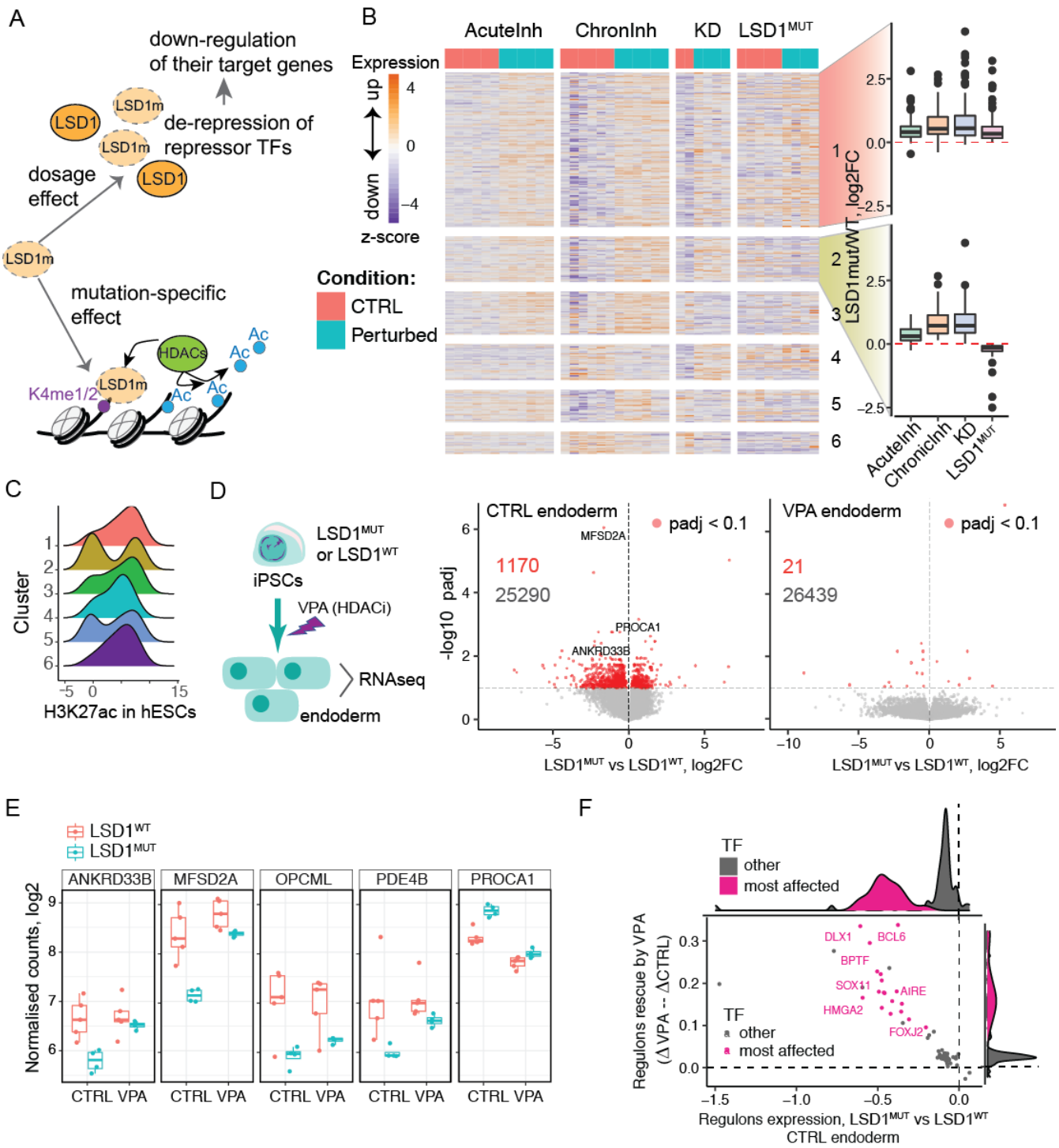
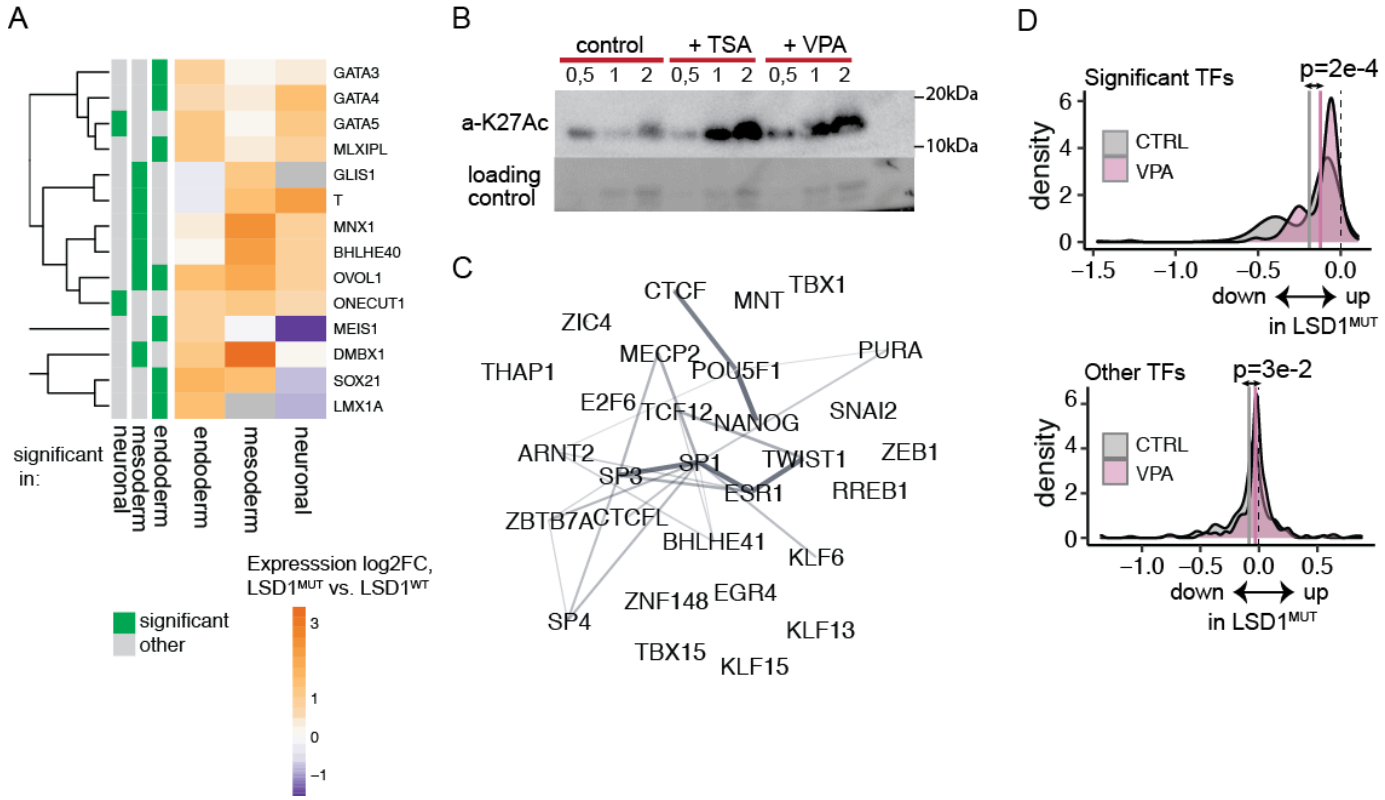


