nature cell biology

C/EBPα creates elite cells for iPSC reprogramming by upregulating Klf4 and increasing the levels of Lsd1 and Brd4

Bruno Di Stefano[1,2,8,9,10], Samuel Collombet[3,8], Janus Schou Jakobsen[4,5,6,8], Michael Wierer[7], Jose Luis Sardina[1,2], Andreas Lackner[1,2,9], Ralph Stadhouders[1,2], Carolina Segura-Morales[1,2], Mirko Francesconi[1,2], Francesco Limone[1,2], Matthias Mann[7], Bo Porse[4,5,6], Denis Thieffry[3] and Thomas Graf[1,2,10]

Reprogramming somatic cells into induced pluripotent stem cells (iPSCs) is typically inefficient and has been explained by elite-cell and stochastic models. We recently reported that B cells exposed to a pulse of C/EBPα (Bα′ cells) behave as elite cells, in that they can be rapidly and efficiently reprogrammed into iPSCs by the Yamanaka factors OSKM. Here we show that C/EBPα post-transcriptionally increases the abundance of several hundred proteins, including Lsd1, Hdac1, Brd4, Med1 and Cdk9, components of chromatin-modifying complexes present at super-enhancers. Lsd1 was found to be required for B cell gene silencing and Brd4 for the activation of the pluripotency program. C/EBPα also promotes chromatin accessibility in pluripotent cells and upregulates *Klf4* by binding to two haematopoietic enhancers. Bα′ cells share many properties with granulocyte/macrophage progenitors, naturally occurring elite cells that are obligate targets for leukaemic transformation, whose formation strictly requires C/EBPα.

The ability to reprogram somatic into pluripotent cells has revolutionized stem cell research with major implications for almost all fields of modern biology. Reprogramming to a pluripotent state can be achieved by overexpressing the transcription factors Oct4, Sox2, Klf4 and c-Myc (OSKM; refs 1,2). The resulting induced pluripotent stem cells (iPSCs) resemble embryonic stem cells (ESCs), being capable of contributing to chimaeric animals, including the germline[3,4].

iPS reprogramming of mouse embryo fibroblasts (MEFs) is accompanied by transcriptional and epigenetic remodelling[5,6], initiated by the downregulation of the somatic transcriptional program[7,8], followed by a mesenchymal-to-epithelial transition[9,10] (MET). After these changes, pluripotency genes, including *Oct4*, *Nanog* and *Sox2*, become expressed in a small number of cells in the course of about two weeks[7,8]. In pluripotent cells, ESC-specific super-enhancers[11] (ESC-SEs) bound by Brd4, Med1 and Cdk9, maintain the identity of the cells[12,13]. Increasing evidence indicates that reprogramming to pluripotency is a complex process, where multiple players synergistically establish new transcriptional networks and remove epigenetic barriers[14]. Among the factors that have been shown to affect the efficiency and kinetics of reprogramming are cell cycle regulators[15,16], chromatin remodellers[17–19] and facilitators of the MET transition[9,10,20,21].

However, a comprehensive understanding of the molecular mechanism of iPS reprogramming is still lacking, in large part because of the low reprogramming efficiency of most somatic cell types[14]. Yamanaka proposed two alternative explanations for this situation: according to the elite-cell model, reprogramming would take place only in a few predisposed cells within a population; and according to the stochastic model most or all cells are competent for reprogramming at low probabilities[22]. We have recently developed an approach that generates the equivalent of 'elite'-type cells, by transiently expressing in pre-B cells (henceforth called B cells) the transcription factor C/EBPα (ref. 23). Following OSKM activation,

[1]Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Dr. Aiguader 88, Barcelona 08003, Spain. [2]Universitat Pompeu Fabra (UPF), Barcelona 08003, Spain. [3]Institut de Biologie de l'Ecole Normale Supérieure (IBENS), CNRS UMR8197, INSERM U1024, Ecole Normale Supérieure, PSL Research University, F-75005 Paris, France. [4]The Finsen Laboratory, Rigshospitalet, Faculty of Health Sciences, University of Copenhagen, Ole Maaløes Vej 5, Copenhagen 2200, Denmark. [5]Biotech Research and Innovation Centre (BRIC), University of Copenhagen, Ole Maaløes Vej 5, Copenhagen 2200, Denmark. [6]Danish Stem Cell Center (DanStem), Faculty of Health Sciences, University of Copenhagen, 3B Blegdamsvej, Copenhagen 2200, Denmark. [7]Max Planck Institute of Biochemistry, 82152 Munich, Germany. [8]These authors contributed equally to this work. [9]Present addresses: Department of Molecular Biology, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts 02114, USA (B.D.S.); Max F. Perutz Laboratories, University of Vienna, Vienna Biocenter (VBC), Dr. Bohr-Gasse 9/3, 1030 Vienna, Austria (A.L.).
[10]Correspondence should be addressed to B.D.S. or T.G. (e-mail: distefano@molbio.mgh.harvard.edu or thomas.graf@crg.eu)

Received 29 July 2015; accepted 8 February 2016; published online 14 March 2016; DOI: 10.1038/ncb3326

