Co-repressor CBFA2T2 regulates pluripotency and germline development

Shengjiang Tu1,2, Varun Narendra1,2, Masashi Yamaji3, Simon E. Vidal3, Luis Alejandro Rojas1,2, Xiaoshi Wang5, Sang Yong Kim6, Benjamin A. Garcia5, Thomas Tuschi13, Matthias Stadtfeld4 and Danny Reinberg1,2

Developmental specification of germ cells lies at the heart 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)1–3, 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 in common with embryonic stem (ES) cells, including OCT4 (encoded by Pou5f1), SOX2, NANOG and PRDM14 (refs 2–4). A biochemical mechanism by which these transcription factors converge on chromatin to produce the dramatic rearrangements underlying ES-cell- and PGC-specific transcriptional programs remains poorly understood. Here we identify a novel co-repressor protein, CBFA2T2, that regulates pluripotency and germline specification in mice. 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 transcription factors at the crossroads of the fundamental developmental plasticity between uni- and pluripotency.

The germ line first segregates from somatic lineages via the specification of PGCs between embryonic day (E)6.25–7.25 in mice1–2. PRDM14 regulates pluripotency5–7, and is the only known transcription factor to specifically regulate germ cell specification4. To understand better the mechanism(s) underlying PGC development, we sought PRDM14-interacting proteins in the human germ-cell tumour cell line NCCIT. NCCIT cells stably expressing Flag–PRDM14 were subjected to affinity purification and proteomic analysis. In contrast with previous reports8–9, 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–haemagglutinin (HA)-tagged CBFA2T2 confirmed strong interaction with PRDM14 (Extended Data Fig. 1a). CBFA2T2, CBFA2T3 and ETO (also known as RUNX5/1T1) comprise a homologous gene family frequently targeted for translocation events in acute myeloid leukaemia10–13. Despite 85% sequence similarity among homologues, their ubiquitous expression and capacity to form heterotetramers14–16, 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 induced pluripotent stem (iPS) cell reprogramming17 and PGC specification18.

Further reciprocal immunoprecipitations confirmed endogenous PRDM14 and CBFA2T2 interaction in both NCCIT cells and mouse ES cells (Fig. 1a and Extended Data Fig. 1b, c). Gel filtration of the Flag eluate gave evidence of a larger than 600 kDa complex (Extended Data Fig. 1d), possibly due to CBFA2T2 oligomerization14. Moreover, a

---

1Howard Hughes Medical Institute, New York University School of Medicine, New York, New York 10016, USA. 2Department of Biochemistry and Molecular Pharmacology, New York University School of Medicine, New York, New York 10016, USA. 3Department of Biochemistry and Molecular Pharmacology, New York University School of Medicine, New York, New York 10016, USA. 4Epigenetics Program, Department of Biochemistry and Biophysics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. 5Epigenetics Program, Department of Biochemistry and Biophysics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. 6Rodent Genetic Engineering Core, NYU School of Medicine, New York, New York 10016, USA.

---

© 2016 Macmillan Publishers Limited. All rights reserved
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 colocalization on chromatin genome wide, we performed chromatin immunoprecipitation followed by sequencing (ChIP-seq) in wild-type NCCIT cells. In the case of CBFA2T2, 2,077 statistically enriched regions (SERs) were identified using a stringent P-value threshold of $1 \times 10^{-15}$, of which 1,384 overlapped with a PRDM14-binding event (Extended Data Fig. 2a and Supplementary Table 1). Global mapping of SERs to their nearest promoters identified 1,022 PRDM14/CBFA2T2 co-targeted genes (Fig. 1b), many of which are transcription factors involved in lineage commitment (Extended Data Fig. 2b, c). By 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 short hairpin RNAs (shRNAs) or short interfering RNAs (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 naive pluripotency in mouse embryonic stem (mES) cells. To examine such a role for CBFA2T2, we generated Cbfa2t2- and Prdm14-knockout cells in KH2 mES cells using CRISPR–Cas9 genome editing. Guide RNAs (gRNAs) targeting the sixth exon (common to all Cbfa2t2 isoforms) or the second exon of Prdm14, produced multiple lines harbouring distinct frameshift mutations and loss of the targeted protein (Fig. 2a and Extended Data Fig. 3a, b). Colonies of Prdm14- and Cbfa2t2-knockout mES cells 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) (Extended Data Fig. 3d), as shown in the case of Prdm14-knockout lines. After exposure to 2i-free conditions, three different knockout lines for both Prdm14 and Cbfa2t2, alongside wild-type cells, were subjected to RNA sequencing (RNA-seq) analyses. Eighty-five per cent of genes differentially expressed in a Prdm14-knockout setting were also dysregulated upon loss of Cbfa2t2 expression (Fig. 2b, Extended Data Fig. 3e and Supplementary Table 2). Moreover, the directionality of differential gene expression was nearly identical across mutants (Fig. 2c and Extended Data Fig. 3f). In both knockout ES cells, numerous pluripotency genes, including Klf4, Pou5f1, Nr0b1 (also known as Dax1), Lin28a and Myc, were downregulated, whereas lineage commitment genes such as Elf5, Cdx1 and Pitx2 were upregulated. Similar to the case with PRDM14 (ref. 5), Cbfa2t2 overexpression enhanced iPS cell reprogramming efficiency (Extended Data Fig. 3g, h). Thus, the Cbfa2t2 co-repressor contributes positively to pluripotency.