Figure 4. Increased HDAC activity leads to over-repression of TF targets in LSD1^{MUT}. A) scheme of dosage and mutation-specific effects in LSD1^{MUT}; B) heatmap of gene expression changes in the control and perturbed iPSCs (acute or chronic chemical inhibition of LSD1, knockdown of LSD1 by shRNA and CPRF patient-derived cells). Normalized expression values are shown as z-scores within each treatment batch. Boxplots of log2 fold changes between perturbed and control cells in each condition are shown on the right for the two first clusters. C) Density plots of H3K27ac signal in human embryonic stem cells (ENCODE) at regulatory regions linked to the genes in the heatmap on panel B. Cluster 2 shows strong bimodal distribution of low- and high-K27ac regions. D) CPRF patient-derived iPSCs were differentiated within 48h to endoderm in the presence of HDAC-inhibitor VPA and gene expression was measured by RNA-seq. Violin plots show the distribution of gene expression changes in LSD1^{MUT} vs. LSD1^{WT} control and VPA-treated endoderm. Significantly differential genes are highlighted in red. E) Expression counts of selected top differentially expressed genes in control endoderm. F) Scatter plot of significantly affected by LSD1^{MUT} TF regulons in control endoderm, showing the difference □ in TF regulon expression between LSD1^{MUT} and LSD1^{WT} control endoderm (X-axis) against the size effect of the VPA-mediated rescue of TF regulon expression (□ in VPA - □ in control endoderm).



