Published in final edited form as:

Nature. ; 534(7607): 387–390. doi:10.1038/nature18004.

CO-REPRESSOR CBFA2T2 REGULATES PLURIPOTENCY AND GERMLINE DEVELOPMENT

Shengjiang Tu[1,2], **Varun Narendra**[1,2], **Masashi Yamaji**[3], **Simon E Vidal**[4], **Luis Alejandro Rojas**[1,2], **Xiaoshi Wang**[5], **Sang Yong Kim**[6], **Benjamin A Garcia**[5], **Thomas Tuschl**[3], **Matthias Stadtfeld**[4], and **Danny Reinberg**[1,2,*]

[1]Howard Hughes Medical Institute, NYU School of Medicine, New York, NY 10016

[2]Department of Biochemistry and Molecular Pharmacology, NYU School of Medicine, New York, NY 10016

[3]Howard Hughes Medical Institute and Laboratory for RNA Molecular Biology, The Rockefeller University, New York, 10065

[4]Department of Cell Biology, NYU School of Medicine, New York, NY 10016

[5]Epigenetics Program, Department of Biochemistry and Biophysics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104

[6]Rodent Genetic Engineering Core, NYU School of Medicine, New York, NY 10016 USA

SUMMARY

*Correspondence and requests for materials should be addressed to D.R. danny.reinberg@nyumc.org.

Supplementary Information is available in the online version of the paper.

Author Contributions

S.T. designed and performed majority of experiments; V.N. analyzed ChIP-seq and RNA-seq data; S.V.V., S.T. and M.S. performed iPSC reprogramming experiments; X.S. and B.A.G. quantified histone modifications. S.K. did CRISPR zygotic injection. S.T., S.K., M.Y. and L.A.R. characterized mice. M.Y., S.T., and T.T. designed and performed PGC experiments. S.T., V.N., M.Y. and D.R. wrote the manuscript.

Sequencing Data are deposited at Gene Expression Omnibus with accession number GSE71676.

The authors declare no competing financial interests.

SI Figure 1: Source data of uncropped Western blot images with size marker
Supplementary Table 1
PRDM14 and CBFA2T2 target genes identified by ChIP-seq in NCCIT cell.
Supplementary Table 2
Differentially expressed genes in *Prdm14* and *Cbfa2t2* Knockout lines.
Supplementary Table 3
qPCR and PCR primers used in ChIP-qPCR and off-target verification.
Supplementary Table 4
PRDM14 and CBFA2T2 target genes identified by ChIP-seq in mouse ESC.
SI Source Data Table 1: Source data in excel file used in plots of Figure 1
SI Source Data Table 2: Source data in excel file used in plots of Figure 4
SI Source Data Table 3: Source data in excel file used in plots of Extended Data Fig. 1
SI Source Data Table 4: Source data in excel file used in plots of Extended Data Fig. 3
SI Source Data Table 5: Source data in excel file used in plots of Extended Data Fig. 5

Developmental specification of germ cells lies at the core of inheritance as germ cells contain all of the genetic and epigenetic information transmitted between generations. The critical developmental event distinguishing germline from somatic lineages is the differentiation of primordial germ cells (PGCs)[1,2], precursors of sex specific gametes that produce an entire organism upon fertilization. Germ cells toggle between uni- and pluripotent states as they exhibit their own "latent" form of pluripotency. For example, PGCs express a number of transcription factors (TFs) in common with embryonic stem cells (ESCs), including OCT4, SOX2, NANOG and PRDM14[2–4]. A biochemical mechanism by which these TFs converge on chromatin to produce the dramatic rearrangements underlying ESC- and PGC-specific transcriptional programs remains poorly understood. Here, we discover a novel co-repressor protein, CBFA2T2, that regulates pluripotency and germline specification. *Cbfa2t2*$^{-/-}$ mice display severe defects in PGC maturation and epigenetic reprogramming. CBFA2T2 forms a biochemical complex with PRDM14, a germline-specific transcription factor. Mechanistically, CBFA2T2 oligomerizes to form a scaffold upon which PRDM14 and OCT4 are stabilized on chromatin. Thus, in contrast to the traditional "passenger" role of a co-repressor, CBFA2T2 functions synergistically with TFs at the crossroads of the fundamental developmental plasticity between uni- and pluripotency

---

The germline first segregates from somatic lineages via the specification of PGCs between embryonic day (E) 6.25 to 7.25 in mice[1,2]. PRDM14 regulates pluripotency[5–7], and is the only known TF to specifically regulate germ cell specification[4]. To better understand the mechanism(s) underlying PGC development, we sought PRDM14-interacting proteins in the germ cell tumor cell line, NCCIT. NCCIT cells stably expressing FLAG-PRDM14 were subjected to affinity purification and proteomic analysis. Contrasting to previous reports[7–9], neither EZH2 nor other Polycomb Repressive Complex 2 (PRC2) components co-purified with PRDM14. Instead, the strongest identified interaction involved a co-repressor protein, CBFA2T2 (Extended Data Fig. 1a). Reciprocal affinity purification of FLAG-HA tagged CBFA2T2 confirmed strong interaction with PRDM14 (Extended Data Fig. 1a). CBFA2T2, CBFA2T3, and ETO comprise a homologous gene family frequently targeted for translocation events in acute myeloid leukemia[10–13]. Despite 85% sequence similarity among homologs, their ubiquitous expression and capacity to form hetero-tetramers[14–16], ETO and CBFA2T3 were barely detectable (Extended Data Fig. 1a). This specificity for CBFA2T2 aligns with published microarray data indicating that it is the only family member upregulated during iPSC reprogramming[17] and PGC specification[18].

Further reciprocal immunoprecipitations confirmed endogenous PRDM14 and CBFA2T2 interaction in both NCCIT cells and mouse ESCs (mESCs) (Fig. 1a, Extended Data Fig. 1b, 1c). Gel filtration of the FLAG eluate gave evidence of a larger than 600 kDa complex, (Extended Data Fig. 1d), possibly due to CBFA2T2 oligomerization[14]. Moreover, a GAL4 recruitment assay demonstrated that GAL4-PRDM14 recruited CBFA2T2, but not EZH2, to the chromatinized luciferase promoter (Extended Data Fig. 1e).

To ascertain PRDM14 and CBFA2T2 co-localization on chromatin genome-wide, we performed Chromatin Immunoprecipitation followed by deep sequencing (ChIP-seq) in WT NCCIT cells. In the case of CBFA2T2, 2,077 statistically enriched regions (SERs) were identified using a stringent p-value threshold of 1E-15, of which 1,384 overlapped with a

PRDM14 binding event (Extended Data Fig. 2a, Supplementary Table 1). Global mapping of SERs to nearest promoters identified 1022 PRDM14/CBFA2T2 co-targeted genes (Fig. 1b), many of which are TFs involved in lineage commitment (Extended Data Fig. 2b, 2c). In contrast, PRDM14 exhibited very limited overlap with PRC2 or Polycomb Repressive Complex 1 (PRC1) (Extended Data Fig. 2a). Interestingly, PRDM14 and CBFA2T2 co-bind the genomic loci from which they are transcribed (Fig. 1c).

The sequence-specific DNA-binding capacity of PRDM14 coupled to the co-repressor activity of CBFA2T2 suggested a hierarchical model of chromatin recruitment. We performed knockdowns of *PRDM14* or *CBFA2T2* using shRNAs or siRNAs. As expected, *PRDM14* knockdown resulted in a loss of CBFA2T2 localization at 11/11 common target genes (Fig. 1d). Surprisingly, *CBFA2T2* knockdown caused a reciprocal loss of PRDM14 binding to the same genes (Fig. 1e), with minimal effect on *PRDM14* expression (Extended Data Fig. 2d). Thus, PRDM14 localization to chromatin depends on its DNA binding activity and its association with CBFA2T2.

PRDM14 is required to repress lineage commitment genes and ensures naïve pluripotency in mESCs[6,7]. To examine such a role for CBFA2T2, we generated *Cbfa2t2* and *Prdm14* knockout (KO) cells in KH2 mESCs[19] using CRISPR/Cas9 genome editing[20]. gRNAs targeting the sixth exon (common to all *Cbfa2t2* isoforms) or the second exon of *Prdm14*, produced multiple lines harboring distinct frameshift mutations and loss of the targeted protein (Fig. 2a, Extended Data Fig. 3a, 3b). Colonies of *Prdm14* and *Cbfa2t2* KO mESCs displayed a flattened morphology (Extended Data Fig. 3c). Both mutant lines ceased to grow and could not be maintained in the absence of kinase inhibitors of MAPK/ERK and GSK3 (2i)[21] (Extended Data Fig. 3d), as shown in the case of *Prdm14* KO lines[7]. After exposure to 2i-free conditions, three different KO lines for both *Prdm14* and *Cbfa2t2*, alongside wild-type (WT) cells were subjected to RNA-seq analyses. 85% of genes differentially expressed in a *Prdm14* knockout setting were also dysregulated upon loss of *Cbfa2t2* expression (Fig. 2b, Extended Data 3e, Supplementary Table 2). Moreover, the directionality of differential gene expression was nearly identical across mutants (Fig. 2c, Extended Data Fig. 3f). In both KO ESCs, numerous pluripotency genes including *Klf4, Pou5f1, Nr0b1* (*Dax1*), *Lin28a* and *Myc* were downregulated, whereas lineage commitment genes such as *Elf3, Cdx1*, and *Pitx2* were upregulated. Similar to the case with PRDM14[5], CBFA2T2 overexpression enhanced iPSC reprogramming efficiency (Extended Data Fig. 3g, 3h). Thus, the CBFA2T2 co-repressor contributes positively to pluripotency.

Given that *Prdm14*$^{-/-}$ mice displayed a major defect in germline development[4], we tested the contribution of CBFA2T2 to both somatic and germline development by generating *Cbfa2t2* KO mice via CRISPR zygotic injection[22]. C57BL/6 zygotes were co-injected with *Cas9* mRNA and one of the gRNAs used in mESCs to target exon 6 of *Cbfa2t2* (Fig. 3a). We obtained multiple pups possessing a genetic lesion that caused a frameshift mutation and a dysfunctional truncated protein (Extended Data Fig. 4a). Genetic targeting was specific as the 10 most likely off-target genomic regions were unperturbed (Supplementary Table 3). Intercrossing of *Cbfa2t2*$^{+/-}$ mice produced pups in a roughly normal Mendelian ratio [25% +/+ (30), 58% +/− (70), 17% −/− (21)] and *Cbfa2t2*$^{-/-}$ animals appeared normal. However,

crosses of those mice (2 female and 3 male) with WT C57BL/6 counterparts failed to produce pups over two months.