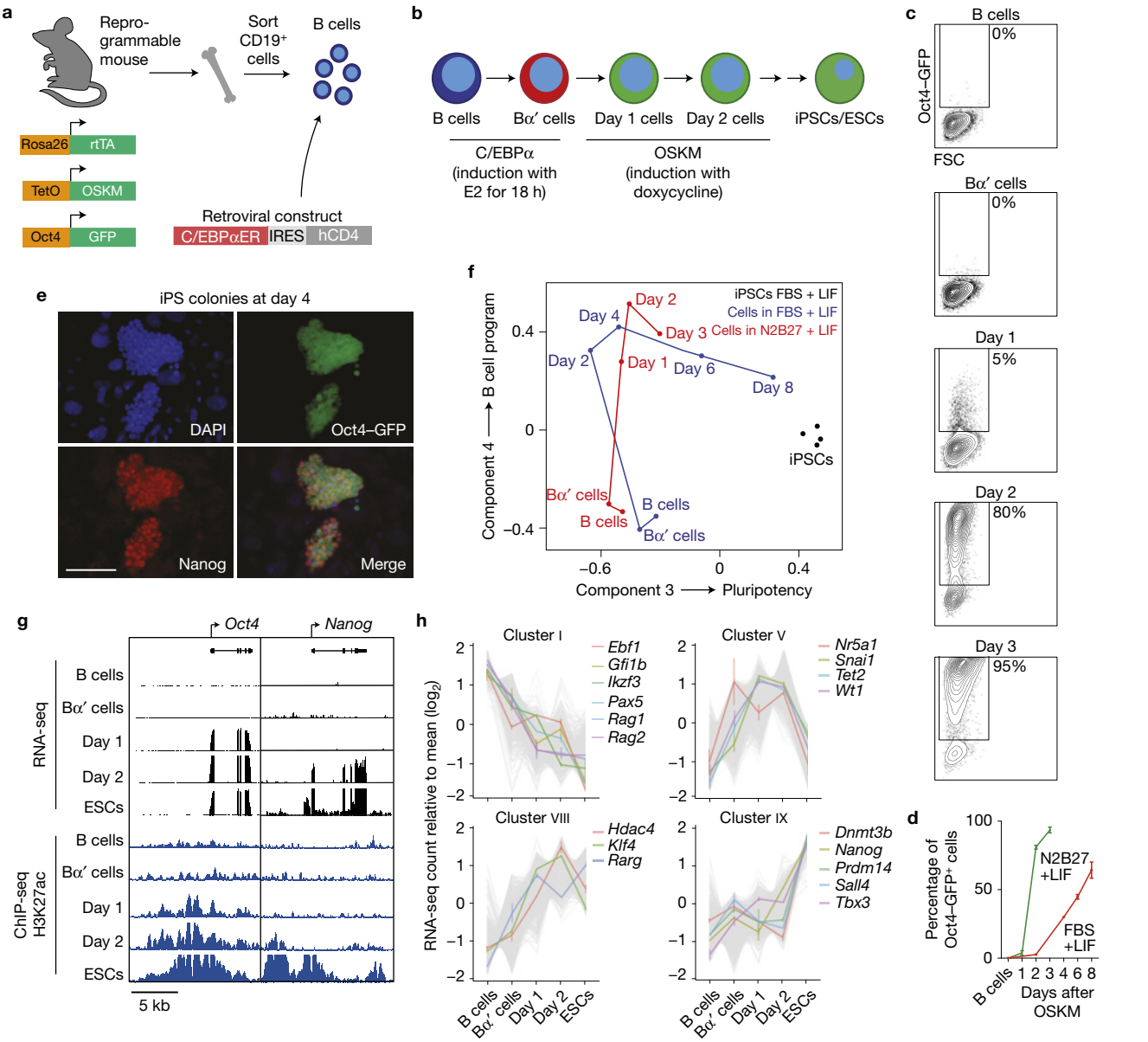


Figure 1 An improved ultrafast reprogramming system of B cells to pluripotency. (**a**) A schematic of the tools used. (**b**) An outline of the reprogramming strategy used. (**c**) Representative FACS analysis of Oct4–GFP expression during B to iPS cell reprogramming. (**d**) Comparison of Oct4–GFP kinetics of the data shown in **c** with data from ref. 23. Error bars indicate s.d. ($n=3$ biological independent experiments). (**e**) Representative immunofluorescence image showing Nanog expression in colonies of Bα′ cells treated for four days with doxycycline (scale bar, 100 μm). (**f**) Independent component analysis of Agilent microarray expression data obtained from cells during reprogramming in serum-free (N2B27+LIF) and serum-containing[23] (FBS+LIF) conditions compared with four iPSC lines. Each time point represents the average from duplicates. (**g**) RNA-seq and H3K27ac ChIP-seq data for the *Oct4* and *Nanog* loci at the indicated times. ESCs were used as controls. (**h**) Representative RNA-seq expression kinetics of four selected gene clusters (see Supplementary Fig. 1B for a complete overview).

these Bα′ cells—unlike control B cells—can be converted into iPSCs at nearly 100% efficiency, also showing that cell reprogramming is basically a deterministic process[23]. Similar efficiencies have been described for genetically modified MEFs and cells treated with specific compound combinations[24–26], including fibroblasts and granulocyte/macrophage progenitors[27,28] (GMPs).

Here we have studied the changes induced in B cells by a C/EBPα pulse as well as during the subsequent rapid onset of OSKM-induced reprogramming, by dynamically monitoring transcription-factor-induced changes in transcription, protein expression, enhancer activity and chromatin accessibility.

RESULTS

Changes in culture conditions result in an ultrafast reprogramming system

We have now improved our B cell reprogramming system by culturing the cells under conditions that favour naive pluripotency[29]. Briefly, pre-B cells from reprogrammable mice crossed with Oct4–GFP mice

were infected with C/EBPαER retrovirus (Fig. 1a), seeded on MEFs in serum-free medium supplemented with LIF and treated for 18 h with β-oestradiol (E2) to activate C/EBPα, followed by doxycycline to activate OSKM (Fig. 1b). Two to five per cent of the cells became Oct4–GFP positive within 1 day and 95% after 3 days (Fig. 1c,d), and Nanog-positive iPSC colonies could already be detected at day 4 (Fig. 1e). A comparison of the Oct4–GFP kinetics with cells cultured in medium containing FBS and LIF showed a 4 to 5 day acceleration under the new culture conditions[23] (Fig. 1d), which was also confirmed by an independent component analysis of the transcriptome (Fig. 1f and Supplementary Table 1). The resulting iPSCs were found to contribute to chimaera formation after blastocyst injection (Supplementary Fig. 1A).

RNA-seq experiments showed the activation of *Oct4* within 1 day after OSKM induction and of *Nanog* within 2 days, correlating with the genes' decoration by the activating mark H3K27ac (Fig. 1g). The 12,781 genes whose expression changed >2-fold at any given time point were subdivided into nine groups by C-means clustering, and analysed for Gene Ontology (GO) annotations (Fig. 1h and Supplementary Fig. 1B,C). Genes in clusters I and II became downregulated following the C/EBPα pulse/OSKM expression and were enriched for the GO terms 'B cell activation' and 'immune response'. In contrast, genes in cluster III became transiently upregulated and were enriched for 'epigenetic regulation of gene expression' and 'chromatin organization', whereas cluster V was enriched for 'mesenchymal–epithelial transition (MET)' and 'cell adhesion'. Finally, genes in clusters VIII and IX were upregulated by OSKM and enriched for 'stem cell maintenance' and 'blastocyst formation', containing the pluripotency genes *Nanog* and *Klf4*.

Pluripotency genes were activated in three waves, initiated by *Oct4*, *Lin28a*, *Zfp296*, *Gdf3* and *Tdh* at day 1, followed by *Nanog* at day 2, and by *Sall4*, *Sox2*, *Esrrb* and *Zfp42* (*Rex1*) at day 4 (Supplementary Fig. 1D). Changes in expression of pluripotency genes were confirmed at the protein level (Supplementary Fig. 1E).

In the following, we will use this ultrafast reprogramming system to investigate changes at the protein and chromatin levels, both after the initial expression of C/EBPα and following OSKM induction.