Supp. Fig. 4. A) TF own expression in differentiation, log2 fold change between LSD1^{MUT} and LSD1^{WT}. Significant expression changes are indicated in green on the left columns. B) Western blot showing increased H3K27ac levels in patient-derived iPSCs treated with HDAC inhibitors TSA and VPA. C) Top TFs linked to differential genes in control endoderm that were rescued by VPA treatment (TFs ranked by a number of differential genes they are linked to). Connections are experimentally confirmed interactions from the STRING database [24]. D) Distributions of TF regulons expression changes in LSD1^{MUT} vs. LSD1^{WT} in control or VPA-treated endoderm (grey or purple color, respectively). TFs are split into those whose regulons are significantly affected in expression (top plot) or others (bottom plot). Average expression change across all TF regulons in control or VPA-treated endoderm are indicated as dashed vertical lines.

Discussion

Recently discovered de novo point mutations in a repressive histone lysine demethylase LSD1 cause a rare neurodevelopmental disorder CPRF, but the molecular mechanisms underlying the disorder are not known. Using patient-derived primary fibroblasts, iPSCs and differentiated cell lineages we studied the effects of two distinct CPRF mutations in LSD1 on epigenomic and transcriptional landscape, as well as on TF-target gene interactions. We found that in defined cell types (primary fibroblasts and iPSCs) LSD1^{MUT} leads to overall upregulation of TF regulons, consistent with the reduction of the repressive function of LSD1 upon mutations. In contrast, during early iPSCs differentiation to endoderm, mesoderm and neuronal lineage LSD1^{MUT} have a dominant effect of stronger repression of TF regulons, mostly affecting cell proliferation related genes. We have identified HDACs as important co-factors of mutated LSD1 and have shown that inhibition of HDAC activity by a chemical compound VPA partially rescues the repressive phenotype in LSD1^{MUT} endoderm differentiation.

Overall, these results show that LSD1^{MUT} is linked to misregulation of the global chromatin landscape, which affects motif accessibility of a specific set of TFs and expression of genes related to cell proliferation, differentiation and development.

LSD1 demethylation activity is inhibited in the presence of acetylated histones [25], suggesting that the acetylation levels at the LSD1-bound regions in different cell states (fibroblasts, iPSCs or during cell differentiation) could influence the effects of LSD1^{MUT}. During cell differentiation H3K27ac levels decrease at enhancers of proliferation-related genes [16] (due to reduced recruitment/activity of p300 and increased deacetylation by HDACs), enabling LSD1 to demethylate these regions and the inactive LSD1^{MUT} protein causes the LSD1-HDAC complex to reside longer at its target regions, causing even stronger repression of TF regulons. This transient effect is later compensated by the functional LSD1^{WT} copy after differentiation (e.g. in fibroblasts) and the upregulation of TF regulons is likely caused by overall increase in poised enhancers marked by H3K4me1 due to lack of LSD1^{MUT} demethylation activity and “aberrant” expression of the associated genes.

Although our analysis of gene expression and chromatin accessibility changes in LSD1^{MUT} across different cell types has shown mis-regulation of many TFs, genes and regulatory regions, there was no bias towards increased accessibility or expression, as would be expected from the impairment of a repressive chromatin factor. For example, a previous study has shown increased global chromatin accessibility in cancer cells treated with LSD1 inhibitors [26]. This suggests that the molecular phenotypes of LSD1^{MUT} identified in this work are mutation-specific and represent a combination of the direct effects and secondary effects caused by directly affected TFs or genes. Our GRN was able to identify enhancer-mediated TF-target gene connections, which are directly affected by LSD1^{MUT} and dissect the LSD1 effects into dosage loss of function of LSD1 and dominant effects of LSD1^{MUT}. Our data therefore suggests that the molecular effects underlying CPRF pathology cannot be fully dissected using only loss of function model systems due to profound mutation-specific transcriptional effects.

LSD1^{MUT} caused changes in activity of a specific set of TFs and in expression of their regulons in all analysed cell types, but the set of TFs and direction of the expression changes of their regulons varied between the cell types. Such cell type specific differences could be explained by distinct enhancers, which are functional at a particular cell state [27], or by relocation of LSD1 to different enhancers during cell state transitions. Moreover, interactions of LSD1 with its co-factors and TFs can vary between the cell types, for example, through tissue-specific TFs [28], introducing another source of variability in the molecular phenotype of LSD1^{MUT}. Sensitivity of a particular set of TFs to LSD1^{MUT} might reflect the susceptibility of TF binding sites to the chromatin environment altered by LSD1^{MUT}. Indeed, among the TFs most affected by LSD1^{MUT} during cell differentiation were chromatin-interacting factors AIRE and BPTF, which are sensitive to the methylation state of H3K4 when binding to chromatin, and BCL6 which recruits histone deacetylases.