To pinpoint the germline defect underlying the fertility phenotype of *Cbfa2t2*$^{-/-}$ mice, we analyzed anatomical and histological phenotypes of the reproductive organs. Female *Cbfa2t2*$^{-/-}$ adult mice have underdeveloped ovaries (Fig. 3b), exhibiting a total absence of follicles with hematoxylin and eosin (H&E) stainings (Fig. 3c). Similarly, testes of male *Cbfa2t2*$^{-/-}$ mice were reduced to ~30% of WT (Fig. 3d, Extended Data Fig. 4b). Total number of sperm was reduced to less than 10% of WT, while remaining sperm were largely immotile (Extended Data Fig. 4c) and unable to bind the *zonae pellucidae* of oocytes during *in vitro* fertilization. H&E staining of sections of *Cbfa2t2*$^{-/-}$ testes showed that 41% of seminiferous tubules did not contain spermatogenic cells (Fig. 3e). Furthermore, P0 male *Cbfa2t2*$^{-/-}$ testes were almost completely devoid of gonocytes (Extended Data Fig. 4d). This data contrasts a previous study claiming that *Cbfa2t2*$^{-/-}$ mice are fertile[23]. This discrepancy may be due to differing purity of the genetic background.

To understand the germline phenotype observed in both sexes, we examined PGC development in *Cbfa2t2*$^{-/-}$ embryos. Alkaline phosphatase (AP) staining of the genital ridge of E11.5 *Cbfa2t2*$^{-/-}$ embryos showed greater than 95% reduction in the number of PGCs relative to WT (Fig. 3f). This defect occurs even earlier, at E7.25-E8.75 (Fig. 3g, Extended Data Fig. 4e, 4f). In accordance, SOX2 is not activated in the mutants (Extended Data Fig. 4g). Thus, CBFA2T2 is a novel factor required for specification and development of PGCs, and the defect in this process results in germ cell depletion.

We next mapped the genomic localizations of PRDM14 and CBFA2T2 in mESCs, relative to that of OCT4, SOX2, and NANOG (OSN) from published ChIP-seq data[24]. CBFA2T2 and PRDM14 colocalize broadly across the genome in mESCs (Fig. 4a), and also exhibit significant overlap with OSN, as reported for PRDM14 in human ESCs[5]. As in the case of NCCIT cells, CBFA2T2/PRDM14 target genes include numerous lineage commitment TFs and chromatin regulators (Extended Data Fig. 5a, Supplementary Table 4), many of which are co-occupied by OCT4, including the histone H3K9 methyltransferase gene Ehmt1[25] (Fig. 4b).

The Nervy homology 2 domain (NHR2) of ETO – a CBFA2T2 homolog – is required for self-renewal of hematopoietic stem cells in leukemia[14]. NHR2 functions in homo- and heterotypic oligomerization by forming a four-helix bundle tetrameric structure. A seven amino acid ‘m7’ substitution within NHR2 disrupted oligomerization[14,16]. To test whether CBFA2T2 oligomerization contributes to ESC pluripotency, mESCs harboring the ‘m7’ mutation in CBFA2T2 were generated using CRISPR/Cas9 technology (Fig. 4c, Extended Data Fig. 5b). Similar to *Cbfa2t2*$^{-/-}$ cells, *Cbfa2t2-m7* cells exhibited a flattened morphology (Extended Data Fig. 5c) and a total abrogation of CBFA2T2 occupancy at a number of target genes (Fig. 4e). Furthermore, while PRDM14 and OCT4 protein levels were unperturbed as was biochemical interaction with PRDM14 (Extended Data Fig. 5d and Fig. 4d, respectively), CBFA2T2 oligomerization was required to stabilize PRDM14 and OCT4 on chromatin. ChIP-qPCR showed a significant reduction in PRDM14 and OCT4 occupancy across 12/12 target genes tested (Fig. 4f, 4g). Importantly, PRDM14/CBFA2T2-

independent OCT4 targets retained OCT4 binding (Extended Data Fig 5e). Thus, CBFA2T2 oligomerization is a critical molecular event underpinning a pluripotent network, providing a scaffolding function to stabilize essential TFs such as PRDM14 and OCT4 at their target sites.

CBFA2T2/PRDM14 targets comprise numerous components of the chromatin modifying machinery, such as EHMT1 (GLP) (Fig. 4b, Extended Data Fig. 5a, Supplementary Table 4). During PGC development, H3K9me2 levels are reduced[26], potentially due to repression of H3K9 methyltransferase EHMT1, via presently unknown mechanism[27]. Here, knockout of *Prdm14* or *Cbfa2t2* in mESCs caused derepression of *Ehmt1* (Extended Data Fig. 5f). Quantitative analysis showed a specific increase in H3K9me2 and H3K9me3 levels in *Prdm14*$^{-/-}$, *Cbfa2t2*$^{-/-}$ and *Cbfa2t2*-m7 mutant mESCs (Extended Data Fig. 5g). Importantly, CBFA2T2/PRDM14-mediated repression was required to maintain appropriate levels of H3K9me2 in PGCs *in vivo*. PGCs in E8.0 *Cbfa2t2*$^{-/-}$ embryos exhibited a specific GLP derepression (Fig. 4h), with resultant increased H3K9me2 levels at E8.75 (Fig. 4i, Extended Data Fig. 5h, 5i). Thus, direct control of global levels of chromatin modifications is likely another mechanism by which PRDM14 and CBFA2T2 regulate the delicate balance between self-renewal and lineage specification (Fig. 4j).

In summary, CBFA2T2, a co-repressor protein, is a novel factor regulating pluripotency and is essential for germline development. In contrast to the long held notion that co-repressors have a passive role in TF recruitment, CBFA2T2, without intrinsic DNA binding capacity, is required to stabilize both PRDM14 and OCT4 on chromatin via its oligomerization. While PRDM14 and OCT4 may independently bind DNA, their affinity based “on-rate” is insufficient to generate a functional regulatory influence on transcription. Instead, CBFA2T2 oligomerization provides a larger interaction surface to limit “off-rate” from chromatin, allowing for stable TF binding (Fig. 4j). Such a model may extend to numerous TFs for which associated co-repressors or co-activators have yet to be identified.

METHODS

No statistical methods were used to predetermine sample size. For mouse studies, no randomization or blinding was done.

Cell Lines and Cultures

NCCIT cells (CRL-2073) were obtained from ATCC and grown in RPMI1640 media with 10% FBS, L-glutamine, penicillin/streptomycin, and sodium pyruvate. 293Trex cells (ThermoFischer, #R710-07) were grown in DMEM with 10% FBS, L-glutamine, penicillin/streptomycin. KH2 ESCs, described previously[19], were grown in DMEM supplemented with 15% FBS, L-glutamine, penicillin/streptomycin, non-essential amino acids, 0.1 mM β-mercaptoethanol, LIF, and 2i inhibitors (1 μM MEK1/2 inhibitor (PD0325901) and 3 μM GSK3 inhibitor (CHIR99021)). On feeder conditions, 2i was omitted. Human fibroblast BJ cells were obtained from ATCC and maintained on fibroblast medium: DMEM knockout media with 10% FBS sera, 1% non-essential amino acids, L-glutamine, penicillin/streptomycin, non-essential amino acids, and 0.1 mM β-mercaptoethanol. Human iPS culture medium: Advanced DMEM/F12 +20% Knockout Serum replacement, L-glutamine,

penicillin/streptomycin, non-essential amino acids, 0.1 mM β-mercaptoethanol +10 ng/ml FGF2 (Peprotech). Cell lines were verified by Western Blots and PCR, and tested for mycoplasma contamination.

To generate PRDM14 or CBFA2T2 NCCIT stable lines, CAG-eGFP vector[28] was used to clone N-ter Flag tagged and C-term HA tagged target gene constructs. After 2 weeks puromycin selection, single colonies were picked and expanded. Similarly, Gal4-PRDM14 293 Trex stable line was generated by transfecting pcDNA4-T0 plasmid (ThermoFisher, #V1020-20) with an N-ter Gal4, C-ter HA fusion PRDM14 construct. For *Prdm14* or *Cbfa2t2* KO KH2 lines, gRNAs were cloned into pSpCas9 (BB)-2A-GFP vector (Addgene, px458)[20]. For *Prdm14* KO, the gRNA sequence was: GCGATGGCCTTACCGCCCTC. For *Cbfa2t2* KO, 2 gRNAs were used: ACTCTCTTGGCAGGCGGTTC and CTGGCCCCCAGGATTCATAA. For *Cbfa2t2* m7 knock-in lines, a gRNA sequence AGAGAAAACTAGGCGCTCCA targeting the NHR2 domain was chosen for cloning into the Cas9 vector. For this knock-in, a donor 723 bp gBlock DNA centered at the NHR2 domain sequence was PCR amplified and purified. Mouse KH2 ESCs were transfected with the above Cas9-gRNA-EGFP plasmids with Lipofecatamine 2000. In the case of m7 knock-in, the 723 bp template was included. Medium to high GFP population was FACS sorted and seeded at 20,000 cells per 15 cm plate. Seven days later, ESC single colonies were selected, expanded and genotyped.

Antibodies

Human PRDM14 antigen (N-ter residue 1-243) was generated by using PreScission protease to cleave the recombinant fusion protein GST-PRDM14(1-243). Rabbit polyclonal antibody was affinity purified by affi-gel 15 matrix conjugated with a his6 tagged antigen of PRDM14(1-243). Mouse PRDM14 antibody was described previously[29]. Briefly, N-ter construct (amino acids 1-231) of mouse Prdm14 was cloned into pET30a vector (Novagen, #69909-3). The corresponding His6 tagged protein was overexpressed and purified for rabbit polyclonal antigen production. The antibody was purified with Affi-gel 15 as mentioned above. In-house rabbit Ezh2 antibody was reported previously[30]. Other antibodies used in this study were from commercial sources with the following catalog numbers: Anti-CBFA2T2: Bethyl A303-593A, anti-OCT4: Santa Cruz #sc-5279, anti-Ring1B: Bethyl #A302-869A, anti-SUZ12: Cell Signaling #3737S, anti-HA: Abcam #ab9110, anti-tubulin Abcam #ab6046, anti-TRA-1-81 biotin: eBiosceince, #13-8883-82, Alexa Fluor 660 Streptavidin: LifeTechnologies s21377, HRP Streptavidin: Biolegend #405210, MVH: Abcam #ab13840.

Nuclear Extracts, Immunoprecipitation, and Affinity Purification

Nuclear extracts were prepared with buffer A and buffer C, essentially as described[31]. Cytosol fraction was removed by buffer A (20 mM Tris, pH 7.9, 10 mM KCl, 0.5 mM DTT, 0.2 mM PMSF, 1 μg/ml Pepstatin A, 1 μg/ml Leupeptin, 1 μg/ml Aprotinin). The pellet was resuspended in buffer C (20 mM HEPES, pH 7.5, 20% glycerol, 420 mM NaCl, 0.5 mM DTT, 0.2 mM EDTA, 0.2 mM PMSF, 1 μg/ml Pepstatin A, 1 μg/ml Leupeptin, 1 μg/ml Aprotinin) and snap frozen with liquid Nitrogen. For immnuoprecipitation, 2-5 μg antibody was incubated overnight with 0.8 mg nuclear extract and immobilized on 40 μl of protein A:

protein G beads (3:1 volume ratio). After 6 washes with BC350 (20 mM Tris, pH 7.9, 350 mM NaCl, 0.1% NP40, 0.2 mM PMSF, 0.2 mM DTT), the immunoprecipitate was separated by SDS-PAGE for Western Blot analysis. For loading controls, 5% of input nuclear extract was used. For FLAG affinity purification, 10 mg nuclear extract was incubated with 100 μl FLAG M2 beads overnight, and washed 6 times with BC350, as described above. Immunoprecipitate was eluted with 0.2 mg/ml Flag peptide in BC100 (20 mM Tris, pH 7.9, 100 mM NaCl, 0.2 mM PMSF, 0.2 mM DTT). The Flag eluate was run into an SDS-PAGE gel for 1 cm. The upper gel slices containing proteins were excised and subjected to trypsin digestion and LC-MS analysis. Digested peptides were desalted and concentrated using C18 stagetips for LC-MS/MS analysis. 120 minute gradients (6-75% acetonitrile) were used (nanoLC1000, Thermo Scientific) and spectra were recorded on an Orbitrap Velos (Thermo Scientific) by selecting the 15 most intense precursor ions for fragmentation in each full scan.