Given that Prdm14−/− mice displayed a major defect in germline development, we tested the contribution of Cbfa2t2 to both somatic and germline development by generating Cbfa2t2-knockout mice via CRISPR zygotic injection. C57BL/6 zygotes were co-injected with Cas9 messenger RNA and one of the gRNAs used in mES cells to target exon 6 of Cbfa2t2 (Fig. 3a). We obtained multiple pups possessing a germline mutation that caused a frameshift mutation and a dysfunctional truncated protein (Extended Data Fig. 4a). Genetic targeting was specific, as the ten most likely off-target genomic regions were unperturbed (Supplementary Table 3). Intercrossing of Cbfa2t2−/− mice produced pups in a roughly normal Mendelian ratio (25% +/+; 50%; 58% +/− (70%; 17% −/− (21)) and Cbfa2t2−/− animals appeared normal. However, crosses of those mice (two female and three male) with wild-type C57BL/6 counterparts failed to produce pups over 2 months.

To pinpoint the germline defect underlying the fertility phenotype of Cbfa2t2−/− mice, we analysed anatomical and histological phenotypes of the reproductive organs. Female Cbfa2t2−/− adult mice have underdeveloped ovaries (Fig. 3b), exhibiting a total absence of follicles (Fig. 3c). Similarly, testes of male Cbfa2t2−/− mice were reduced to ~30% of wild type (Fig. 3d and Extended Data Fig. 4b). Total number of sperm was reduced to less than 10% of wild type, while remaining sperm were largely immotile (Extended Data Fig. 4c) and unable to bind the zona pellucida 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, postnatal day 0 (P0) male Cbfa2t2−/− testes were almost completely devoid of gonocytes (Extended Data Fig. 4d). These data contrast with a previous study claiming that Cbfa2t2−/− mice are fertile. 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 staining of the genital ridge of E11.5 Cbfa2t2−/− embryos showed greater than 95% reduction in the number of PGCs relative to wild type (Fig. 3f). This defect occurs even earlier, at E7.25–8.75 (Fig. 3g and Extended Data Fig. 4e, f). 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 mES cells, relative to that of OCT4, SOX2 and NANOG (OSN) from published ChIP-seq data. CBFA2T2 and PRDM14 colocalize broadly across the genome in mES cells (Fig. 4a), and also exhibit...
sections of ovaries and testes, respectively, stained by H&E. Scale bars, respectively, 100 μm. Genital ridges of Cbfa2t2−/− and Cbfa2t2+/+ embryos at E11.5 stained by alkaline phosphatase. Scale bar, 1 mm. g. Alkaline phosphatase 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.

The nervy homology 2 domain (NHR2) of ETO—a CBFA2T2 homologue—is required for self-renewal of haematopoietic stem cells in leukaemia14,16. 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 by forming a four-helix bundle tetrameric structure. To test whether CBFA2T2 oligomerization contributes

Figure 3 | Cbfa2t2−/− mice are defective in their germ line. a, Schematic of Cbfa2t2-knockout mouse generation by CRISPR zygotic injection. nt, nucleotide. b, d, Image of dissected ovaries (n = 8) and testes (n = 4), respectively, in Cbfa2t2-knockout mice. Scale bars, 1 mm. c, 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 considered overlap with OSN, as reported for PRDM14 in human ES cells5. As in the case of NCCIT cells, CBFA2T2–PRDM14 target genes include numerous lineage-commitment transcription factors and chromatin regulators (Extended Data Fig. 5a and Supplementary Table 4), many of which are co-occupied by OCT4, including the histone H3K9 methyltransferase gene Ehmt1 (Fig. 4b)22.