The C/EBPα pulse elevates the levels of multiple proteins, without concomitant transcriptional changes

To analyse proteome dynamics during reprogramming, we performed shotgun proteomics by label-free quantification of samples from the four reprogramming time points and ESCs. We detected 7,497 proteins that included 520 transcription factors, 295 kinases, 141 phosphatases and 96 isomerases (Supplementary Fig. 2A), showing an excellent correlation between duplicates (Supplementary Fig. 2B). C-means clustering of the proteins that changed >2-fold at any given time point resulted in ten groups (Fig. 2a and Supplementary Fig. 2C). Downregulated proteins (clusters a,b) were highly enriched for the GO terms 'B cell immunity' and 'immune system' (Fig. 2b). The early upregulated proteins were found to be associated with the terms 'RNA splicing' and 'protein degradation' (Supplementary Table 2). In contrast, late upregulated proteins (clusters i,j) were enriched for 'stem cell maintenance' and 'stem cell development' and contained Oct4 and Sox2. Surprisingly, 439 out of 538 proteins elevated in Bα′ cells (clusters g,h) were not regulated at the RNA level (Fig. 2c). These belonged to the categories 'chromatin remodelling' and 'histone modifications' and included the epigenetic factors Lsd1, Hdac1, Brd4 and Med1. The observed accumulation at the protein level of the histone demethylase Lsd1 (Kdm1A) and the histone deacetylase Hdac1 (Fig. 3a) could also be confirmed by western blot (Supplementary Fig. 3A).

A protein complex containing Lsd1, Hdac1 and C/EBPα mediates B cell gene silencing

Lsd1 is known to demethylate H3K4me1/2 (ref. 30) and to be part of a complex that includes the histone deacetylase Hdac1 (ref. 31). Hdac1 in turn has been described to interact with C/EBPα (ref. 32). We therefore investigated whether C/EBPα can interact with both Lsd1 and Hdac1 and whether the complex is required to silence B cell gene expression (Supplementary Fig. 3B) during reprogramming. Gel-filtration chromatography showed that C/EBPα, Lsd1 and Hdac1 co-elute in a high-molecular-weight complex (Supplementary Fig. 3D). Furthermore, C/EBPα immunoprecipitation followed by mass spectrometry revealed an interaction with Lsd1 and Hdac1 (Fig. 3b and Supplementary Fig. 3E), a finding confirmed by western blot (Supplementary Fig. 3F). Our proteomic analyses also showed an interaction of C/EBPα with Wdr5 and other members of the MLL complex, as recently reported[33]. Finally, an antibody against Lsd1 co-immunoprecipitated Hdac1 and C/EBPα (Fig. 3c), but not the control proteins Pcna and Parp1 (Supplementary Fig. 3G), further supporting the proposed association between C/EBPα, Lsd1 and Hdac1.

As the Lsd1/Hdac1 complex has been shown to regulate the inactivation of super-enhancers[31] (SEs), we tested the effect of C/EBPα on B cell SEs (B-SEs), examining 514 regions with high H3K27ac levels[11,34]. This revealed a marked reduction in B-SE activity, as seen by the loss of H3K27ac and Brd4 (Fig. 3d and Supplementary Fig. 3H). We also found a significant decrease in the activation-associated marker H3K4me2 at regulatory elements of the B cell genes *Ebf1*, *Fox1*, *Gfi1b* and *Ikzf3* (ref. 35), as well as of the lymphoid *Rag1* and *Ciita* genes (Fig. 3e).

To test whether Lsd1 is required for the C/EBPα-mediated B cell silencing (Supplementary Fig. 3B) and decrease in B-SE activity, we tested the effect of S2101, a compound that specifically blocks the enzymatic activity of Lsd1 (ref. 36). We found that it indeed prevented the C/EBPα-induced decommissioning of B cell enhancers (Fig. 3e) and, at least in part, B cell silencing (Fig. 3f), as did Hdac1 inhibition (Supplementary Fig. 3I). Lsd1 was also found to be required for the fast reprogramming of B cells into iPS cells as S2101 treatment impaired Oct4–GFP upregulation and markedly reduced the number of iPS colonies (Fig. 3g–i). This was not due to an adverse effect of Lsd1 inhibition on B cell viability (Supplementary Fig. 4A) or proliferative capacity (Supplementary Fig. 4B). Knockdown experiments confirmed these observations, yielding reduced numbers of iPS colonies (Fig. 3j and Supplementary Fig. 4C–E). Interestingly, OSKM-induced reprogramming of B cells not exposed to C/EBPα was not inhibited by S2101 (Supplementary Fig. 4F), indicating that Lsd1 acts downstream of C/EBPα.

These results show that inhibition of Lsd1 and Hdac1 impairs enhancer decommissioning and B cell gene silencing in Bα′ cells.