PRDM14 and CBFA2T2 Knockdowns

For *PRDM14* knockdown experiments, the following shRNA sequences from Open Biosystems TRC pLKO.1 shRNA libraries were used:

Human *PRDM14* shRNA 1: TTCTGTAGTGTCCATAGGACG

Human *PRDM14* shRNA 2: AACATGAAGAATGTGGATCCG

Human *PRDM14* shRNA 3: TTGAAGGGAGTCTTTATCCAG

Lentiviruses from the above shRNAs, as well as empty pLKO.1 vector were produced from 293T cells. Four million 293T cells were seeded on a 10 cm plate. Next day, 2.3 μg plasmid (shRNA or control), 1.6 μg psPAX2, and 1.1 μg pMD2.G 2nd generation packaging plasmids (Addgene, #12260, #12259) were transfected with Lipofectamine 2000. 48 hrs and 60 hrs post-transfection, supernatants were collected. Viral particles were filtered through 0.45 μM filters and enriched 100-fold by centrifugation at 20000 rpm for 1.5 hrs. For transduction of NCCIT cells by lentivirus, 0.20 million cells/well were seeded in 6-well plates. The next day, 20 μl virus was transduced with polybrene at a final concentration of 8 μg/ml. 48 hrs post-tranduction, 1 μg/ml Puromycin was added to select for transduced NCCIT cells. Transduced cells were expanded and harvested in one week for ChIP-qPCR.

For *CBFA2T2* knockdown experiments, the following On-TARGETplus siRNA sequences from Thermo Scientific were used:

GAUCAUCGUUUGACAGAAA,

CAGAUUCUCUCAGCAAUGA,

UAGAGGAUAUUGCAACUUC,

CCACAGAGAUUCAGCAAUG.

Qiagen AllStars negative siRNA was used as control. siRNA (final concentration, 10 nM) was transfected with 4 μl Lipofectamine RNAiMAX (Life Technologies) in each well of 6-well plates. 70 hrs post-transfection, cells were harvested for Western Blot analysis. For

ChIP chromatin preparation, *CBFA2T2* siRNA transfection was scaled up to two 10 cm plates. Cells were split once before harvesting chromatin.

ChIP-qPCR and ChIP-seq

Chromatin immunoprecipitation (ChIP) was done as described with biological replicates[32]. Briefly, cells were crosslinked with 1% formaldehyde, lysed, and sonicated in buffer 3 (10 mM Tris, pH 7.9, 1 mM EDTA, 0.5 mM EGTA, 0.5% N-lauroylsarcosine) down to a desired chromatin size. 40 µl of 3:1 mixture of protein A and protein G Sepharose beads were blocked with 0.1 mg/ml salmon sperm DNA and 1 mg/ml BSA. Approximately 100 µg chromatin was pre-cleared by half of the blocked beads, and incubated with 2-5 µg antibody overnight in Tris buffer (10 mM Tris, 10 mM EDTA, 1% Triton X-100, 0.1% sodium deoxycholate (DOC), and protease inhibitors). After 4 hrs incubation with the remaining protein A/G beads, samples were washed 6 times with RIPA buffer (0.5 M LiCl, 50 mM HEPES, pH 7.5, 1 mM EDTA, 1% NP-40, 0.7% DOC, and protease inhibitors). After a brief wash with TE buffer (10 mM Tris, 1 mM EDTA, 50 mM NaCl), samples were resuspended in 200 µl of $T_{50}E_{10}S_1$ (50 mM Tris, pH 8.0, 10 mM EDTA, 1% SDS) and incubated at 65° C overnight to reverse crosslinks. Samples were digested at 55° C for 3 hrs with 10 µg each of RNAse A and proteinase K. Digested samples were PCR column purified and diluted into 300 µl water for qPCR. For qPCR quantification, 5 µl SYBR Green I Master mix (Roche), ROX reference dye, 3 µl water, 1 µl 5 µM primer pair, and 1 µl DNA were mixed for PCR amplification. For Gal4-ChIP-qPCR, the primer sequences were: Gal4ChIPLucP5F: CACCGAGCGACCCTGCATAAGC and Gal4ChIPLucP5R: GCT TCT GCC AAC CGA ACG GAC. Other qPCR primers are listed in Supplementary Table 3.

For PRDM14, CBFA2T2, and OCT4 ChIP-Seq, 40 µl protein G beads per IP were used, and salmon sperm DNA was omitted in the blocking step. DNA was eluted in 30 µl elution buffer for library construction as described below[33].

RNA-seq

Mouse KH2 cells and KO lines were grown on feeders for 3 consecutive passages on 6-well plates with FBS sera media supplemented with LIF. After the 4th passage (Day 5), feeder cells were removed by trypsinizing and then plating the cells on a T25 flask for 35 min. The un-attached ESCs were spun down and lysed with TRIzol. Standard RNA-Seq procedure was used[33].

Library Construction

Libraries for ChIP-seq were prepared according to manufacturer's instructions (Illumina). Briefly, IP'ed DNA (~5 ng) was end-repaired using End-It Repair Kit (Epicenter), tailed with deoxyadenine using Klenow exo- (New England Biolabs), and ligated to custom adapters with T4 Rapid DNA Ligase (Enzymatics). Fragments of 200-400 bp were size-selected using Agencourt AMPure XP beads, and subjected to PCR amplification using Q5 DNA polymerase (New England Biolabs). Libraries were quantified by qPCR using primers annealing to the adapter sequence and sequenced at a concentration of 12 pM on an Illumina HiSeq. Barcodes were utilized for multiplexing. For RNA-seq libraries, polyA+ RNA was isolated using Dynabeads Oligo(dT)25 (Invitrogen) and constructed into strand-specific

libraries using the dUTP method. Once dUTP-marked double-stranded cDNA was obtained, the remaining library construction steps followed the same protocol as described above for ChIP-seq libraries.

Bioinformatic Analysis

Sequenced reads from ChIP-seq experiments were mapped to the hg19 or mm9 genome with Bowtie, using the parameters –v2 and –m4. Normalized genome-wide read densities were computed using a custom script and visualized on the UCSC genome browser after extending to the estimated size of a ChIP fragment (~200nt). Enriched regions (ERs) were identified using MACS 1.40rc2 with default parameters and an input control, and then filtered for enriched regions (ERs) with at least 10 tags and an unadjusted p-value of <1E-15 for ChIPs performed in NCCIT cells, and <1E-5 for ChIPs performed in KH2 mESCs. ERs were associated to gene targets using the HOMER tool. Heatmaps were generated using a custom code in which reads were mapped to non-overlapping 10 bp bins within peak-centered windows of 5–10 kb. Normalized cumulative read density across these bins is depicted. All GO analysis was performed using the DAVID tool[34].

RNA-seq reads were assigned to genes using DEGseq (R package)[35] and the ENSEMBL annotation. FDR adjusted p-values for differential gene expression were calculated with DEseq (R package). Genes were considered “differentially expressed” if their adjusted p-value was < 0.00001.

Human iPSC Reprogramming

PRDM14 and CBFA2T2 lentiviral constructs were cloned into pHAGE-EF1α-IRES-td tomato and pHAGE-EF1α-IRES-zGreen vectors, respectively. OKSM polycistronic vector under EF1α control was used for the reprogramming experiments[36]. Lentiviruses were produced as described previously[37]. Supernatants were collected every 12 hrs on two consecutives days starting 48 hrs after transfection. Viral particles were concentrated by centrifugation at 20000 rpm for 1.5 hrs. Virus titer was quantified by flow cytometry and immunofluorescence, and high titer viruses (greater than 60% transduction efficiency) were chosen for further human iPSC reprogramming experiments. Human fibroblast BJ cells were seeded at 0.14 million cells/well of a 6-well plate. 28 hrs after seeding, 10–15ul of concentrated OKSM, OKSM+PRDM14, or OSKM+CBFA2T2 virus combinations were used to transduce human fibroblasts with polybrene (final concentration of 8 μg/ml). 40 hrs post-transduction, td-Tomato and GFP positive cells were sorted and seeded onto irradiated MEF feeders (GLOBASTEM). Cells undergoing reprogramming were maintained on human fibroblast medium for the first week, and transferred into 100% human iPS culture medium at the end of the second week. At the end of the third week on feeders, human iPS cells were tested for live Tra-1-81 staining. Anti-human TRA-1-81 (Podocalyxin) Biotin solution (1:200 dilution in 4% FBS PBS solution) was directly added to each well. After washing, fluorescent secondary antibody Alexa Fluor 660 Streptavidin (1:200) was used for imaging. For colony counting, TRA-1-81 staining with secondary antibody conjugated with peroxidase HRP was used with substrate DAB (Vector Labs, #SK-4100)[38]. Error bars are based on 3 biological replicates of each condition.

Cbfa2t2 KO Mice

We generated *Cbfa2t2* KO mice via zygotic injection[22]. T7-gRNA DNA template was PCR amplified from one of the Cas9-gRNA plasmids for generation of *Cbfa2t2* KO ESC. The gRNA sequence is ACTCTCTTGGCAGGCGGTTC. The primer sequences are ttaatacgactcactatagGGAGAATGGACTATAAGGACCACGAC and GCGAGCTCTAGGAATTCTTAC. Subsequently, T7-gRNA was generated by *in vitro* transcription with MEGAshortscript T7 kit. Similarly, Cas9 mRNA was generated with mMESSAGE mMACHINE T7 ULTRA kit (Life Technologies). Injection mix contained 100 ng/μl Cas9 mRNA, 50 ng/μl CBFA2T2 KO gRNA. Cytoplasmic injection was performed on 102 C57BL/6 zygotes. Of those, 72 embryos were transferred to three pseudopregnant female mice. A total of 30 pups were born and genotyped. Genotyping primers were TAGCAGTCTTCCTGCTTTGG and CTTCTCGGTGTTCTAGCATCTT. Ten top potential off-target sites were tested by PCR sequencing. The 10 primer pair sequences are listed in Supplementary Table 3. Crossing CRISPR mutant mice containing one allele of in-del mutation with WT C57BL/6 mice resulted in *Cbfa2t2*$^{+/-}$ mice. Intercrossing of these mice produced full *Cbfa2t2* KO mice (*Cbfa2t2*$^{-/-}$). Mouse studies were approved by NYUMC IACUC.