Figure 4 | Mechanism of CBFA2T2–PRDM14 complex chromatin binding and direct regulation of PGC epigenetic reprogramming. a. Heat map depicting CBFA2T2, PRDM14, OCT4, SOX2 and NANOG ChIP-seq read density centred about the top 299 CBFA2T2 SERS in mES cells. b. Representative genome browser tracks at the indicated Ehmt1 locus in mES cells. c, Domain annotation of wild-type, Cbfa2t2-knockout (CBF-KO) and oligomerization-mutant m7 proteins. The 7 amino acids mutated in Cbfa2t2-knockout (CBF-m7) are depicted as lines within NHR2. d, Immunoprecipitation (IP) against the indicated proteins in Cbfa2t2-m7 mES cells followed by western blot. e–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 the Supplementary Information. h, EHMT1 expression (red) in AP-2γ-positive PGCs (green, arrowheads) in Cbfa2t2−/− and Cbfa2t2+/− embryos at E8.0, late head-fold (LHF) stage. i, Immunofluorescence analysis of H3K9me2 (red) of AP-2γ-positive (green; arrowheads) PGCs in Cbfa2t2−/− and Cbfa2t2+/− embryos at E8.75. Line plot analysis on yellow-arrowed area are shown on the right. Scale bars, 10 μm. Data are representative of three independent experiments. j, Model depicting the co-repressor CBFA2T2 oligomerization to stabilize associated transcription factors (PRDM14 and OCT4) on chromatin.
to ES cell pluripotency, mES cells harbouring the m7 mutation in CBF2AT2 were generated using CRISPR-Cas9 technology (Fig. 4c and Extended Data Fig. 5b). Similar to Cbf2at2−/− cells, Cbf2at2-m7 cells exhibited a flattened morphology (Extended Data Fig. 5c) and a total abrogation of CBF2AT2 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), CBF2AT2 oligomerization was required to stabilize PRDM14 and OCT4 on chromatin. ChiP with quantitative polymerase chain reaction (ChiP-qPCR) showed a significant reduction in PRDM14 and OCT4 occupancy across 12/12 target genes tested (Fig. 4g, i). Importantly, PRDM14–CBF2AT2-independent OCT4 targets retained OCT4 binding (Extended Data Fig. 5e). Thus, CBF2AT2 oligomerization is a critical molecular event underpinning a pluripotent network, providing a scaffolding function to stabilize essential transcription factors such as PRDM14 and OCT4 at their target sites.

CBF2AT2–PRDM14 targets comprise numerous components of the chromatin modifying machinery, such as EHMT1 (also known as GLP) (Fig. 4h, Extended Data Fig. 5a and Supplementary Table 4). During PGC development, H3K9me2 levels are reduced26, potentially due to repression of the H3K9 methyltransferase EHMT1 via a presently unknown mechanism27. Here, knockout of Prdm14 or Cbf2at2 in mES cells caused derepression of Ehmt1 (Extended Data Fig. 5f). Quantitative analysis showed a specific increase in H3K9me2 and H3K9me3 levels in Prdm14−/−, Cbf2at2−/− and Cbf2at2-m7 mutant mES cells (Extended Data Fig. 5g). Importantly, CBF2AT2–PRDM14-mediated repression was required to maintain appropriate levels of H3K9me2 in PGCs in vivo. PGCs in E8.0 Cbf2at2−/− embryos exhibited a specific EHMT1 derepression (Fig. 4h), with resultant increased H3K9me2 levels at E8.75 (Fig. 4i and Extended Data Fig. 5h, i). Thus, direct control of global levels of chromatin modifications is probably another mechanism by which PRDM14 and CBF2AT2 regulate the delicate balance between self-renewal and lineage specification (Fig. 4j).

In summary, CBF2AT2, a co-repressor protein, is a novel factor regulating pluripotency and is essential for germine development. In contrast to the long-held notion that co-repressors have a passive role in transcription factor recruitment, CBF2AT2, 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 for stable transcription factor binding (Fig. 4j). Such a model may extend to numerous transcription factors tested, including CBFA2T2, whose binding to specific target sites.

Note added in proof: During the revision of this manuscript, another study utilized an in vitro differentiation system to determine the involvement of CBF2AT2 in PGC formation, in accordance with our in vivo findings28.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 11 August 2015; accepted 13 April 2016.

Published online 8 June 2016.