Over-repression of TF regulons during cell differentiation is an unexpected dominant effect of a mutation in a repressive factor LSD1 and our study has shown that this effect is linked to HDACs, known co-factors of LSD1. HDAC inhibitors were previously found to rescue the deficiency in H3K4 methyltransferase KMT2D by balancing the lack of active H3K4 marks [29], whereas in this work the rescue of a repressor LSD1 impairment by inhibiting another repressor

(the HDAC) was not an anticipated effect. The success of HDAC inhibition points to higher activity or recruitment of HDACs in the presence of LSD1^{MUT} protein and further studies are required to dissect the exact details of this process. In the defined cell states (fibroblasts and iPSCs) we did not observe over-repression of target genes, suggesting that this molecular phenotype might be transient and compensated later by the intact copy of the LSD1. Nevertheless, the defects in early cell differentiation might have profound consequences on tissue formation and organism development.

Overall, we identified a group of TFs whose activity was impaired in cell-type specific manner due to LSD1^{MUT} and have shown that an FDA-approved drug VPA can rescue the TF regulons expression defects during cell differentiation caused by LSD1^{MUT}, which makes VPA a promising candidate for potential treatment of the CPRF syndrome patients.

Methods

Cell lines and culture conditions

Primary skin fibroblasts were obtained from 4 individuals … Fibroblasts were cultured at 37 C under 5% CO2 in DMEM media supplemented with 10% fetal bovine serum. iPSCs were generated from primary fibroblasts using Sendai virus reprogramming kit (Thermo Fisher), yielding 2-3 different clonal lines for each individual. iPSCs were maintained in Essential 8 (E8) media (LifeTech) on vitronectin coated plates following the manufacturer's instructions. iPSCs were passaged at least twice after thawing prior to the start of differentiation into endoderm, mesoderm and neuronal lineages. For endoderm differentiation we used a PSC Definitive Endoderm (DE) Induction Kit (Gibco, ThermoFisher Scientific Cat. No. A3062601), cells were collected after 2 days. For mesoderm differentiation we used a PSC cardiomyocyte differentiation kit (ThermoFisher Scientific Cat. No. A29212), cells were collected after 4 days. For neuronal differentiation we used a Gibco PSC neural induction medium (ThermoFisher Scientific Cat. No. A1647801) and cells were collected after 7 days. For the VPA treatment iPSCs were differentiated to endoderm using a 2-day DE Induction Kit following the manufacturer's instructions and from day 0 of differentiation 400uM VPA in DMSO was added to cell cultures.

Healthy human iPSC cell lines for Figure 4 were cultured under feeder-free conditions on plates coated with human-qualified Matrigel (BD Biosciences) diluted 1:40 in mTeSR-1 medium (StemCell Technologies). iPSCs were passaged upon treatment with Accutase (Sigma) and resuspending them in mTeSR-1 medium supplemented with 5 µM Y-27632 (Sigma) Six iPSC lines from 3 different individuals were treated with irreversible LSD1 inhibitor DDP26095 (10 uM), produced by the IEO Drug Discovery Unit–Drug Development Program (IEO DDU-DDP) for 24 h (acute inhibition) or 7 days (chronic inhibition). To induce a constitutive knock-down of LSD1, iPSC lines were transduced with two different validated pLKO.1 TRC vectors containing shRNAs (sh1 nnnnnnnnnn, sh2 nnnnnnnnn) against *kmd1a* and maintained with mTeSR-1 medium with puromycin (1 ug/ml) for selection. A pLKO.1 TRC vector containing a scrambled short hairpin sequence was used as a negative control.