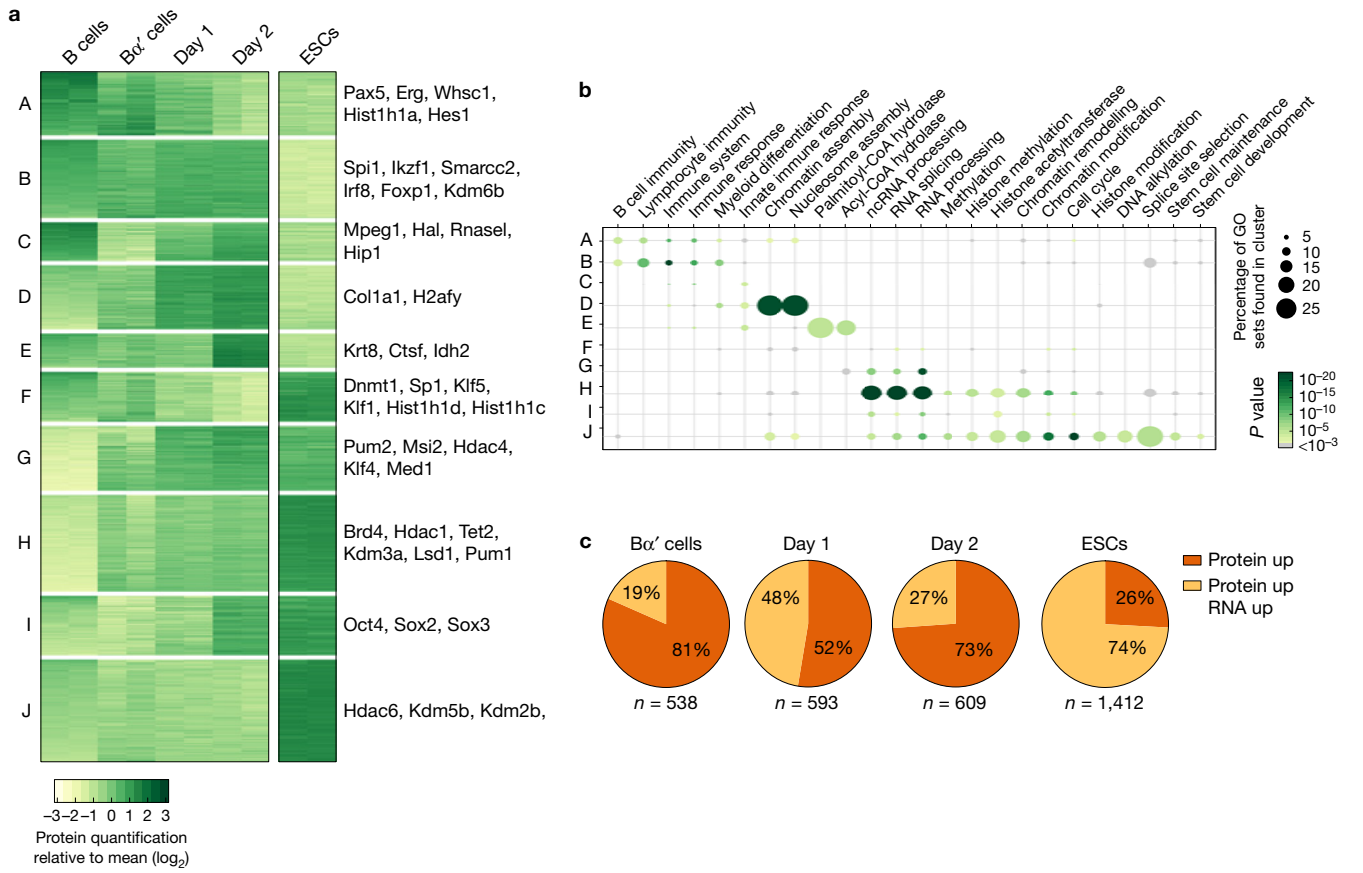


Figure 2 C/EBPα- and OSKM-induced proteome changes during reprogramming. (**a**) C-means clustering of proteins whose abundance changes >2-fold during reprogramming. Selected members of each cluster are indicated. (**b**) GO analysis of protein clusters shown in **a**. The size of each circle represents the proportion of GO sets found in each cluster; the intensity of the colour represents the *P* value as determined by a hypergeometric test. (**c**) Percentage of proteins upregulated during reprogramming for each time point with respect to the previous one; that is, the values shown for Bα′ cells are relative to B cells. In orange are proteins whose abundance increases without concomitant transcriptome change; in ochre are proteins that become upregulated at both the protein and transcriptional levels.

C/EBPα post-transcriptionally increases Brd4, Med1 and Cdk9 protein levels and Brd4 is required for reprogramming

The C/EBPα pulse also caused a marked and coordinated post-transcriptional increase of Brd4, Med1 and Cdk9 as detected by mass spectrometry and western blot (Fig. 4a and Supplementary Fig. 3A,C). This is intriguing as recent studies have shown that a complex containing Brd4, Med1 and Cdk9 is required for the self-renewal capacity and pluripotency of ESCs (ref. 12) and for reprogramming of MEFs (ref. 13), and that Brd4 inhibition results in the complex's disruption[12]. As Brd4 is also required for SE activity[37], we analysed the changes of H3K27ac decoration at ESC-SEs during reprogramming. Notably, some ESC-SEs became already activated in Bα′ cells (Fig. 4b), including the *Lefty* and *Rarg* enhancers, and 19/32 showed increased Brd4 binding. This finding raised the possibility that C/EBPα recruits Brd4 to acetylated histone tail residues[38]. To test this hypothesis, we performed ChIP-seq experiments with B cells and Bα′ cells and found that C/EBPα and Brd4 co-occupy approximately 25% of the newly bound sites, including ESC-SEs of *Id1*, *Iqgap1*, *Rarg* and *Egln3* genes (Fig. 4c,d and Supplementary Fig. 4G). Co-immunoprecipitations further confirmed that the two proteins form a complex (Fig. 4e), as was also recently described[38].

Next we tested whether Brd4 inhibition impairs iPS reprogramming by using the compound JQ1, known to effectively inhibit the function of bromodomain and extraterminal domain (BET) family proteins, including Brd4 (ref. 39). JQ1 was either administered only during the pulse or during a two-day doxycycline treatment. Cells were monitored for the expression of Oct4–GFP by fluorescence-activated cell sorting (FACS) or of key pluripotency genes. Strikingly, JQ1 treatment completely inhibited Oct4–GFP activation (Fig. 4f) and significantly reduced the upregulation of *Oct4*, *Lin28a*, *Zfp296*, *Lefty1* and *Cdh1* when administered during both the C/EBPα pulse and OSKM activation (Fig. 4g). Finally, we observed a threefold reduction of iPSC colonies when the drug was administered during the C/EBPα pulse and a tenfold impairment when provided together with doxycycline (Fig. 4h and Supplementary Fig. 4H), a finding confirmed with an short hairpin RNA (shRNA) against Brd4 (Fig. 4i and Supplementary Fig. 4C–E). Of note, the concentration of JQ1 used did not severely affect B cell survival or proliferation (Supplementary Fig. 4A,B).

We conclude that C/EBPα induces a post-transcriptional accumulation of the Brd4–Med1–Cdk9 complex and re-localizes Brd4 to ES-SEs. Our data also show that Brd4 is strictly required for iPS cell formation from B cells.