1. Saitou, M., Barton, S. C. & Surani, M. A. A molecular programme for the specification of germ cell fate in mice. *Nature* **418**, 293–300 (2002).
2. Surani, M. A., Hayashi, K. & Hajkova, P. Genetic and epigenetic regulators of pluripotency. *Cell* **128**, 747–762 (2007).
3. Yeom, Y. I. et al. Germline regulatory element of Oct-4 specific for the totipotent cycle of embryonal cells. *Development* **122**, 881–894 (1996).
4. Yamaji, M. et al. Critical function of Prdm14 for the establishment of the germ cell lineage in mice. *Nature Genet.* **40**, 1016–1022 (2008).
5. Chia, N. Y. et al. A genome-wide RNA screen reveals determinants of human embryonic stem cell identity. *Nature* **468**, 316–320 (2010).
6. Ma, Z., Swigut, T., Valouev, A., Rada-Iglesias, A. & Wysocka, J. Sequence-specific regulator Prdm14 safeguards mouse ESCs from entering extramammary endoderm fates. *Nature Struct. Mol. Biol.* **18**, 120–127 (2011).
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* **12**, 368–382 (2013).
8. Chan, Y. S. et al. A PRC2-dependent repressive role of PRDM14 in human embryonic stem cells and induced pluripotent stem cell reprogramming. *Stem Cells** **31**, 682–693 (2013).
9. Payer, S. et al. Tbx RNA and the germline factor PRDM14, link X reactivation and stem cell reprogramming. *Mol. Cell* **52**, 805–818 (2013).
10. De Braekeleer, E. et al. RUNX1 translocations and fusion genes in malignant hemopathies. *Future Oncol.* **7**, 77–91 (2011).
11. Guastadisegni, M. C. et al. CBF2AT2 and C20orf112: two novel fusion partners of RUNX1 in acute myeloid leukemia. *Leukemia** **24**, 1516–1519 (2010).
12. Hijikoshi, H. et al. The t(8;21) translocation in acute myeloid leukemia results in production of an AML1-MTG8 fusion transcript. *EMBO J.* **12**, 2715–2721 (1993).
13. Calabi, F. & Cilli, V. CBF2AT1, a gene rearranged in human leukemia, is a member of a multigene family. *Genomics* **52**, 332–341 (1998).
14. Liu, Y. et al. The tetramer structure of the Nervy homology domain, NHR2, is critical for AML1/ETO’s activity. *Cancer Cell* **9**, 249–260 (2006).
15. Hu, B. A. & Lazar, M. A. ETO interacting proteins. *OncoGene** **23**, 4270–4274 (2004).
16. Sun, X. J. et al. A stable transcription factor complex nucleated by oligomeric AML1-ETO controls leukaemogenesis. *Nature* **500**, 93–97 (2013).
17. Mikkelsen, T. S. et al. Dissecting direct reprogramming through integrative genomic analysis. *Nature* **454**, 49–55 (2008).
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.* **22**, 1617–1635 (2008).
19. Beard, C., Hochmedinger, K., Plath, K., Wutz, A. & Jaenisch, R. Efficient method to generate single-copy transgenic mice by site-specific integration in embryonic stem cells. *Genesis* **44**, 23–28 (2006).
20. Ran, F. A. et al. Genome engineering using the CRISPR/Cas9 system. *Nature Protocols** **8**, 2281–2308 (2013).
21. Ying, Q. L. et al. The ground state of embryonic stem cell self-renewal. *Nature* **453**, 519–523 (2008).
22. Yang, H., Wang, H. & Jaenisch, R. Generating genetically modified mice using CRISPR/Cas-mediated genome engineering. *Nature Protocols** **9**, 1956–1968 (2014).
23. Amann, J. M. et al. Mtgr1 is a transcriptional corepressor that is required for maintenance of the secretory cell lineage in the small intestine. *Mol. Cell. Biol.* **25**, 9576–9585 (2005).
24. Whyte, W. A. et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307–319 (2013).
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.* **19**, 619–626 (2005).
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.* **278**, 440–458 (2005).
27. Hajkova, P. et al. Chromatin dynamics during epigenetic reprogramming in the mouse germ line. *Nature* **452**, 877–881 (2008).
28. Nady, N. et al. ETO family protein Mtgr1 mediates Prdm14 functions in stem cell maintenance and primordial germ cell formation. *Elife* **4**, e10150 (2015).