Chromatin accessibility analysis by ATAC-seq

ATAC-seq experiment was performed on fresh fibroblasts in 2 biological replicates per individual according to the protocol [30]. Briefly, 50.000 cells were harvested and washed once in cold PBS for 5 min at 500 g at 4C. The pellet was gently resuspended in 50ul of cold lysis buffer containing 0.1% NP-40 and immediately spun down at 500 g for 10 min at 4 C. Transposition reaction mix containing transposition buffer and 2.5 ul of Tn5 enzyme per sample (Illumina) was added to the pellet and incubated for 30 min at 37 C. The reaction was stopped and DNA was purified using a Qiagen MinElute kit and eluted in 10ul of the EB buffer. DNA was PCR amplified for a total of 11-13 cycles using barcoded primers (Illumina Nextera XT Index Kit v2) and purified using Ampure beads (1.4:1 beads:sample ratio). The quality and concentrations of the eluted libraries was determined using an Agilent Bioanalyzer HS chip. Libraries were sequenced on Illumina NextSeq500 in paired-end mode.

ATAC-seq data analysis

ATAC-seq data were processed following the steps in [31] using the custom Snakemake [32] pipeline. Briefly, reads were trimmed using Trimmomatic [33] and aligned to the mouse genome (mm10) using Bowtie2 with parameters -X 2000 --very-sensitive. The read start sites were adjusted due to the transposon insertion specifics of Tn5. Mitochondrial reads and duplicated reads were removed with Picard tools, peak calling was performed using macs2 with parameters --nolambda --nomodel --slocal 10,000.
To look for differentially accessible regions a consensus peak set using samples from all individuals and replicates was produced using the R package DiffBind (minimum of 2 samples to retain a peak), resulting in 96,467 peaks. The accessibility signal was corrected with loess and significantly differentially accessible regions were defined by an FDR threshold of 0.05.

ChromHMM analysis (for Supp.Fig.1B). To analyse the distribution of differential ATAC-seq peaks across different chromatin states we downloaded a 15-state core ChromHMM model (computed on foreskin fibroblast data, cell type E055) https://egg2.wustl.edu/roadmap/web_portal/chr_state_learning.html. The enrichments of peaks (differential peaks more accessible in $LSD1^{MUT}$ or less accessible in $LSD1^{MUT}$ and non-differential peaks) were calculated using the ChromHMM software v.1.22 using the OverlapEnrichment command and default settings http://compbio.mit.edu/ChromHMM/.

TF activity analysis

Differential TF activity between $LSD1^{MUT}$ and $LSD1^{WT}$ was analysed as in [34] using diffTF tool [14]. Briefly, this method uses predicted TF binding sites defined by a genome-wide PWM scanner and overlaps these with accessible chromatin in ATAC-seq peaks. As a PWM database source, we used Jaspar CORE 2018 database (Khan et al., 2018), extended TF binding sites by 50 bp in each direction and calculated a fold change in ATAC-seq reads between $LSD1^{MUT}$ and $LSD1^{WT}$ conditions, such that the final TF activity value for each TF corresponds to the mean of genome-wide differences in accessibility at the TF's binding sites between the conditions. Statistical significance was calculated as a Cohen's distance of the weighted mean differences distribution to the matching distribution calculated for a permuted binding site of the corresponding TF. A list of significantly differentially active TFs with TF activity values is available in Table S1. Gene

ontology analysis on significant TFs was performed using clusterProfiler R package [35] and all TFs as a background.

Ageing-related TFs (as well as other genes in the subsequent RNA-seq analysis of the fibroblast data) were excluded from the fibroblasts analysis. The lists of ageing-related genes were taken from a curated database of human ageing genes GenAge https://genomics.senescence.info/genes/allgenes.php and from a list of genes whose expression was found significantly differential in ageing skin [36].