Sperm Counts and Motility Analysis and *In Vitro* Fertilization (IVF)

Individual caudal epididymis was minced in 90 μl MBCD medium. After 30 min incubation at 37 °C, sperm were separated by pipetting and passaging through a 70 μm filter. Sperm counts and motility assessment were performed by using the DRM-600 CELL-VU Sperm Counting Cytometer. For IVF, 28 egg donors were used in each experiment.

Tissue Staining of Sections

Mutant testes were weighed before fixation. Ovaries and Testes were fixed in Bouin's fixative for 2 to 6 hrs, washed with PBS overnight, dehydrated with ethanol solution, embedded in paraffin and sectioned at 5 μm. Sections were stained by hematoxylin-eosin. Postnatal day 0 (P0) testes were fixed with 4% PFA for 15 min. Slides of 10 μm cryosections were stained with MVH antibody (1:250).

Whole Mount Immunofluorescence Analysis and AP Staining

Embryo isolation and staging were done as described previously[4]. The immunofluorescence analysis and AP staining were performed essentially as described previously[4]. The primary antibodies used were as follows: anti-AP2γ rabbit polyclonal, 1:500 (cat#sc-8977; Santa Cruz Biotechnology); anti-ETMT1/GLP mouse monoclonal, 5μg/mL (cat#PP-B0422-00; R&D Systems); anti-H3K9me2 mouse monoclonal, 1:500 (cat#ab1220; Abcam); anti-SOX2 goat polyclonal, 1:200 (cat#sc-17320, Santa Cruz Biotechnology). The following secondary antibodies from Molecular Probes (Eugene, OR) were used at a dilution of 1:500: Alexa Fluor 488 goat anti-rabbit IgG; Alexa Fluor 568 goat anti-mouse IgG. The stained embryos were mounted with VECTASHIELD Antifade Mounting Medium (cat#H-1000; VECTOR LABORATORIES). The immunofluorescence images and the AP staining images were taken by a confocal microscope (Zeiss LSM880) and a stereomicroscope (Leica M80), respectively. The image analyses were done by using ImageJ/Fiji software.

Histone Modification Quantification

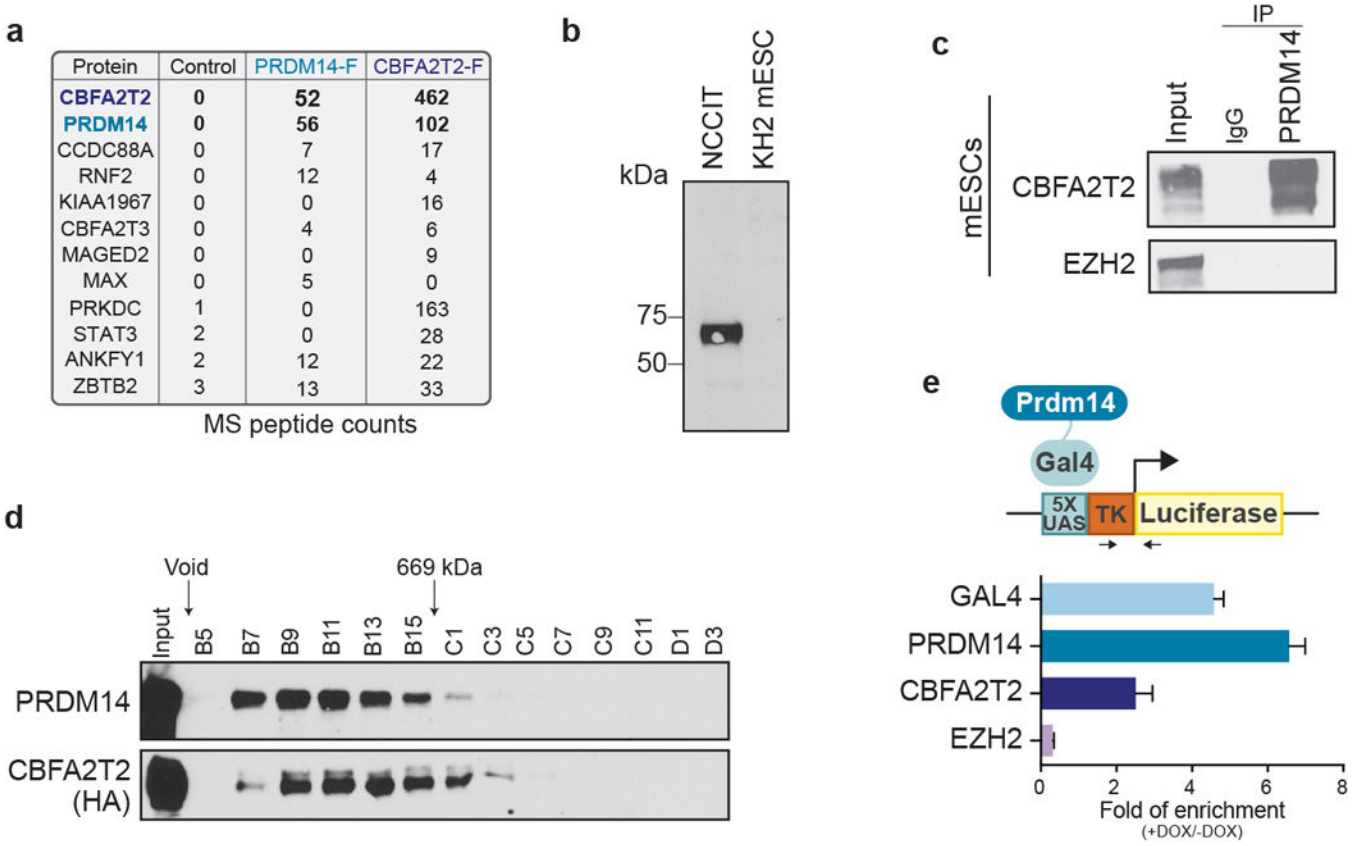
Histones from Mouse KH2 and KO mutant ESCs were purified by acid extraction[39]. Approximately 100 μg histones were derivatized with propionic anhydride. The reaction was repeated two times and then trypsinized. The newly formed N-termini were then derivatized with propionic anhydride twice. The resulting peptides were purified with C18 stage-tip for Mass Spectrometry analysis. Desalted histone peptides (1 μg) were then loaded onto and separated by reversed-phase HPLC on a Thermo Scientific™ EASY-nLC 1000 system with a 75 um i.d. × 15 cm Reprosil-Pur C18-AQ 3 μm nanocolumn run at 300 nL/min. Peptides were eluted with a gradient from 2% to 30% ACN (35 min) and to 98% ACN over 20 min in 0.1% formic acid. The HPLC was coupled to a Thermo Scientific™ Orbitrap Elite™ Hybrid Ion Trap-Orbitrap mass spectrometer. In each cycle, one full MS Orbitrap detection was performed with the scan range of 290 to 1400 m/z, a resolution of 60 K and AGC of 1×e6. Then, data dependent acquisition mode was applied with a dynamic exclusion of 30 s. MS2 scans were followed on parent ions from the most intense ones. Ions with a charge state of one were excluded from MS/MS. An isolation window of 2 m/z was used. Ions were fragmented using collision induced dissociation (CID) with a collision energy of 35%. Iontrap detection was used with normal scan range mode and normal scan rate. The resolution was set to be 15 K with AGC of 1×e4. Targeted scans were performed on a number of peptides to increase the identification of low-abundance modifications. Histone PTM quantification was performed by using in-house developed software EpiProfile[40].

Extended Data

a

Protein	Control	PRDM14-F	CBFA2T2-F
CBFA2T2	**0**	**52**	**462**
PRDM14	**0**	**56**	**102**
CCDC88A	0	7	17
RNF2	0	12	4
KIAA1967	0	0	16
CBFA2T3	0	4	6
MAGED2	0	0	9
MAX	0	5	0
PRKDC	1	0	163
STAT3	2	0	28
ANKFY1	2	12	22
ZBTB2	3	13	33

MS peptide counts



Extended Data Figure 1. Biochemical interaction between PRDM14 and CBFA2T2

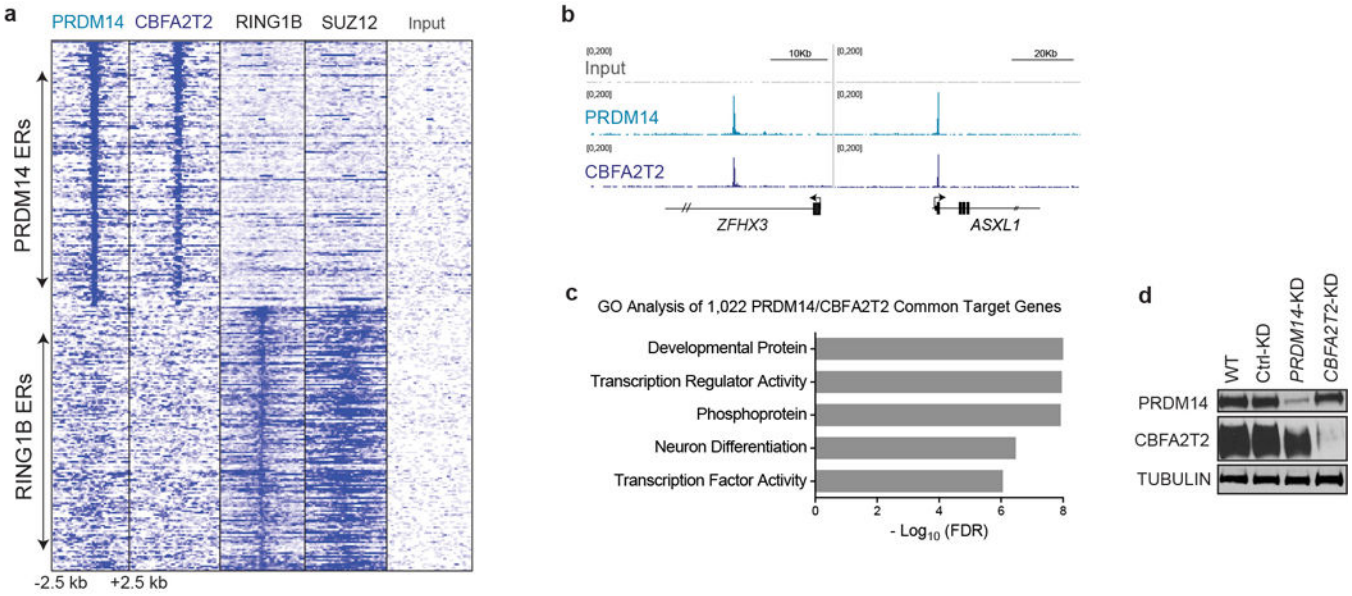
a, Mass spectrometry peptide counts from FLAG affinity purification from NCCIT control cells and stable lines expressing FLAG-HA-PRDM14 (PRDM14-F), and FLAG-HA-CBFA2T2 (CBFA2T2-F).

b, Characterization of in-house human PRDM14 antibody. Western blot performed using 30 μg of NCCIT and KH2 mESC cell lysate. Human PRDM14 antibody is specific and does not cross-react with mouse PRDM14.

c, Immunoprecipitation using antibodies against the indicated endogenous proteins in mESC cells.

d, Western blot of Superose 6 column fractionation of FLAG-purified CBFA2T2 complex in NCCIT cells stably expressing FLAG-HA-CBFA2T2.

e, Chromatin immunoprecipitation analysis using the indicated antibodies in 293T-REx harboring a UAS-TK-Luciferase transgene. Fold enrichment represents the ratio of enrichment by ChIP-qPCR upon induction of GAL4-PRDM14 expression via doxycycline addition. Positions of the primer set are indicated by small arrows in the schematic. qPCR source data are included in SI Source Data Table3.