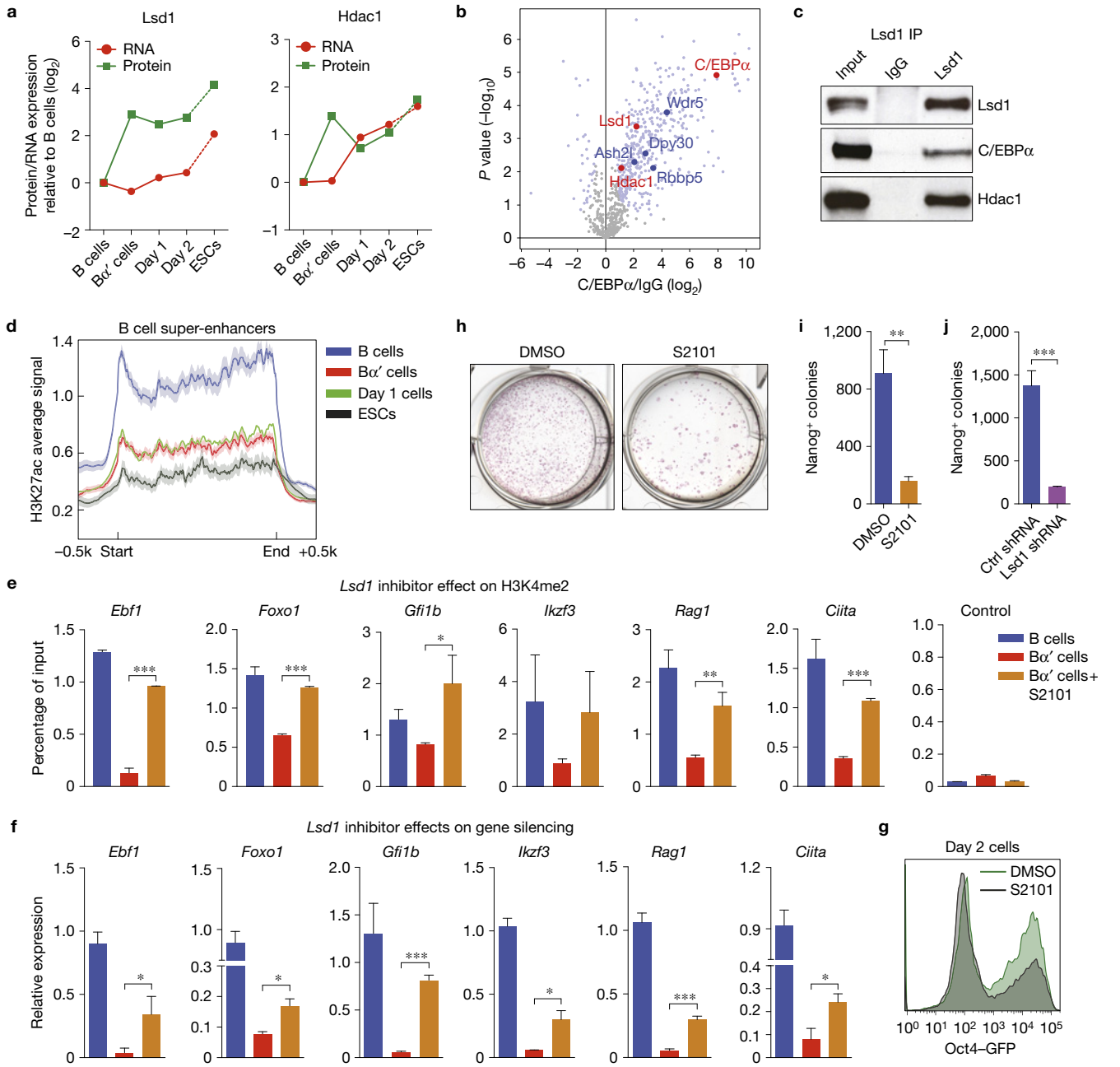


Figure 3 C/EBPα-induced Lsd1 and Hdac1 upregulation and requirement of Lsd1 for silencing of the B cell program. (**a**) Expression of Lsd1 and Hdac1 mRNA and proteins during reprogramming compared with ESCs. The data represent the average from two biologically independent samples. (**b**) C/EBPα immunoprecipitation mass-spectrometry data (based on three experiments using a specific antibody and three using IgG control serum). Statistical significance was determined using a two-tailed unpaired Student's *t*-test ($n = 3$ biologically independent samples). Small grey dots: non-significantly enriched proteins (FDR > 0.05). Small blue dots: significantly enriched proteins (FDR < 0.05). Large blue dots: proteins from the MLL complex and Top2a. Red dots: see main text. (**c**) Representative western blot of co-immunoprecipitation experiments. Extracts from Bα′ cells were probed with Lsd1, C/EBPα or Hdac1 antibodies. Unprocessed original scans of blots are shown in Supplementary Fig. 8. (**d**) H3K27ac levels at B cell super-enhancers in B cells, Bα′ cells, day 1 cells and ESCs. (**e**) H3K4me2 at selected B cell enhancers as measured by ChIP-qPCR in B cells, Bα′ cells and Bα′ cells treated with the Lsd1 inhibitor S2101 during the pulse. Error bars indicate s.d. ($n = 3$ biologically independent samples). Statistical significance was determined using a two-tailed unpaired Student's *t*-test ($^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$). (**f**) Gene expression by qRT-PCR of samples shown in **d**. Values were normalized against *Pgk* expression. Error bars indicate s.d. ($n = 3$ biologically independent samples). Statistical significance was determined using a two-tailed unpaired Student's *t*-test ($^{*}P < 0.05$, $^{***}P < 0.001$). (**g**) Representative Oct4–GFP FACS analysis of B cells treated with S2101 at day 2 of reprogramming. (**h**) Representative alkaline-phosphatase-positive iPS colonies obtained from reprogramming of Bα′ cells induced with OSKM and treated with S2101 or dimethylsulphoxide as a control. (**i**) Nanog^{+} iPSC colony counts at day 12 of reprogramming of cells treated with S2101 or dimethylsulphoxide as control. Error bars indicate s.d. ($n = 3$ biological independent experiments). Statistical significance was determined using a two-tailed unpaired Student's *t*-test ($^{**}P < 0.01$). (**j**) Nanog^{+} iPSC colony counts at day 12 of reprogrammed cells with a knockdown of Lsd1. Error bars indicate s.d. ($n = 3$ biological independent experiments). Statistical significance was determined using a two-tailed unpaired Student's *t*-test ($^{***}P < 0.001$).