Gene expression profiling by RNA-seq

Samples for all RNA-seq experiments in this work were collected in two biological replicates from each of 2-3 clonal lines or individuals in case of primary fibroblasts (except for the VPA-treated endoderm where individual cell lines were used as biological replicates). RNA was isolated from 600.000-1.200.000 snap-frozen cell pellets using Qiagen RNeasy Mini kit and treated with TURBO DNA-free kit (Ambion) to remove DNA. The quality of RNA was verified on Agilent Bioanalyzer. 1ug of total RNA was used as input for poly-A selection (NEBNext PolyA mRNA magnetic isolation module), cDNA synthesis and library preparation with NEBNext Ultra RNA Library Prep Kit for Illumina according to the manufacturer's instructions. Library quality was determined using Agilent Bioanalyzer and concentration was measured by Qubit fluorometer. Libraries were sequenced on Illumina NextSeq500 in single-end mode.
RNA-seq reads were adapter trimmed using Trimmomatic and aligned to the human genome (hg19) with STAR aligner version 2.5.2b https://github.com/alexdobin/STAR using default parameters. Gene counts were produced with a summarizeOverlaps function from the GenomicAlignments R package and Gencode version 19 annotation of all human genes. Differential expression analysis was performed using an R package DESeq2 [37] in each cell type separately (fibroblasts, iPSCs, differentiating endoderm or mesoderm or neuronal cells) between $LSD1^{MUT}$ and $LSD1^{WT}$ cell lines. Fibroblasts: samples from primary cell cultures of two individuals grown in two different flasks were used as biological replicates (2x2=4 replicates per condition); iPSCs and VPA-treatment endoderm: 2-3 clonal lines derived from each individual were used as biological replicates; untreated endoderm and mesoderm and neuronal lineages: 2-3 clonal lines derived from each individual were grown in two flasks each and used as biological replicates($LSD1^{MUT}$: 2 x 2flasks x 2individuals = 8 replicates, $LSD1^{WT}$: 2 x 2flasks x 1individual + 3 x 2flasks x 1individual = 10 replicates). Significantly differential genes were defined by adjusted p-value cutoff < 0.1.
Gene ontology enrichment analyses were done with clusterProfiler R package [35] using MSigDB database of Reactome-Kegg pathways or "Biological process" GO annotation, p-adjusted cutoff < 0.1.

Integration of RNA-seq data of LSD1 inhibition and knock-down of normal iPSCs

Differential gene expression analysis on all iPSC samples (from the RNA-seq experiments on patient-derived iPSCs, as well as on normal human iPSC lines treated with LSD1 inhibitor or sh-RNA against LSD1) was performed using edgeR package [38] and glmRLT function with a model.matrix(~experiment+condition2), where "experiment" refers to $LSD1^{MUT}$/LSD1inhibition/LSD1knockdown and "condition2" is control samples vs. any of the

perturbations. Significantly differential genes in at least one of the conditions were defined by FDR < 0.05. For clustering, the edgeR log fold changes per condition were then sqrt-transformed (i.e. `sign(x)*sqrt(abs(x))`) to reduce the impact of a few very large fold changes, and clustered using cluster::clara with k=8. Two of the clusters were merged based on visual similarity and one cluster was not included in the visualization due to the very low number of genes it contained (only two genes), which resulted in 6 final clusters on Figure 4B.

Integration of publicly available H3K27ac ChIP-seq data

To estimate the levels of histone acetylation at the iPSC state across different gene clusters from Fig. 4B we downloaded raw H3K27ac ChIP-seq data of a matching cell type (human embryonic stem cells, cell line H1) from the Roadmap Epigenomics Project https://www.ncbi.nlm.nih.gov/sra/SRX012366[accn]. The raw sequences were adapter trimmed (using Trimmomatic v.0.33 [33]) and aligned to hg19 using bowtie2 (v.2.3.2 [39]) with parameters -X 2000 -t. We then calculated the levels of H3K27ac at the enhancers linked to the genes in six clusters from Fig. 4B using the DiffBind R package (version 3.0.15 [40]), which we visualized as density plots in Fig. 4C.

Gene regulatory network analysis

Gene regulatory network was constructed from the matched RNA-seq and ATAC-seq data of healthy human samples and cell lines described in Table 1 and Supp.Table S2 using an unpublished R package https://grp-zaugg.embl-community.io/grn/index.html (Kamal *et al, in preparation*). The studies were selected to include relevant cell types for the CPRF syndrome (brain tissues, differentiation time course of pluripotent stem cells, mesoderm lineage) and have paired RNA-seq and ATAC-seq data from the same samples of normal cells or tissues (unmutated/unperturbed). Raw input data was downloaded from the original publications and uniformly processed with the same pipelines for RNA-seq and ATAC-seq processing, as the fibroblast data described above. The resulting consensus set of ATAC-seq peaks across all 63 samples (peaks retained in the final peakset if they were identified in at least 2 samples) includes 500,000 peaks and counts within this consensus peakset were used as input for the network construction.

The computational approach to construct a gene regulatory network consists of two steps: building TF-enhancer links and enhancer-gene links. The details of this approach are described in [15]. Briefly, in the first step the expression of each TF was correlated with accessibility of each of the enhancer regions (=ATAC-seq peak) with and without a known binding site of the TF (foreground and background, respectively). Known binding sites were defined using the HOCOMOCO database v.10. Significantly correlated TF-enhancer links were identified using empirical FDR of 20% (calculated separately for each TF) and an absolute correlation Pearson's coefficient > 0.4. In the second step chromatin accessibility at the ATAC-seq peaks was correlated with the expression of all genes < 500kb away from the peak and peak-gene links were retained if they were positively and significantly (P < 0.05) correlated (our assumption is that accessibility at the regulatory region positively correlates with expression of the linked gene), and if their Pearson's correlation coefficient was > 0.4. This approach resulted in a total of 1.3 million TF-enhancer-gene links, connecting 393 TFs with 15,128 unique genes via 96,050 unique ATAC-seq peaks.