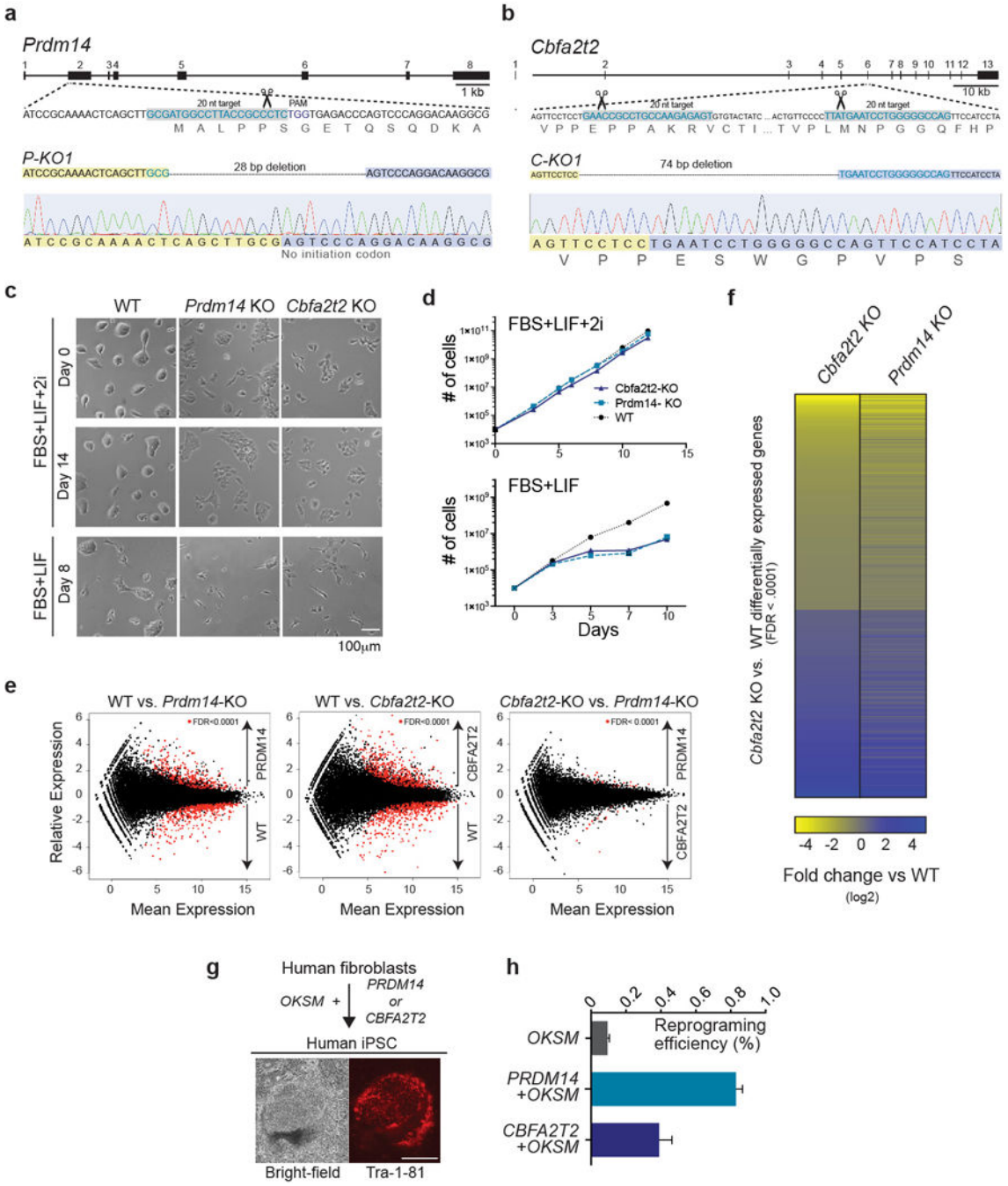
Extended Data Figure 2. PRDM14 and CBFA2T2 exhibit an overlapping and interdependent distribution on chromatin in NCCIT cells

a, Heatmap depicting PRDM14, CBFA2T2, RING1B and SUZ12 read density across a 5kb window centered about the PRDM14 (top) or RING1B (bottom) SERs identified in NCCIT cells.

b, Representative genome browser tracks depicting SERs at the indicated genomic loci.

c, Gene Ontology analysis of PRDM14/CBFA2T2 common target genes.

d, Western blot analysis of PRDM14 and CBFA2T2 protein levels in knockdown experiments (Fig. 1d, 1e).



Extended Data Figure 3. Characterization of KO ESC mutants and quantification of human iPSC reprogramming efficiency

a,b, Strategy for generating *Prdm14* and *Cbfa2t2* knockout mESCs via CRISPR/Cas9 genome-editing. Sequencing chromatograms confirming homozygous disruption of the locus are depicted.

c, *Cbfa2t2* and *Prdm14* KO ESCs require 2i to maintain growth. ESC lines generated under FBS+LIF+2i conditions were continuously cultured in FBS+LIF+2i (top, middle), or switched to FBS+LIF (bottom). Eight days after 2i withdrawal (FBS+LIF), well-formed

ESC colonies were undetectable; instead, mutant ESCs appeared to be differentiated. Scale bar: 100 μm.

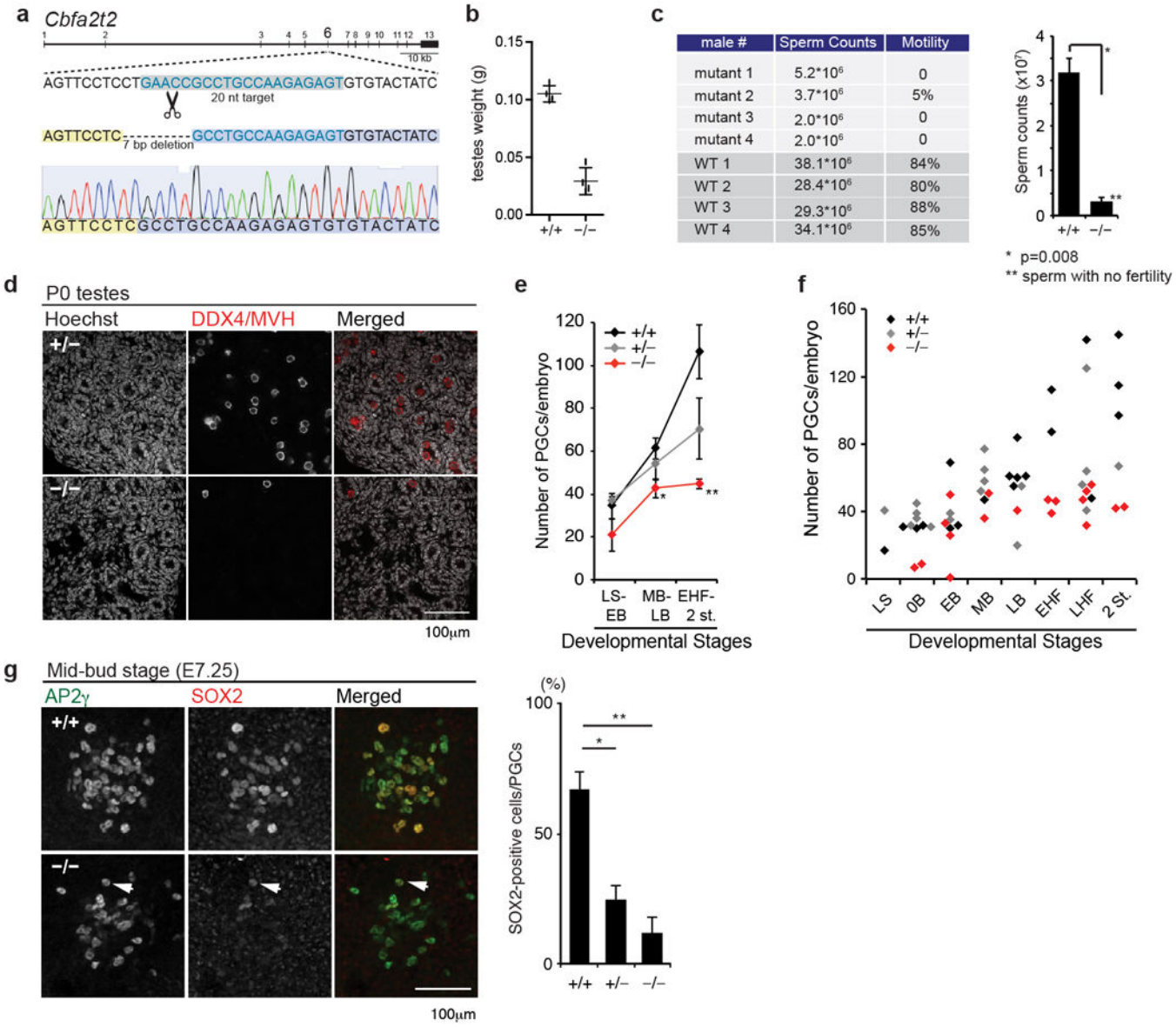
d, Proliferation rates of WT and mutant KO ESCs as sescibed in c. Data were obtained from 3 biological replicates. Please note error bars were shown in the plots. Due to the logarithmic scale used here, some error bars are very small and might be invisible.

e, RNA-seq MA plot in the indicated ESCs. Data is representative of three biological replicate experiments for each line. Mean abundance is plotted on the x-axis and enrichment (both in log 2 scale) is plotted on the y-axis. Genes depicted in red are differentially expressed with a FDR < 0.0001.

f, Heatmap showing relative expression of all differentially expressed genes as sescribed in Fig. 2c. Only difference is now the heatmap is centered on CBFA2T2 differentially expressed genes, rather than PRDM14 DEGs.

g, Scheme of human fibroblast reprogramming to induced pluripotent stem cells. Fibroblasts were transduced with lentiviruses expressing polycistronic *Oct4/Klf4/Sox2/Myc* (OKSM) and either *Prdm14* or *Cbfa2t2*. Three weeks later, bright-field images of successfully reprogrammed colonies (left) and live TRA-1-81 staining (right) were recorded. Scale bar: 500 μm.

h, Quantification of human iPSC reprogramming efficiency based on TRA-1-81 staining with secondary antibody conjugated with peroxidase HRP and substrate DAB. Error bars are based on 4 biological replicates of each condition. The source data is included in SI Source Data Table 4.



Extended Data Figure 4. *Cbfa2t2*$^{-/-}$ mouse genotypes and sperm defects

a, One representative *Cbfa2t2*$^{-/-}$ mouse genotype wherein a 7 bp fragment is deleted.

b, Testes of multiple WT (n=4) and *Cbfa2t2*$^{-/-}$ (n=4) male mice at 8 weeks old were dissected and weighed.

c, Number of sperm in the epididymis of *Cbfa2t2*$^{+/+}$ (n=4) and *Cbfa2t2*$^{-/-}$ (n=4) mice is shown with SEM. *p*-value was determined by student's *t*-test.