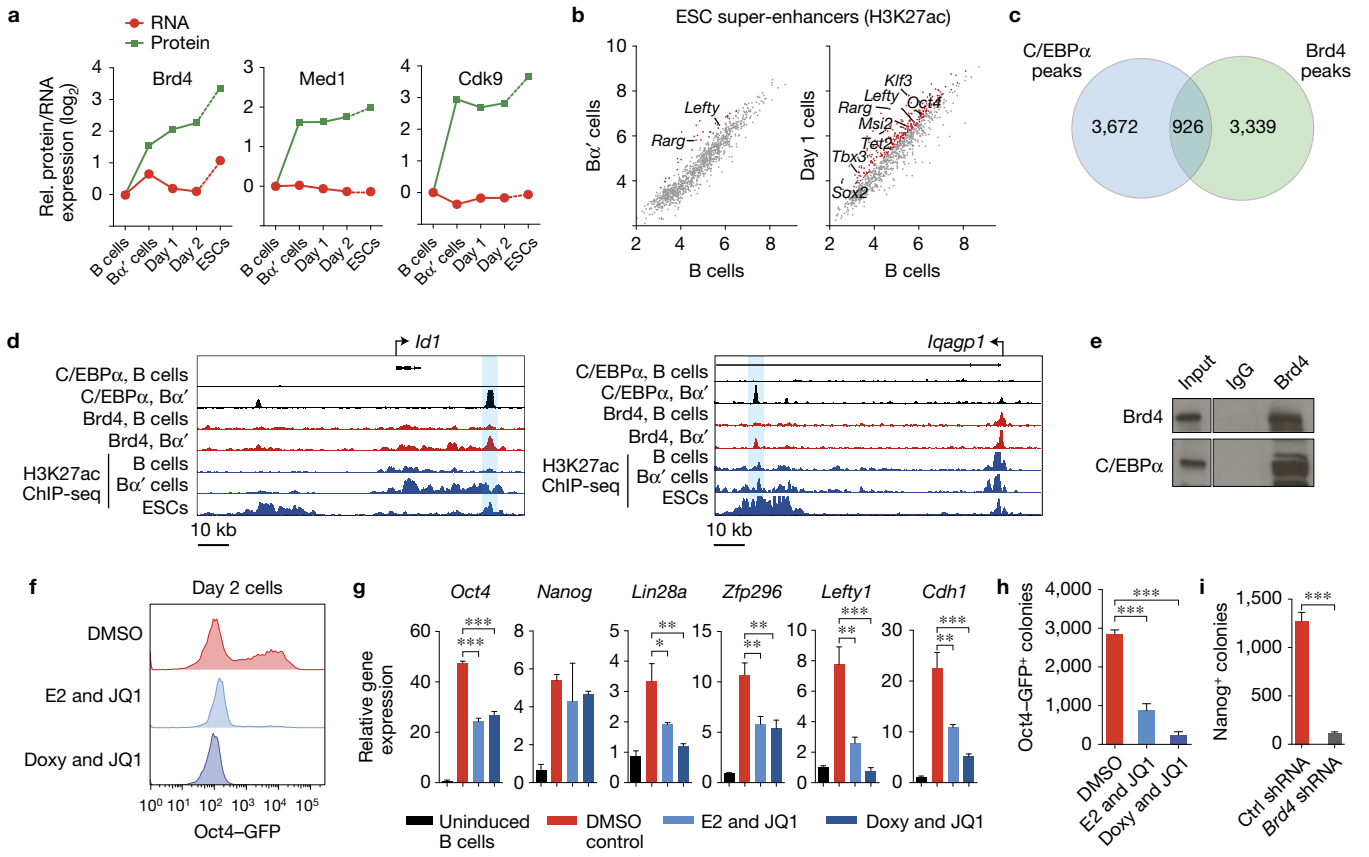


Figure 4 C/EBPα-induced Brd4, Med1 and Cdk9 upregulation and requirement of Brd4 for activation of the pluripotency program. (**a**) Expression of Brd4, Med1 and Cdk9 mRNA and proteins during reprogramming and in ESCs. The data represent the average from two biologically independent samples. (**b**) Activation of ESC-SEs during reprogramming as measured by H3K27Ac, comparing B with Bα′ cells and B with day 1 cells. Axes show the RPK score (log scale) for H3K27ac. Selected genes associated with super-enhancers showing a >1.5-fold increase in activity compared with B cells are highlighted in red. (**c**) Venn diagram showing the overlap between C/EBPα and Brd4 ChIP-seq peaks in Bα′ cells. (**d**) Representative genome browser screenshots of the *Id1* and *Iqagp1* genes showing C/EBPα, H3K27ac and Brd4 ChIP-seq data. (**e**) Representative western blot of co-immunoprecipitation experiment. Extracts from Bα′ cells were probed with Brd4 or C/EBPα antibodies. Unprocessed original scans of blots are shown in Supplementary Fig. 8. (**f**) Representative Oct4–GFP FACS analysis of B cells treated with JQ1 for 18 h or 2 days. (**g**) Effect of JQ1 on the expression of selected pluripotency genes by qRT-PCR. Values were normalized against *Pgk* expression. Error bars indicate s.d. ($n=3$ biologically independent samples). Statistical significance was determined using a two-tailed unpaired Student's *t*-test ($^{*}P<0.05$, $^{**}P<0.01$, $^{***}P<0.001$). (**h**) Oct4–GFP^{+} iPSC colonies at day 12 of reprogramming, after treatment of B cells with JQ1 during either C/EBPα or OSKM induction. Error bars indicate s.d. ($n=3$ biologically independent samples). Statistical significance was determined using a two-tailed unpaired Student's *t*-test ($^{***}P<0.001$). (**i**) Nanog^{+} iPSC colony counts after reprogramming of B cells with a knockdown for Brd4. Error bars indicate s.d. ($n=3$ biologically independent samples). Statistical significance was determined using a two-tailed unpaired Student's *t*-test ($^{***}P<0.001$).

C/EBPα induces chromatin accessibility at selected pluripotency loci

To study changes in chromatin accessibility, we used ATAC-seq[40] to identify accessible chromatin regions and reveal potential regulatory elements bound by transcription factors. After calling discrete ATAC peaks, we identified 6,319 regions with distinct dynamics during the time course, which were divided by C-means clustering into four groups (Fig. 5a). Cluster I contained 525 regions with ATAC peaks formed *de novo* after the C/EBPα pulse and that reach maximum accessibility in ESCs. Their association with GO categories 'developmental process' and 'embryo development' (Fig. 5a and Supplementary Fig. 5A) suggests that they mark sites associated with pluripotency. In contrast, most ATAC peaks induced by the C/EBPα pulse, represented in clusters II, III and IV, diminished in day 2 cells and were poorly represented in ESCs. GO analyses of genes associated with these regions revealed an enrichment for myeloid genes (Supplementary Fig. 5A) as exemplified by *Id1* and *Ifitm6*, genes that are upregulated by C/EBPα (Supplementary Fig. 5B). We then investigated the presence of transcription-factor-binding motifs within ATAC-seq peaks. As predicted, cluster I regions were strongly enriched for motifs associated with pluripotency transcription factors such as Klf4, Oct4, Sox2 and Essrb. In contrast, newly accessible regions in clusters II, III and IV were enriched for potential binding sites of the myeloid transcription factors PU.1 (Ets), C/EBPα, Irf8 and Runx1 (Fig. 5b, and Supplementary Fig. 5C).