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

Acknowledgements We thank P. Andrews and X.-J. Sun for providing plasmids. We are grateful to L. Vales, M. E. Torres-Padilla and L. Bu for critical comments, H. Zhang for MS analysis, and D. Hernandez, C. Leek, M. Yamaji, A. Paradkar and M. Alu for excellent technical assistance. The work was supported by the Howard Hughes Medical Institute (HHMI) and National Institutes of Health (NIH; R01GM064844-12) (D.R.). B.A.G. acknowledges funding from NIH grant R01GM10174. T.T. was supported by the HHMI, NIH, Starr Foundation, and Tri-Institutional Stem Cell Initiative. M.Y. was a recipient of a Japan Society for the Promotion of Science (JSPS) Research Fellowship.

Author Contributions S.T. designed and performed majority of experiments; V.N. analysed ChiP-seq and RNA-seq data; S.E.V. and D.S. performed iPS cell reprogramming experiments; X.W. and B.A.G. quantified histone modifications. S.Y.K. did CRISPR/Cas-mediated genome engineering, *Nature Protocols* **9**, 1956–1968 (2014).

Author Information Sequencing data have been deposited in the Gene Expression Omnibus under accession number GSE71676. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.R. (danny.reinberg@nyumc.org).

© 2016 Macmillan Publishers Limited. All rights reserved
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, t-glutamine, penicillin/streptomycin and sodium pyruvate. 293T cells (ThermoFisher, #R710-07) were grown in DMEM with 10% FBS, t-glutamine and penicillin/streptomycin. KH2 ES cells, described previously40, were grown in DMEM supplemented with 15% FBS, t-glutamine, penicillin/streptomycin, non-essential amino acids, 0.1 mM β-mercaptoethanol. Human iPS cell culture medium contains advanced DMEM/F12 plus 20% Knockout Serum Replacement, t-glutamine, penicillin/streptomycin, non-essential amino acids, 0.1 mM β-mercaptoethanol. Human fibroblast B cells were obtained from ATCC and maintained on fibroblast medium: DMEM knockout media with 10% FBS sera, 1% non-essential amino acids, t-glutamine, penicillin/streptomycin, non-essential amino acids and 0.1 mM β-mercaptoethanol. Human fibroblast FACS sorted and seeded at 20,000 cells per 15-cm plate. Seven days later, ES-cell nuclei were PCR amplified and purified. Mouse KH2 ES cells were transfected with the knock-in, a donor 723 bp gBlock DNA centred at the NHR2 domain sequence targeting the NHR2 domain was chosen for cloning into the Cas9 vector. For this Cbfa2t2 vector (Addgene, px458) 20. For -knockout KH2 lines, gRNAs were cloned into pSpCas9 (BB)-2A-eGFP or N-terminal GAL4, C-terminal HA fusion PRDM14 construct. 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. ChIP was done as described with biological replicates41. Briefly, cells were crosslinked with 1% formaldehyde, lysed, and sonicated in buffer 3 (10 mM Tris, pH 7.9, 1 mM EDTA, 0.5% N-lauroylsarcosine) down to a desired chromatin size. Forty microlitres of 3:1 mixture of protein A and protein G Sepharose beads were blocked with 0.1 mg ml−1 salmon sperm DNA and 1 mg ml−1 bovine serum albumin (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, 1 mM EDTA, 0.1% Triton X-100, 0.1% sodium deoxycholate (DOC) and protease inhibitors). After 4 h incubation with the remaining protein A/G beads, samples were washed six 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 T2E50S1 (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 h 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 Master mix (Roche), ROX reference dye, 3 μl water, 1 μl 5′/3′ primer pair, and 1 μl DNA were mixed for PCR amplification. For GAL4 ChIP-qPCR, the primer sequences were: GAL4ChIPLucP5R: GCTTCTGCCAACCGAACGGAC. Other qPCR primers are listed in Supplementary Table 3.

ChIP-seq. Mouse KH2 cells and knockout lines were grown on feeders for three 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 unattached ES cells were spun down and lysed with TRIzol. Standard RNA-seq procedure was used44.