TF regulons expression analysis in LSD1^{MUT} vs. LSD1^{WT} cells

TF regulons were defined as all TF-gene links of each TF within the network. To identify which TF regulons were affected by LSD1^{MUT} we calculated aggregated mean expression log2 fold change of each TF regulon in LSD1^{MUT} vs. LSD1^{WT} and tested whether it differs significantly from all expressed genes (background) in the corresponding cell type (Student's t-test). The raw p-values from the test were corrected using Benjamini-Hochberg approach and TFs which passed the FDR < 0.1 were defined as having significantly affected regulons (Fig. 2C and 3C, Supp.Fig. 2A). The difference in means between TF regulons and background genes' expression was further used in the scatter/density plots and in the heatmap (Fig. 2D).

In the VPA treatment experiment significantly affected TF regulons were defined in untreated endoderm and their expression was calculated in VPA-treated samples, preserving the classification into significant and other TFs to keep the comparison between the same sets of TFs before and after treatment.

LDSC analysis

We performed stratified LDSC (S-LDSC [22]) to identify if any of the TF clusters were enriched for heritability in GWAS traits. GWAS summary statistics for 806 traits that included participants of European descent were downloaded from the GWAS catalog (https://www.ebi.ac.uk/gwas/) and harmonized. The summary statistics were converted into the LDSC format as described in XXX. Summary statistics with less than 10,000 individuals or less than 100,000 genetic markers were removed, leaving 442 GWAS traits. We downloaded 54 sets of genomic regions from XXX to serve as a background for the enrichment analyses. Peak sets were created by taking the enhancer regions of the full LSD1 GRN, as well as the enhancer regions connected to the TFs of clusters 1-3 and those connected to the top 10 TFs for the neuronal, endoderm and mesoderm lineages. The peak set for the full LSD1 GRN was added to the background regions, while the other enhancer regions were tested for enrichment of heritability. The LDSC enrichment P-values were corrected for multiple testing within each trait using the Benjamini Hochberg method.

References

1. Wang, P. *et al.* Global analysis of H3K4 methylation defines MLL family member targets and points to a role for MLL1-mediated H3K4 methylation in the regulation of transcriptional initiation by RNA polymerase II. *Mol. Cell. Biol.* **29**, 6074–6085 (2009).
2. Faucher, D. & Wellinger, R. J. Methylated H3K4, a transcription-associated histone

modification, is involved in the DNA damage response pathway. *PLoS Genet.* **6**, (2010).

3. Sha, Q.-Q. *et al.* CFP1 coordinates histone H3 lysine-4 trimethylation and meiotic cell cycle progression in mouse oocytes. *Nat. Commun.* **9**, 3477 (2018).
4. Sze, C. C. *et al.* Histone H3K4 methylation-dependent and -independent functions of Set1A/COMPASS in embryonic stem cell self-renewal and differentiation. *Genes Dev.* **31**, 1732–1737 (2017).
5. Shi, Y. *et al.* Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* **119**, 941–953 (2004).
6. Humphrey, G. W. *et al.* Stable histone deacetylase complexes distinguished by the presence of SANT domain proteins CoREST/kiaa0071 and Mta-L1. *J. Biol. Chem.* **276**, 6817–6824 (2001).
7. Wang, Y. *et al.* LSD1 is a subunit of the NuRD complex and targets the metastasis programs in breast cancer. *Cell* **138**, 660–672 (2009).
8. Wang, J. *et al.* Opposing LSD1 complexes function in developmental gene activation and repression programmes. *Nature* **446**, 882–887 (2007).
9. Wang, J. *et al.* The lysine demethylase LSD1 (KDM1) is required for maintenance of global DNA methylation. *Nat. Genet.* **41**, 125–129 (2009).
10. Kerenyi, M. A. *et al.* Histone demethylase Lsd1 represses hematopoietic stem and progenitor cell signatures during blood cell maturation. *elife* **2**, e00633 (2013).
11. Amente, S., Lania, L. & Majello, B. The histone LSD1 demethylase in stemness and cancer transcription programs. *Biochim. Biophys. Acta* **1829**, 981–986 (2013).
12. Chong, J. X. *et al.* Gene discovery for Mendelian conditions via social networking: de novo variants in KDM1A cause developmental delay and distinctive facial features. *Genet. Med.* **18**, 788–795 (2016).
13. Pilotto, S. *et al.* LSD1/KDM1A mutations associated to a newly described form of intellectual disability impair demethylase activity and binding to transcription factors. *Hum. Mol. Genet.*

25, 2578–2587 (2016).