C/EBPα directly regulates Klf4 expression and together they increase chromatin accessibility

Klf motif enrichment in ATAC cluster I suggested that Klf4 might already be expressed in Bα′ cells. We found that it was

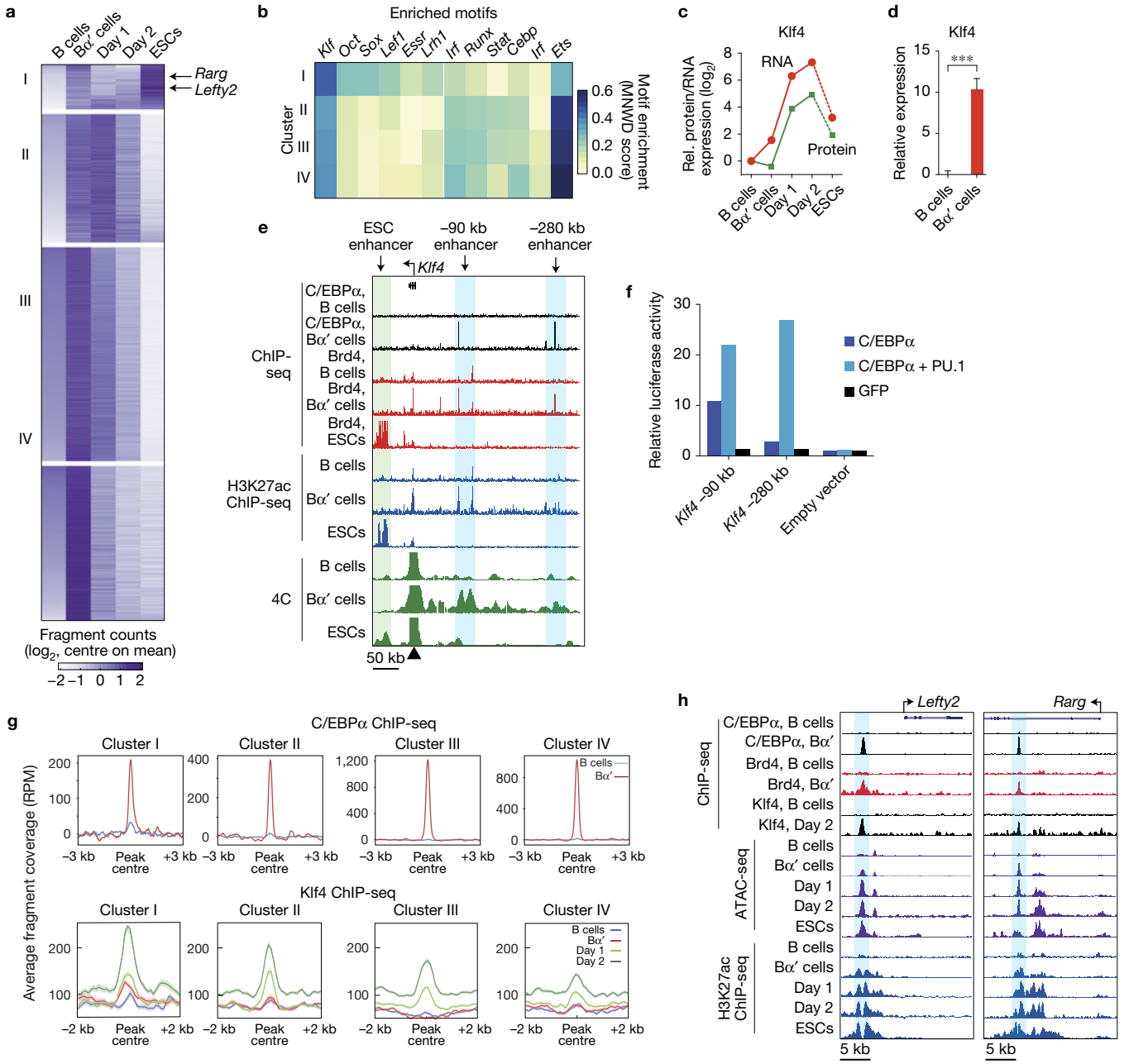


Figure 5 Analysis of chromatin accessibility changes and transcription factor binding in C/EBPα-pulsed B cells. (**a**) C-means clustering of newly formed accessible chromatin regions (as measured by ATAC-seq) in Bα′ cells compared with B cells, yielding 4 clusters. (**b**) Motif analysis of ATAC-seq peaks within the clusters identified in **a**. (**c**) mRNA and protein expression kinetics of Klf4 during reprogramming. The data represent the average from two biologically independent samples. (**d**) Comparison of Klf4 expression by qRT-PCR between B cells and Bα′ cells. Error bars indicate s.d. ($n=3$ biologically independent samples). Statistical significance was determined using a two-tailed unpaired Student's *t*-test ($^{***}P<0.001$). (**e**) Genome browser screenshot of the *Klf4* locus, showing ChIP-seq data for C/EBPα, Brd4 and H3K27ac, as well as 4C data using the *Klf4* promoter as the view point (black triangle at the bottom). (**f**) Luciferase reporter assay of 293T cells transfected with luciferase constructs containing only a minimal promoter (empty vector) or also one of the two putative *Klf4* enhancers at −90 kb and −280 kb predicted to be regulated by C/EBPα and PU.1. Plasmids encoding C/EBPα, C/EBPα and PU.1 or GFP were co-transfected. The data represent the average from two biologically independent samples. (**g**) C/EBPα and Klf4 ChIP-seq analysis in B, Bα′, Day 1 and Day 2 cells, showing average peak distribution. (**h**) Representative genome browser screenshots of *Rarg* and *Lefty2*, showing ChIP-seq data for C/EBPα, Klf4 and H3K27ac in Bα′ cells, day 2 cells and ESCs.

indeed transcriptionally upregulated >2-fold by C/EBPα (Fig. 5c,d), a finding confirmed by western blot (Supplementary Fig. 3A). Remarkably, C/EBPα binds to two regions upstream of Klf4 (−90 kb and −280 kb) that are also occupied by Brd4 and PU.1 and were enriched for H3K27ac, suggesting that these sites represent active enhancers (Fig. 5e and Supplementary Fig. 5D). Circularized chromosome conformation capture (4C-Seq) experiments support this interpretation, revealing that the *Klf4* promoter interacts with

these two regions in Bα′ cells but not in B cells (Fig. 5e and Supplementary Fig. 5D). Interestingly, in ESCs these enhancers showed no activity and no looping with the *Klf4* promoter, which interacted instead with a +50 kb enhancer (Fig. 5e). In addition, co-transfection of the −90 kb enhancer linked to a luciferase reporter with C/EBPα enhanced reporter activity, whereas both C/EBPα and PU.1 were necessary to activate the −280 kb enhancer (Fig. 5f). Their differential responsiveness to the two transcription factors can be explained by the observation that in B cells the −280 kb, but not the −90 kb enhancer, is primed by endogenous PU.1 (Supplementary Fig. 5D), a transcription factor known to synergize with C/EBPα (ref. 41).