Library construction. Libraries for ChIP-seq were prepared according to manufacturer’s instructions (Illumina). Briefly, immunoprecipitated DNA (~5 ng) was end-repaired using End-It Repair Kit (Epiconet), tailed with deoxyadenylate using Klenow exo-- (New England Biolabs), and ligated to custom adaptors with T4 Rapid DNA Ligase (Enzymatics). Fragments of 200–400 bp were size-selected using agarose gels, and subjected to PCR amplification using Q5 DNA Polymerase (NEB). The PCR products were purified and end-repaired using Blunt/Adaptor II Kit (Epicenter) and subjected to PCR amplification using Klenow exo-- (New England Biolabs). Libraries were quantified by qPCR using primers annealing to the adaptor sequence and sequenced at a concentration of 12 pM on an Illumina HiSeq. Barcodes were used for multiplexing. For RNA-seq libraries, polya+ RNA was isolated using Dynabeads Oligo(dT)25 (Invitrogen) and

Analysis. One hundred and twenty minute gradients (6–7.5% acetonitrile) were used (nanofLC1000, 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: TTCCTGATGTGTCATACGAG; human PRDM14 shRNA 2: AACATGAAGAATGTGGATCCG; human PRDM14 shRNA 3: TGGTCCAGGAGTTCTTACCGAG. Lentiviruses from these 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), 1.1 μg spPAZ2, and 1.1 μg PMD2.G 2nd generation packaging plasmids (Addgene, #12260, #12259) were transfected with Lipofectamine 2000. Forty eight hours and sixty hours post-transfection, supernatants were collected. Viral particles were filtered through 0.45 μm filters and enriched 100-fold by centrifugation at 20,000 r.p.m. for 1.5 h. For transduction of NCCIT cells, lentivirus, 0.20 million cells per well were seeded in 6-well plates. The next day, 20 μl virus was transduced with polyacrylamine at a final concentration of 8 μg ml−1. Forty eight hours post-transduction, 1 μg ml−1 Puromycin was added to select for transduced NCCIT cells. Transduced cells were expanded and harvested in 1 week for ChIP-qPCR.

For Cbfa2t2 knockdown experiments, the following On-TARGETplus siRNA sequences from Thermo Scientific were used: GAUACUGCUUUGACGAAAG; CAGAAUUCUCACGAAAAGUA; UAGAGGAAUUGCAACUG; CCACAGGAGAUCCAGAGAG.

© 2016 Macmillan Publishers Limited. All rights reserved
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 earlier 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 (~200 nt). Enriched regions (ERs) were identified using ChIPseek with default parameters and an input control, and then filtered for ERs with at least 10 tags and an unadjusted \(P\) value of \(<1 \times 10^{-15}\) for ChIPs performed in NCCIT cells, and \(<1 \times 10^{-5}\) for ChIPs performed in HK2 mES cells. 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-centred windows of 5–10 kb. Normalized cumulative read density across these bins is depicted. All Gene Ontology analysis was performed using the DAVID tool.

RNA-seq reads were assigned to genes using DESeq (R package) and the ENSEMBL annotation. FDR-adjusted \(P\) values for differential gene expression were calculated with DESeq (R package). Genes were considered to be ‘differentially expressed’ if their adjusted \(P\) value was \(<0.0001\).

Human iPS cell reprogramming. PRDM14 and CBAF2T2 lentiviral constructs were cloned into pHAGE-EF1α-iRES tdTomato and pHAGE-EF1α-iRES zGreen vectors, respectively. OKSM polyclinosteric vector under EF1α control was used for the reprogramming experiments. Lentiviruses were produced as described previously. Supernatants were collected every 12 h on two consecutive days starting 48 h after transfection. Viral particles were concentrated by centrifugation at 20,000 r.p.m. for 1 h. Virus titre was quantified by flow cytometry and immunofluorescence, and high-titre viruses (greater than 60% transduction efficiency) were chosen for further human iPSC cell reprogramming experiments. Human fibroblast BJ cells were seeded at 0.14 million cells per well of a 96-well plate. Twenty-eight hours after seeding, 10–15 μl of concentrated OKSM, OKSM plus PRDM14, or OKSM plus CBAF2T2 virus combinations were used to transduce human fibroblasts with polybrene (final concentration of 8 μg ml\(^{-1}\)). Forty hours post-transduction, tdTomato and GFP-positive cells were sorted and seeded onto irradiated mouse embryonic fibroblast (MEF) feeders (GLOBESTEM). Cells undergoing reprogramming were maintained on human fibroblast medium for the first week, and then 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 (Podoclycin) Biotin solution (1:20 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 HRP was used with substrate DAB (Vector Labs, #SK-4100). Error bars are based on three biological replicates of each condition.