14. Berest, I. *et al.* Quantification of Differential Transcription Factor Activity and Multiomics-Based Classification into Activators and Repressors: diffTF. *Cell Rep.* **29**, 3147-3159.e12 (2019).
15. Reyes-Palomares, A. *et al.* Remodeling of active endothelial enhancers is associated with aberrant gene-regulatory networks in pulmonary arterial hypertension. *Nat. Commun.* **11**, 1673 (2020).
16. Whyte, W. A. *et al.* Enhancer decommissioning by LSD1 during embryonic stem cell differentiation. *Nature* **482**, 221–225 (2012).
17. Bock, C. *et al.* Reference Maps of human ES and iPS cell variation enable high-throughput characterization of pluripotent cell lines. *Cell* **144**, 439–452 (2011).
18. Koh, A. S., Kingston, R. E., Benoist, C. & Mathis, D. Global relevance of Aire binding to hypomethylated lysine-4 of histone-3. *Proc Natl Acad Sci USA* **107**, 13016–13021 (2010).
19. Hata, K. *et al.* Arid5b facilitates chondrogenesis by recruiting the histone demethylase Phf2 to Sox9-regulated genes. *Nat. Commun.* **4**, 2850 (2013).
20. Li, H. *et al.* Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF. *Nature* **442**, 91–95 (2006).
21. Zhang, H., Okada, S., Hatano, M., Okabe, S. & Tokuhisa, T. A new functional domain of Bcl6 family that recruits histone deacetylases. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1540**, 188–200 (2001).
22. Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).
23. Sixto-López, Y., Bello, M. & Correa-Basurto, J. Exploring the inhibitory activity of valproic acid against the HDAC family using an MMGBSA approach. *J. Comput. Aided Mol. Des.* **34**, 857–878 (2020).
24. Szklarczyk, D. *et al.* STRING v11: protein-protein association networks with increased

coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**, D607–D613 (2019).

25. Shi, Y.-J. *et al.* Regulation of LSD1 histone demethylase activity by its associated factors. *Mol. Cell* **19**, 857–864 (2005).
26. Cusan, M. *et al.* LSD1 inhibition exerts its antileukemic effect by recommissioning PU.1- and C/EBPα-dependent enhancers in AML. *Blood* **131**, 1730–1742 (2018).
27. Heinz, S., Romanoski, C. E., Benner, C. & Glass, C. K. The selection and function of cell type-specific enhancers. *Nat. Rev. Mol. Cell Biol.* **16**, 144–154 (2015).
28. Yokoyama, A. *et al.* Identification of myelin transcription factor 1 (MyT1) as a subunit of the neural cell type-specific lysine-specific demethylase 1 (LSD1) complex. *J. Biol. Chem.* **289**, 18152–18162 (2014).
29. Bjornsson, H. T. *et al.* Histone deacetylase inhibition rescues structural and functional brain deficits in a mouse model of Kabuki syndrome. *Sci. Transl. Med.* **6**, 256ra135 (2014).
30. Buenrostro, J. D., Wu, B., Chang, H. Y. & Greenleaf, W. J. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Curr. Protoc. Mol. Biol.* **109**, 21.29.1-21.29.9 (2015).
31. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).
32. Köster, J. & Rahmann, S. Snakemake--a scalable bioinformatics workflow engine. *Bioinformatics* **28**, 2520–2522 (2012).
33. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
34. Bunina, D. *et al.* Genomic Rewiring of SOX2 Chromatin Interaction Network during Differentiation of ESCs to Postmitotic Neurons. *Cell Syst.* **10**, 480-494.e8 (2020).
35. Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R package for comparing

biological themes among gene clusters. *OMICS* **16**, 284–287 (2012).

36. Glass, D. *et al.* Gene expression changes with age in skin, adipose tissue, blood and brain. *Genome Biol.* **14**, R75 (2013).
37. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
38. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
39. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
40. Ross-Innes, C. S. *et al.* Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature* **481**, 389–393 (2012).