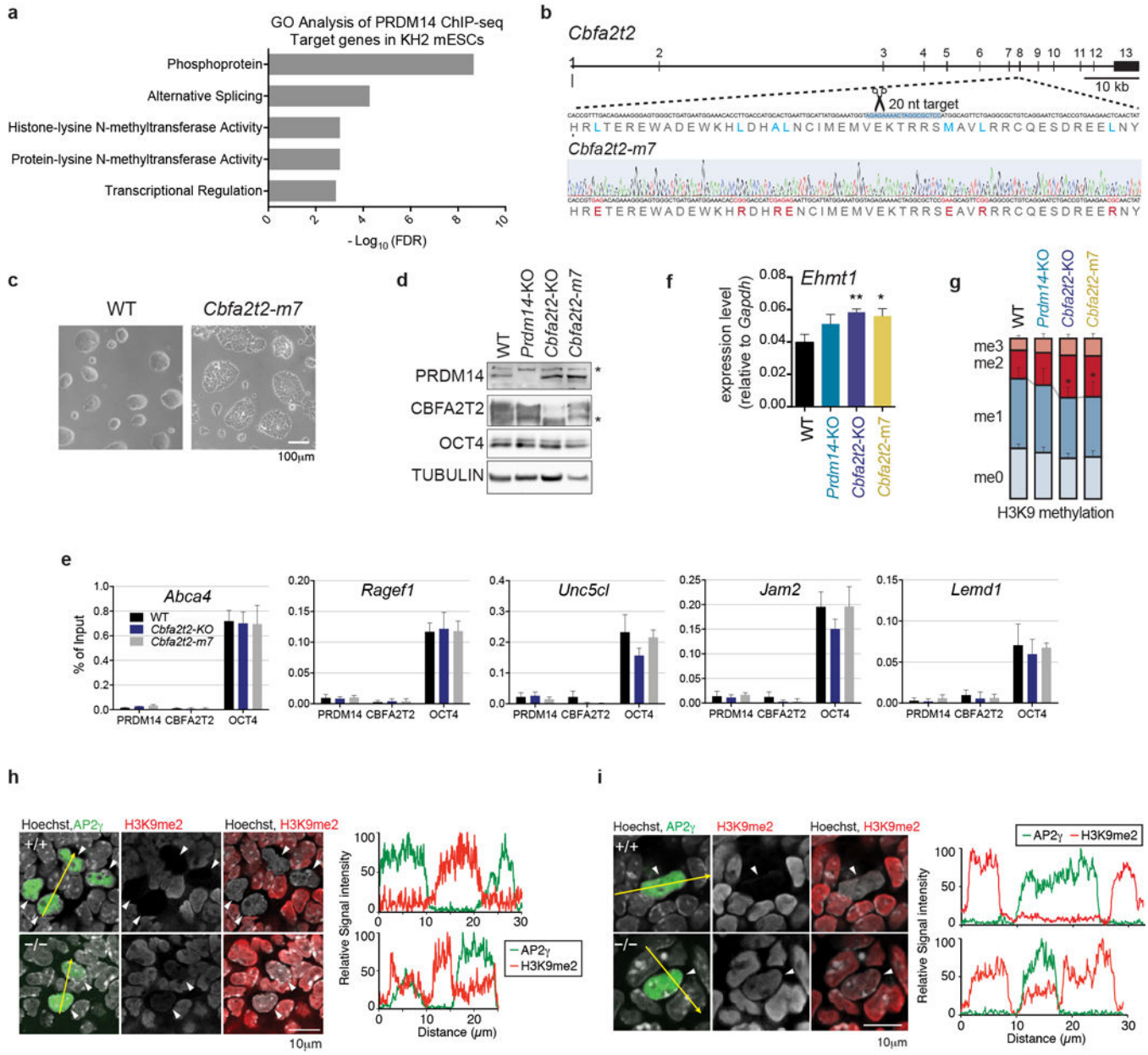
d, Near loss of gonocytes in *Cbfa2t2* KO mutant P0 testes by DDX4 (MVH) staining. Visualization of MVH (red) positive gonocytes in *Cbfa2t2*$^{+/-}$ (upper panels) or *Cbfa2t2*$^{-/-}$ (lower panels) testis at P0 stage. The merged images with Hoechst (left panels; white) are shown on the right panels. Scale bars, 100 mm.

e, Numbers of AP2γ–positive PGCs in *Cbfa2t2*$^{+/+}$ (black), *Cbfa2t2*$^{+/-}$ (gray) and *Cbfa2t2*$^{-/-}$(red) embryos at the indicated embryonic stages. LS, late-streak stage; EB, MB,

and LB, early-, mid-, and late-bud stage; EHF, early-head fold stage; 2 st., 2 somites stage. t-test: *p=0.03, **p=0.003.

f, Numbers of AP2γ–positive PGCs in *Cbfa2t2*$^{+/+}$ (black), *Cbfa2t2*$^{+/-}$ (gray) and *Cbfa2t2*$^{-/-}$(red) embryos at the indicated embryonic stages. LS, late-streak stage; 0B, EB, MB, and LB, zero-, early-, mid-, and late-bud stage; EHF and LHF, early- and late-head fold stage; 2 st., 2 somites stage.

g, Left panel, expression of SOX2 (red) in AP2γ (green)-positive PGCs in *Cbfa2t2*$^{+/+}$ (upper panels) or *Cbfa2t2*$^{-/-}$ (lower panels) embryo at mid-bud (MB) stage, E7.25, shown as z-projection images of posterial confocal sections. Arrow indicates a minor PGC with relatively normal activation of SOX2. Scale bar, 50 mm. Right panel, %SOX2 (red)-positive cells in AP2γ (green)-positive PGCs in the indicated genotypes of *Cbfa2t2* at E7.0-7.25 (zero- to mid-bud stage) are shown with statistical significance (t-test: *p=0.0006, **p=0.0001; *Cbfa2t2*$^{+/+}$, n=7; *Cbfa2t2*$^{+/-}$, n=5; *Cbfa2t2*$^{-/-}$, n=5).



Extended Data Figure 5. *Cbfa2t2* m7 mutant characterization and the related mechanism

a, Gene Ontology analysis of PRDM14 ChIP-seq target genes. PRDM14 target genes are enriched in histone methyltransferase activities by DAVID analysis.

b, *Cbfa2t2* m7 mutant genotyping. The mutant 7 amino acids are in red and corresponding WT residues are highlighted in blue in the displayed protein sequences.

c, Bright-field images of WT and *Cbfa2t2*-m7 mESCs. Scale bar: 100 μm.

d, Western Blot analysis of PRDM14, CBFA2T2, and OCT4 protein levels in *Prdm14* KO, *Cbfa2t2* KO or m7 mutant ESCs under feeder-free FBS+LIF+2i condition. Nonspecific bands are denoted with an asterisk.

e, ChIP-qPCR using antibodies against PRDM14, CBFA2T2, or OCT4 at selected Oct4 target genes. Occupancy is compared between WT, *Cbfa2t2*-KO and *Cbfa2t2*-m7 mESCs. ChIP-qPCR primer sequences are included in Supplementary Table 3.
f, RT-qPCR quantification of *Ehmt1* mRNA levels in WT and mutant lines. P values are 0.004 (**) and 0.0142 (*).The source data is included in SI Source Data Table 5.
g, Mass spectrometry quantification of histone H3K9 modifications in WT and mutant lines. p values are 0.00956, 0.04248 (*). The source data is included in SI Source Data Table 5.
h, i, additional Immunofluorescence analysis of H3K9me2 (red) of AP2γ-positive (green; arrowheads) PGCs in *Cbfa2t2*$^{+/-}$ and $^{-/-}$ embryos at E8.75 as described in Fig 4i.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Drs. Peter Andrews and Xiao-Jian Sun for providing Plasmids. We are grateful to Drs. Lynne Vales and Maria Elena Torres-Padilla, Le Bu for critical comments, Dr. Haiyan Zheng for MS analysis, and Deborah Hernandez, Courtney Leek, Misaki Yamaji, Anika Paradkar, and Mark Alu, for excellent technical assistance. The work was supported by the Howard Hughes Medical Institute and National Institutes of Health (RO1GM064844-12) (D. R.). B.A.G. acknowledges funding from NIH grant R01GM110174. T.T. was supported by HHMI, NIH, Starr Foundation, and Tri-Institutional Stem Cell Initiative. M.Y. was a recipient of JSPS Research Fellowship.

reference

1. Saitou M, Barton SC, Surani MA. A molecular programme for the specification of germ cell fate in mice. Nature. 2002; 418:293–300. [PubMed: 12124616]
2. Surani MA, Hayashi K, Hajkova P. Genetic and epigenetic regulators of pluripotency. Cell. 2007; 128:747–762. [PubMed: 17320511]
3. Yeom YI, et al. Germline regulatory element of Oct-4 specific for the totipotent cycle of embryonal cells. Development. 1996; 122:881–894. [PubMed: 8631266]
4. Yamaji M, et al. Critical function of Prdm14 for the establishment of the germ cell lineage in mice. Nat Genet. 2008; 40:1016–1022. [PubMed: 18622394]
5. Chia NY, et al. A genome-wide RNAi screen reveals determinants of human embryonic stem cell identity. Nature. 2010; 468:316–320. [PubMed: 20953172]
6. Ma Z, Swigut T, Valouev A, Rada-Iglesias A, Wysocka J. Sequence-specific regulator Prdm14 safeguards mouse ESCs from entering extraembryonic endoderm fates. Nat Struct Mol Biol. 2011; 18:120–127. [PubMed: 21183938]
7. Yamaji M, et al. PRDM14 ensures naive pluripotency through dual regulation of signaling and epigenetic pathways in mouse embryonic stem cells. Cell Stem Cell. 2013; 12:368–382. [PubMed: 23333148]
8. Chan YS, et al. A PRC2-dependent repressive role of PRDM14 in human embryonic stem cells and induced pluripotent stem cell reprogramming. Stem Cells. 2013; 31:682–692. [PubMed: 23280602]
9. Payer B, et al. Tsix RNA and the germline factor, PRDM14, link X reactivation and stem cell reprogramming. Mol Cell. 2013; 52:805–818. [PubMed: 24268575]
10. De Braekeleer E, et al. RUNX1 translocations and fusion genes in malignant hemopathies. Future Oncol. 2011; 7:77–91. [PubMed: 21174539]
11. Guastadisegni MC, et al. CBFA2T2 and C20orf112: two novel fusion partners of RUNX1 in acute myeloid leukemia. Leukemia. 2010; 24:1516–1519. [PubMed: 20520637]
12. Miyoshi H, et al. The t(8;21) translocation in acute myeloid leukemia results in production of an AML1-MTG8 fusion transcript. EMBO J. 1993; 12:2715–2721. [PubMed: 8334990]