Cbfα2t2 knockout mice. We generated Cbfα2t2 knockout mice via zygotic injection. T7-gRNA DNA template was PCR amplified from one of the Cas9-gRNA plasmids for generation of Cbfα2t2 knockout ES cells. The gRNA sequence is ACTCCTCCTGCGAGCTGCCT. The primer sequences are TTAATACGACTCACTATAGGGATATGCAATAGCACCAGC and GGAGGCTCTGAAATTCCTAC. 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\(^{-1}\) Cas9 mRNA, 50 ng μl\(^{-1}\) gRNA. Cytoplasmic injection was performed on 102 C57BL/6Ezoytes. Of those, 72 embryos were transferred to three pseudopregnant female mice. A total of 30 pups were born and genotyped. Genotyping primers were TAGAGCTCCTCCTGTGTTG and CTTGGTGGTCTGTAGACTT. Ten potential off-target sites were tested by PCR for each of the ten primer pair sequences are listed in Supplementary Table 3. Crossing CRISPR mutant mice containing one allele of indel mutation with wild-type C57BL/6 mice resulted in Cbfα2t2/− mice. Intercrossing of these mice produced full Cbfα2t2 knockout mice (Cbfα2t2/−). Mouse studies were approved by the New York University Medical Center Institutional Animal Care and Use Committee.

Sperm counts and motility analysis and in vitro fertilization. Individual caudal epididymis was minced in 90 μl MBCD medium. After 30 min incubation at 37°C, sperm were separated by pipetting and passing through a 70-μm filter. Sperm counts and motility assay were performed by using the DRM-600 CELLI-VU Sperm Counting CytoFecton. For in vitro fertilization (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 2–6 h, washed with PBS overnight, dehydrated with ethanol solution, embedded in paraffin and sectioned at 5 μm.

Sections were stained by H&E. 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 alkaline phosphatase staining. Embryo isolation and staging were done as described previously. The immunofluorescence analysis and alkaline phosphatase staining were performed essentially as described previously. The primary antibodies used were as follows: anti-AP-2α rabbit polyclonal, 1:500 (catalogue no. sc-8977; Santa Cruz Biotechnology); anti-ETM1/1/GLP mouse monoclonal, 5 μg ml\(^{-1}\) (catalogue no. PP-B422-00; R&D Systems); anti-Cbfa2t2 mouse monoclonal, 1:500 (catalogue no. ab1220; Abcam); anti-sox2 goat polyclonal, 1:200 (catalogue no. sc-1730, Santa Cruz Biotechnology).

The following secondary antibodies from Molecular Probes 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 (catalogue no. H-1000; Vector Laboratories). The immunofluorescence images and the alkaline phosphatase 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 HK2 and knock-out mutant ES cells were purified by acid extraction. 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 MS analysis. Desalted histone peptides (1 μg) were then loaded onto and separated by reversed-phase high-performance LC (HPLC) on a Thermo Scientific EasyEv-1000 system with a 75 μm i.d. × 15 cm (internal diameter and length) ReproSil-Pur C18 AQ 3 μm nanocolumn run at 300 n1 μml\(^{-1}\). 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 Thomson Scientific Q Exactive OrbiTrap Elite hybrid mass spectrometer. In each cycle, one full MS Orbitrap detection was performed with the scan range of 290 to 1,400 m/z, a resolution of 60 K and AGC of 1 × 10⁶. 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 × 10⁶. 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.
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 mES cell lysate. Human PRDM14 antibody is specific and does not cross-react with mouse PRDM14. c, Immunoprecipitation (IP) using antibodies against the indicated endogenous proteins in mES cells. d, Western blot of Superose 6 column fractionation of Flag-purified CBFA2T2 complex in NCCIT cells stably expressing Flag–HA–CBFA2T2. e, ChIP analysis using the indicated antibodies in 293T-REx harbouring 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 the Supplementary Information.
Extended Data Figure 2 | PRDM14 and CBFA2T2 exhibit an overlapping and interdependent distribution on chromatin in NCCIT cells. 

**a**. Heat map depicting PRDM14, CBFA2T2, RING1B and SUZ12 read density across a 5-kb window centred 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 (GO) analysis of PRDM14 and CBFA2T2 common target genes. 

**d**. Western blot analysis of PRDM14 and CBFA2T2 protein levels in knockdown (KD) experiments (Fig. 1d, e).
Extended Data Figure 3 | Characterization of knockout ES cell mutants and quantification of human iPS cell reprogramming efficiency.