13. Calabi F, Cilli V. CBFA2T1, a gene rearranged in human leukemia, is a member of a multigene family. Genomics. 1998; 52:332–341. [PubMed: 9790752]
14. Liu Y, et al. The tetramer structure of the Nervy homology two domain, NHR2, is critical for AML1/ETO's activity. Cancer Cell. 2006; 9:249–260. [PubMed: 16616331]
15. Hug BA, Lazar MA. ETO interacting proteins. Oncogene. 2004; 23:4270–4274. [PubMed: 15156183]
16. Sun XJ, et al. A stable transcription factor complex nucleated by oligomeric AML1-ETO controls leukaemogenesis. Nature. 2013; 500:93–97. [PubMed: 23812588]
17. Mikkelsen TS, et al. Dissecting direct reprogramming through integrative genomic analysis. Nature. 2008; 454:49–55. [PubMed: 18509334]
18. Kurimoto K, et al. Complex genome-wide transcription dynamics orchestrated by Blimp1 for the specification of the germ cell lineage in mice. Genes Dev. 2008; 22:1617–1635. [PubMed: 18559478]
19. Beard C, Hochedlinger K, Plath K, Wutz A, Jaenisch R. Efficient method to generate single-copy transgenic mice by site-specific integration in embryonic stem cells. Genesis. 2006; 44:23–28. [PubMed: 16400644]
20. Ran FA, et al. Genome engineering using the CRISPR-Cas9 system. Nat Protoc. 2013; 8:2281–2308. [PubMed: 24157548]
21. Ying QL, et al. The ground state of embryonic stem cell self-renewal. Nature. 2008; 453:519–523. [PubMed: 18497825]
22. Yang H, Wang H, Jaenisch R. Generating genetically modified mice using CRISPR/Cas-mediated genome engineering. Nat Protoc. 2014; 9:1956–1968. [PubMed: 25058643]
23. Amann JM, et al. Mtgr1 is a transcriptional corepressor that is required for maintenance of the secretory cell lineage in the small intestine. Mol Cell Biol. 2005; 25:9576–9585. [PubMed: 16227606]
24. Whyte WA, et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell. 2013; 153:307–319. [PubMed: 23582322]
25. Tachibana M, et al. Histone methyltransferases G9a and GLP form heteromeric complexes and are both crucial for methylation of euchromatin at H3-K9. Genes Dev. 2005; 19:815–826. [PubMed: 15774718]
26. Seki Y, et al. Extensive and orderly reprogramming of genome-wide chromatin modifications associated with specification and early development of germ cells in mice. Dev Biol. 2005; 278:440–458. [PubMed: 15680362]
27. Hajkova P, et al. Chromatin dynamics during epigenetic reprogramming in the mouse germ line. Nature. 2008; 452:877–881. [PubMed: 18354397]

Method References

28. Liew CG, Draper JS, Walsh J, Moore H, Andrews PW. Transient and stable transgene expression in human embryonic stem cells. Stem Cells. 2007; 25:1521–1528. [PubMed: 17379764]
29. Burton A, et al. Single-cell profiling of epigenetic modifiers identifies PRDM14 as an inducer of cell fate in the mammalian embryo. Cell Rep. 2013; 5:687–701. [PubMed: 24183668]
30. Kuzmichev A, Jenuwein T, Tempst P, Reinberg D. Different EZH2-containing complexes target methylation of histone H1 or nucleosomal histone H3. Mol Cell. 2004; 14:183–193. [PubMed: 15099518]
31. Dignam JD, Lebovitz RM, Roeder RG. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. Nucleic Acids Res. 1983; 11:1475–1489. [PubMed: 6828386]
32. Takahashi Y, Rayman JB, Dynlacht BD. Analysis of promoter binding by the E2F and pRB families in vivo: distinct E2F proteins mediate activation and repression. Genes Dev. 2000; 14:804–816. [PubMed: 10766737]
33. Narendra V, et al. Transcription. CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation. Science. 2015; 347:1017–1021. [PubMed: 25722416]

34. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 2009; 4:44–57. [PubMed: 19131956]

35. Wang L, Feng Z, Wang X, Wang X, Zhang X. DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. Bioinformatics. 2010; 26:136–138. [PubMed: 19855105]

36. Sommer CA, et al. Induced pluripotent stem cell generation using a single lentiviral stem cell cassette. Stem Cells. 2009; 27:543–549. [PubMed: 19096035]

37. Mostoslavsky G, Fabian AJ, Rooney S, Alt FW, Mulligan RC. Complete correction of murine Artemis immunodeficiency by lentiviral vector-mediated gene transfer. Proc Natl Acad Sci U S A. 2006; 103:16406–16411. [PubMed: 17062750]

38. Onder TT, et al. Chromatin-modifying enzymes as modulators of reprogramming. Nature. 2012; 483:598–602. [PubMed: 22388813]

39. Lin S, Garcia BA. Examining histone posttranslational modification patterns by high-resolution mass spectrometry. Methods Enzymol. 2012; 512:3–28. [PubMed: 22910200]

40. Yuan ZF, et al. EpiProfile Quantifies Histone Peptides With Modifications by Extracting Retention Time and Intensity in High-resolution Mass Spectra. Mol Cell Proteomics. 2015; 14:1696–1707. [PubMed: 25805797]

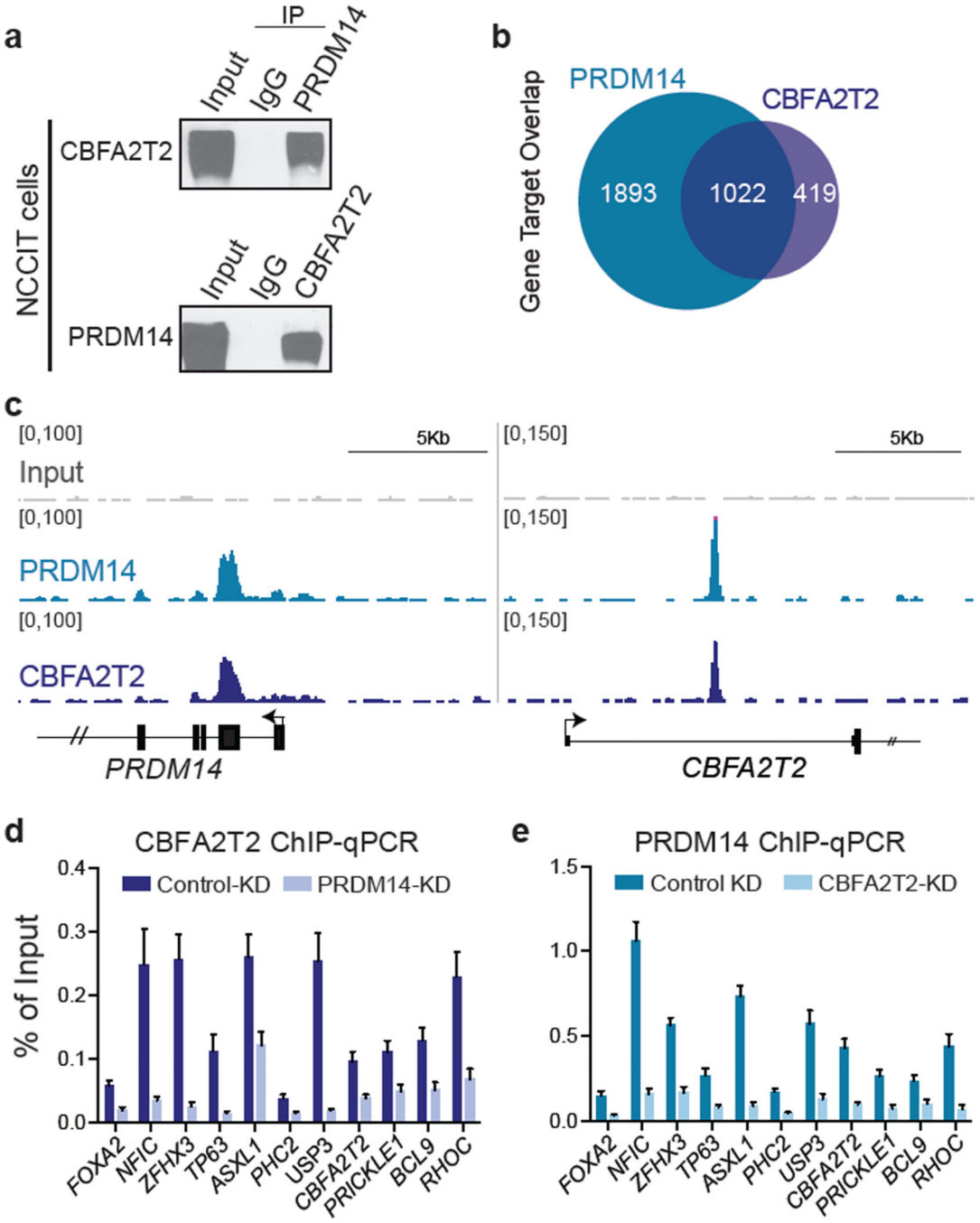


Figure 1. PRDM14 and the co-repressor protein CBFA2T2 interact and bind to chromatin inter-dependently

a, Immunoprecipitation using antibodies against the indicated endogenous proteins in NCCIT cells. For all Western blots, source gel data are included in SI Figure 1.

b, Venn Diagram depicting the overlap of PRDM14 and CBFA2T2 target genes as identified by ChIP-seq.

c, Genome browser tracks showing PRDM14 and CBFA2T2 at their respective genomic loci.

d, e, ChIP-qPCR at SERs found near the 11 indicated genes in NCCIT cells with siRNAs against CBFA2T2 (e) or short hairpin RNAs against PRDM14 (f) (n=3 biological replicates). Error bars, s.d. qPCR source data are included in SI Source Data Table 1.