**a, b,** Strategy for generating Prdm14- and Cbfa2t2-knockout (KO) mES cells via CRISPR–Cas9 genome editing. Sequencing chromatograms confirming homozygous disruption of the locus are depicted. **c,** Cbfa2t2- and Prdm14-knockout ES cells require 2i to maintain growth. ES cell lines generated under FBS plus LIF plus 2i conditions were continuously cultured in FBS plus LIF plus 2i (top, middle), or switched to FBS plus LIF (bottom). Eight days after 2i withdrawal (FBS plus LIF), well-formed ES cell colonies were undetectable; instead, mutant ES cells appeared to be differentiated. Scale bar, 100 μm. **d,** Proliferation rates of wild-type (WT) and mutant knockout ES cells as described in c. Data were obtained from three biological replicates. Please note error bars shown in the plots. Owing to the logarithmic scale used here, some error bars are very small and might be invisible. **e,** RNA-seq MA plot (log ratio (M) versus mean average (A)) in the indicated ES cells. Data are representative of three biological replicate experiments for each line. Mean abundance is plotted on the x axis and enrichment (both in log; scale) is plotted on the y axis. Genes depicted in red are differentially expressed with a FDR < 0.0001. **f,** Heat map showing relative expression of all differentially expressed genes as described in Fig. 2c. The only difference is now the heat map is centred on CBFA2T2 differentially expressed genes, rather than PRDM14 differentially expressed genes. **g,** Scheme of human fibroblast reprogramming to iPS cells. Fibroblasts were transduced with lentiviruses expressing polycistronic OCT4/KLF4/SOX2/c-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 iPS cell reprogramming efficiency based on TRA-1-81 staining with secondary antibody conjugated with horseradish peroxidase (HRP) and substrate DAB. Error bars are based on four biological replicates of each condition. The source data are included in the Supplementary Information.
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 wild-type (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 standard error of the mean. P value was determined by Student’s t-test. d, Near loss of gonocytes in Cbfa2t2-knockout mutant P0 testes by DDX4 (MVH) staining. Visualization of MVH-positive (red) gonocytes in Cbfa2t2\(^{+/+}\) (top) or Cbfa2t2\(^{-/-}\) (bottom) testis at P0 stage. The merged images with Hoechst (left; white) are shown on the right. Scale bars, 100 μm. e, Numbers of AP-2\(^{y}\)-positive PGCs in Cbfa2t2\(^{+/+}\) (black), Cbfa2t2\(^{+/+}\) (grey) 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. Student’s t-test: *P = 0.03, **P = 0.003. f, Numbers of AP-2\(^{y}\)-positive PGCs in Cbfa2t2\(^{+/+}\) (black), Cbfa2t2\(^{-/-}\) (grey) and Cbfa2t2\(^{-/-}\) (red) embryos at the indicated embryonic stages. 0B, zero-bud stage; LHF, late-head fold stage. g, Left, expression of SOX2 (red) in AP-2\(^{y}\)-positive (green) PGCs in Cbfa2t2\(^{+/+}\) (top) or Cbfa2t2\(^{-/-}\) (bottom) embryo at mid-bud stage, E7.25, shown as z-projection images of posterior confocal sections. Arrow indicates a minor PGC with relatively normal activation of SOX2. Scale bar, 50 mm. Right, percentage of SOX2-positive (red) cells in AP-2\(^{y}\)-positive (green) PGCs in the indicated genotypes of Cbfa2t2 at E7.0–7.25 (zero- to mid-bud stage) are shown with statistical significance (Student’s 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 (GO) analysis of PRDM14 ChIP-seq target genes. PRDM14 target genes are enriched in histone methyltransferase activities by DAVID functional annotation tool analysis. b, Cbfa2t2-m7 mutant genotyping. The mutant 7 amino acids are in red and corresponding wild-type (WT) residues are highlighted in blue in the displayed protein sequences. c, Bright-field images of wild-type and Cbfa2t2-m7 mES cells. Scale bar, 100 μm. d, Western blot analysis of PRDM14, CBFA2T2 and OCT4 protein levels in Prdm14-knockout (KO), Cbfa2t2-knockout or m7 mutant E5 cells under feeder-free FBS plus LIF plus 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 wild-type, Cbfa2t2-knockout and Cbfa2t2-m7 mES cells. ChIP-qPCR primer sequences are included in Supplementary Table 3. f, RT-qPCR quantification of Ehmt1 mRNA levels in wild-type and mutant lines. P values are 0.004 (**) and 0.0142 (*). The source data are included in the Supplementary Information. g, Mass spectrometry quantification of histone H3K9 modifications in wild-type and mutant lines. P values are 0.00956, 0.04248 (*). The source data are included in the Supplementary Information. h, i, Additional immunofluorescence analysis of H3K9me2 (red) of AP-2γ-positive (green; arrowheads) PGCs in Cbfa2t2+/− and Cbfa2t2−/− embryos at E8.75 as described in Fig. 4i.