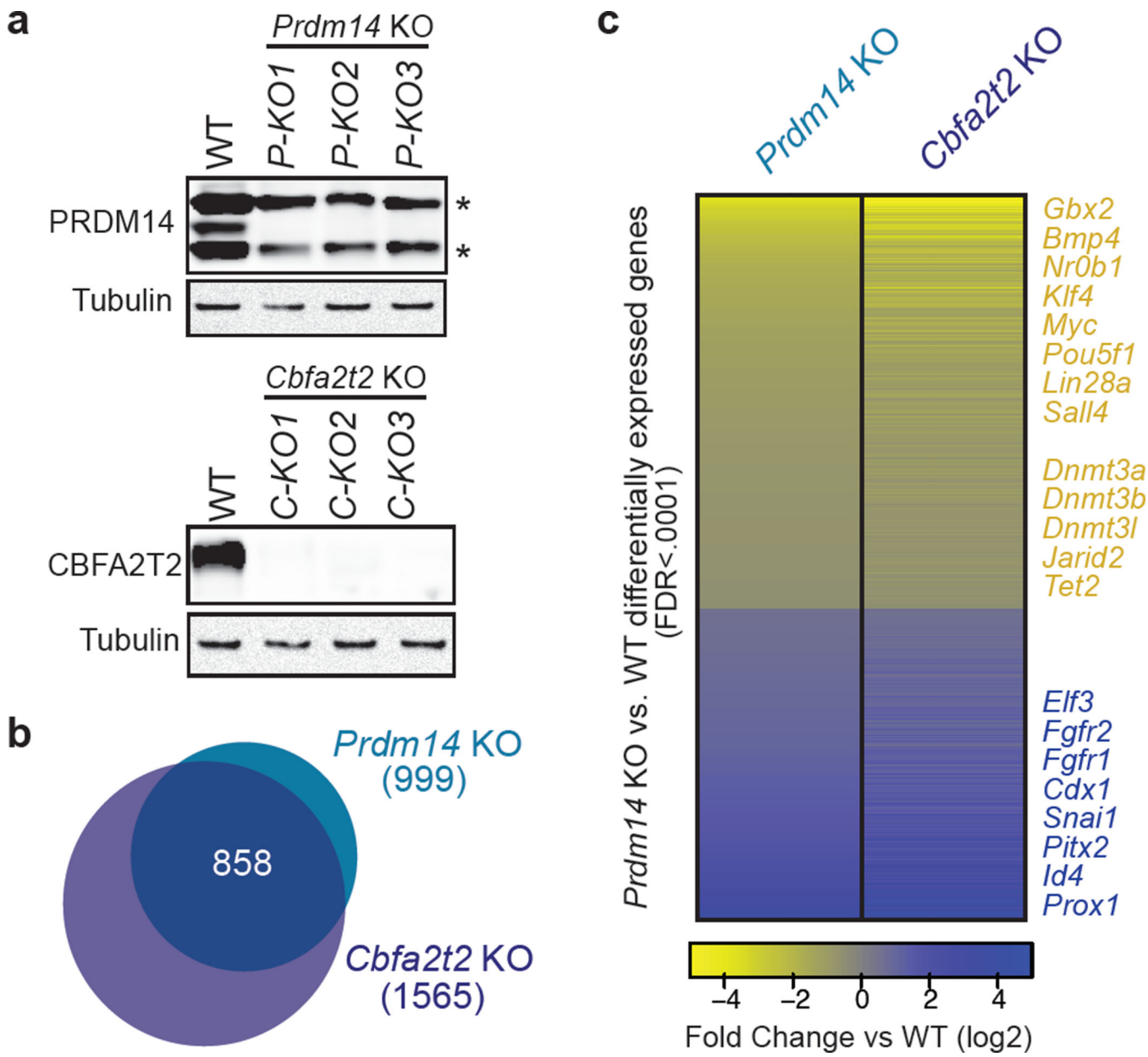


Figure 2. PRDM14 and CBFA2T2 regulate pluripotency

a, Western blots confirming loss of PRDM14 or CBFA2T2 expression in KO mESCs. Nonspecific bands are denoted with an asterisk.

b, Venn diagram depicting the overlap of differentially expressed genes upon deletion of *Prdm14* or *Cbfa2t2* after removal of feeders from FBS+LIF+feeders culture.

c, Heatmap showing relative expression of all differentially expressed genes identified with a false discovery rate less than 1E-3 between WT and either *Prdm14* or *Cbfa2t2* KO mESCs.

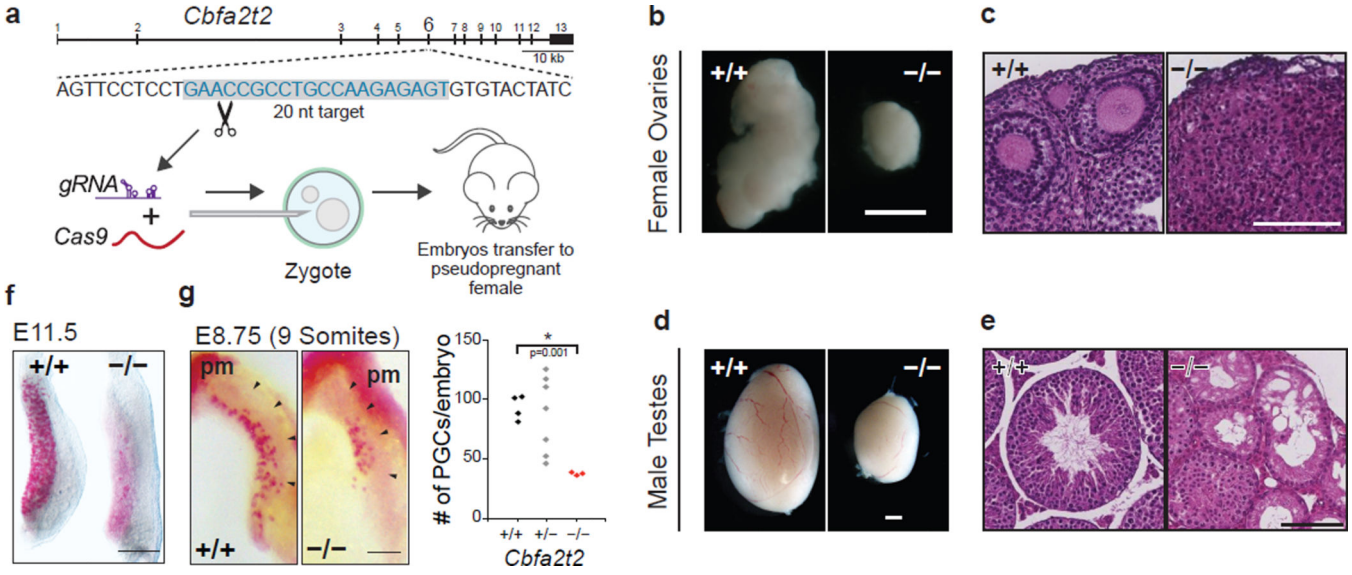


Figure 3. *Cbfa2t2*$^{-/-}$ mice are defective in their germline

a, Schematic of *Cbfa2t2* KO mouse generation by CRISPR zygotic injection.

b and d, Image of dissected ovaries (n=8) and testes (n=4), respectively, *in Cbfa2t2* KO mice. Scale bars, 1 mm.

c and e, Histological sections of ovaries and testes, respectively, stained by H&E. Scale bars, 100 μm.

f. Genital ridges of *Cbfa2t2*$^{+/+}$ and *Cbfa2t2*$^{-/-}$ embryos at E11.5 stained by alkaline phosphatase (AP). Scale bar, 1mm.

g, AP staining of PGCs of E8.75 (9 somites) embryos is shown (n=3). Arrowheads point to the boundary of the developing hindgut. pm, para-axial mesoderm. Scale bar, 100 μm. PGC numbers in each embryo were plotted in the right panel, with the following values: +/+, 93±5; +/−, 87±5; −/−, 38±1.

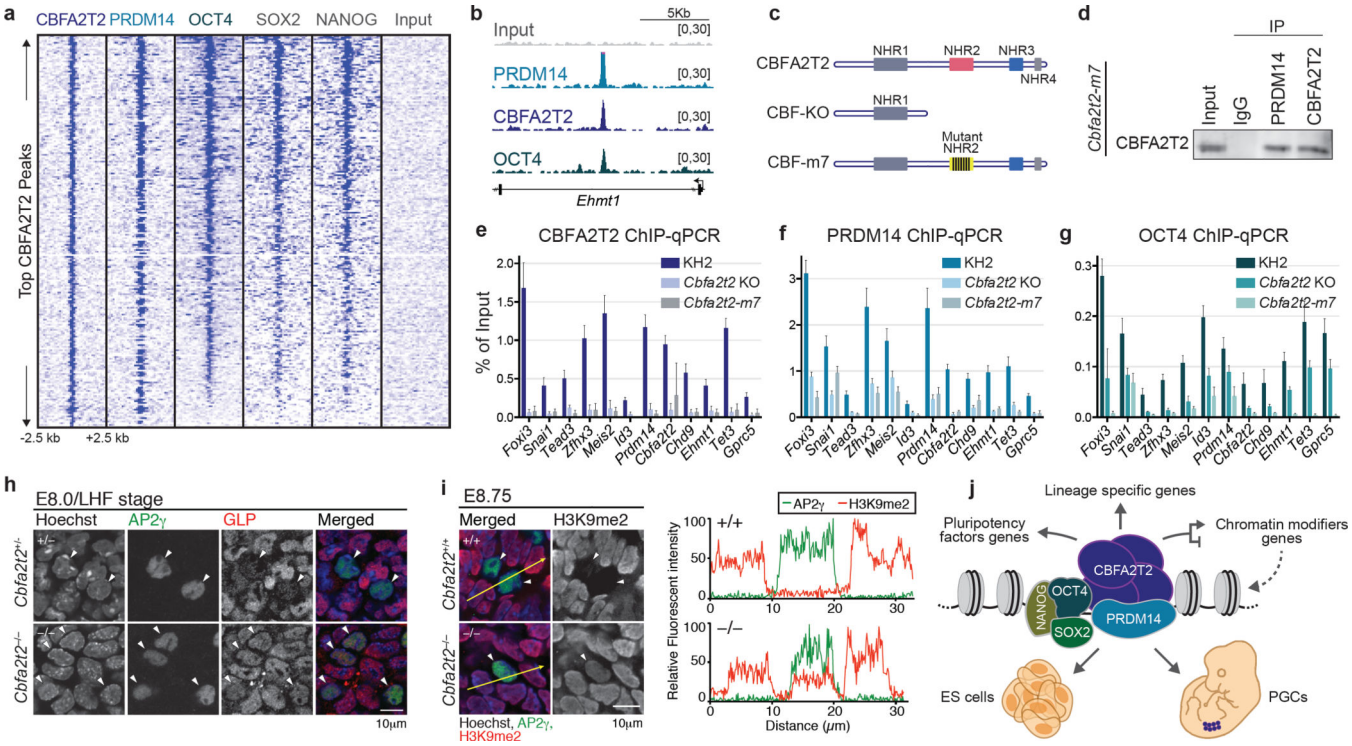


Figure 4. Mechanism of CBFA2T2/PRDM14 complex chromatin binding and direct regulation of PGC epigenetic reprogramming

a, Heatmap depicting CBFA2T2, PRDM14 and OCT4, SOX2, NANOG ChIP-seq read density centered about the top 299 CBFA2T2 SERs in mESCs.

b, Representative genome browser tracks at the indicated *Ehmt1* locus in mESCs.

c, Domain annotation of WT, *Cbfa2t2* KO and oligomerization mutant m7 proteins. The 7 amino acids mutated in *Cbfa2t2*-m7 are depicted as lines within NHR2.

d, Immunoprecipitation against the indicated proteins in *Cbfa2t2-m7* mESCs followed by Western blot.

e, f, g, ChIP-qPCR using antibodies directed against CBFA2T2 (e), PRDM14 (f) or OCT4 (g) at SERs found near the indicated genes (n=3). Error bars, s.d. qPCR source data are included in SI Source Data Table 2.

h, GLP (EHMT1) expression (red) in AP2γ-positive PGCs (green, arrowheads) in *Cbfa2t2*$^{+/-}$ and $^{-/-}$ embryos at E8.0, Late Head-fold (LHF) stage.

i. Immunofluorescence analysis of H3K9me2 (red) of AP2γ-positive (green; arrowheads) PGCs in *Cbfa2t2*$^{+/-}$ and $^{-/-}$ embryos at E8.75. Line plot analysis on yellow arrowed-area are shown on the right. Scale bars, 10 μm. Data are representative of 3 indepenent expeiments.

j. Model depicting the co-repressor CBFA2T2 oligomerizing to stabilize associated TFs (PRDM14 and OCT4) on chromatin.