To explore the possibility of interplay between C/EBPα and Klf4, we compared the binding sites of the two factors and analysed their enrichment at the ATAC peaks within the four clusters. Supporting our motif analyses, Klf4 binding showed the strongest enrichment at cluster I sites, decreasing progressively towards cluster IV, whereas C/EBPα binding showed the reverse trend (Fig. 5g). In ESCs, cluster I sites were specifically enriched for Brd4 and Klf4 binding, highlighting the known interaction between the two factors[13] (Supplementary Fig. 5E). Our data also showed that approximately 5% of regions bound by C/EBPα in Bα′ cells are bound by Klf4 in pluripotent cells (Supplementary Fig. 5F), raising the possibility that here C/EBPα acts as a pioneer factor for subsequent Klf4 binding. Analysis of MNase-seq and C/EBPα ChIP-seq data obtained after C/EBPα activation[41], generated in an inducible pre-B cell line, support this hypothesis (Supplementary Fig. 5G). A pioneering role of C/EBPα is also suggested for the pluripotency-associated genes *Rarg* and *Lefty1/2*. C/EBPα binds to the enhancers of these genes in Bα′ cells, which then become activated, as seen by their decoration with H3K27ac and Brd4 (Fig. 5h). Following OSKM induction, these enhancers become subsequently bound by Klf4 and further activated. Moreover, both regions are bound by Oct4, Nanog, Sox2, Klf4 and Brd4 in ESCs and are also enriched for H3K27ac, confirming that they correspond to active pluripotency enhancers (Supplementary Fig. 6A). Consistent with this interpretation, *Rarg* and *Lefty1* become upregulated during reprogramming (Supplementary Fig. 6B).

In conclusion, C/EBPα creates a large number of newly accessible chromatin regions in B cells and it upregulates *Klf4* by binding to two haematopoietic enhancers. C/EBPα and Klf4 subsequently facilitate accessibility to other pluripotency-associated transcription factors, as exemplified for the *Lefty* and *Rarg* enhancers.

C/EBPα and Klf4 link the elite cell state of Bα′ cells to that of granulocyte/macrophage progenitors

GMPs are the cells within the haematopoietic system most susceptible to OSKM-induced reprogramming, with 25% to >90% % of single seeded cells converting into iPSCs (refs 25–28; summarized in Fig. 6a), thus representing an 'elite' or 'privileged' cell state. C/EBPα-deficient mice lack GMPs (ref. 42), showing that their formation strictly requires C/EBPα. To compare Bα′ cells and GMPs, we performed expression microarray and RNA-seq experiments. *Cebpa* was expressed >100-fold higher in GMPs than in B cells and ESCs (Fig. 6b), whereas *Klf4* was expressed about fivefold higher in Bα′ cells and GMPs compared with B cells (Fig. 6b). Moreover, C/EBPα binds to the −90 and −280 kb Klf4 enhancers in GMPs (ref. 43; Fig. 6c), raising the possibility that C/EBPα and Klf4 co-regulate a substantial proportion of genes in GMPs and Bα′ cells. Indeed, almost 30% of the genes expressed in GMPs—but not B cells—were also found to be expressed in Bα′ cells (Supplementary Fig. 6C). Furthermore, a canonical variate analysis showed that Bα′ cells are most similar to GMPs when compared with all other haematopoietic populations (Supplementary Fig. 6D). We then performed ATAC-seq in GMPs to test whether chromatin accessible regions in Bα′ cells are also accessible in these cells. Strikingly, of 525 newly accessible regions in Bα′ cells, 357 overlapped with ATAC peaks in GMPs (Fig. 6d and Supplementary Fig. 6E,F), which were not detectable in MEFs (ref. 44; Supplementary Fig. 6F). This is exemplified by the pluripotency genes *Lefty* and *Rarg* and the myeloid genes *Tet2*, *Ifitm6* and *Id1* (Supplementary Fig. 6G). Furthermore, the intersection between C/EBPα and ATAC-seq peaks in Bα′ cells shows that C/EBPα binding exhibits a similar enrichment in GMPs (Supplementary Fig. 6H and Fig. 5g).

Rapidly cycling GMPs express elevated levels of Klf4 and Tet2

It was recently reported that a rapidly cycling GMP sub-fraction can be reprogrammed into iPSCs at a >3-fold higher efficiency than the slowly cycling sub-fraction[28]. We therefore determined whether the elite-cell-associated factors Klf4 and Tet2, which have been shown to enhance the reprogramming efficiency of B cells and MEFs (refs 23,45), are differentially expressed in the two cell fractions. For this purpose, GMPs from reprogrammable mice were labelled with CFSE and separated into a CSFE-high fraction (slow-cycling cells) and a CSFE-low fraction (fast-cycling cells; Supplementary Fig. 7A). The fast-cycling GMPs expressed twofold more *Klf4* and 1.5-fold more *Tet2* than the slow-cycling cells, a finding confirmed by RNA-seq and quantitative real-time PCR (qRT-PCR) for *Klf4* (Fig. 6e,f and Supplementary Fig. 7B) whereas the slow-dividing cells expressed higher levels of the cell cycle inhibitor *Cdkn1a* (p21; Fig. 6e). In contrast, the two cell fractions did not differ in their expression of *Cebpa* and *Cebpb* nor of *Lsd1*, *Hdac1*, *Brd4*, *Med1* and *Cdk9* (Supplementary Fig. 7C), in line with the observation that C/EBPα regulates these genes only at the protein level. Finally, we examined the role of Tet2 in the reprogramming to iPSCs of both B cells and GMPs. After Tet2 knockdown, we observed a significant decrease in the number of iPSC colonies for both cell types (Fig. 6g,h and Supplementary Fig. 7D).

DISCUSSION

Our results describe molecular mechanisms by which a pulse of the transcription factor C/EBPα converts B cells into elite cells for reprogramming into iPSCs. C/EBPα initiates the reprogramming cascade by concomitantly boosting the levels of protein complexes required for silencing the B cell program and for establishing the pluripotency program. C/EBPα also controls Klf4 expression by binding to two haematopoietic specific enhancers, allowing the two factors to induce chromatin accessibility at regions that become fully accessible in pluripotent cells (Fig. 7a,b).

Our most surprising observation was that C/EBPα post-transcriptionally increases the levels of hundreds of proteins, including key chromatin-modifying factors. The finding that C/EBPα interacts with Lsd1 (Fig. 3b,c), Hdac1 (Fig. 3b,c and ref. 32) and Brd4 (Fig. 4e) raises the possibility that it could increase protein stability

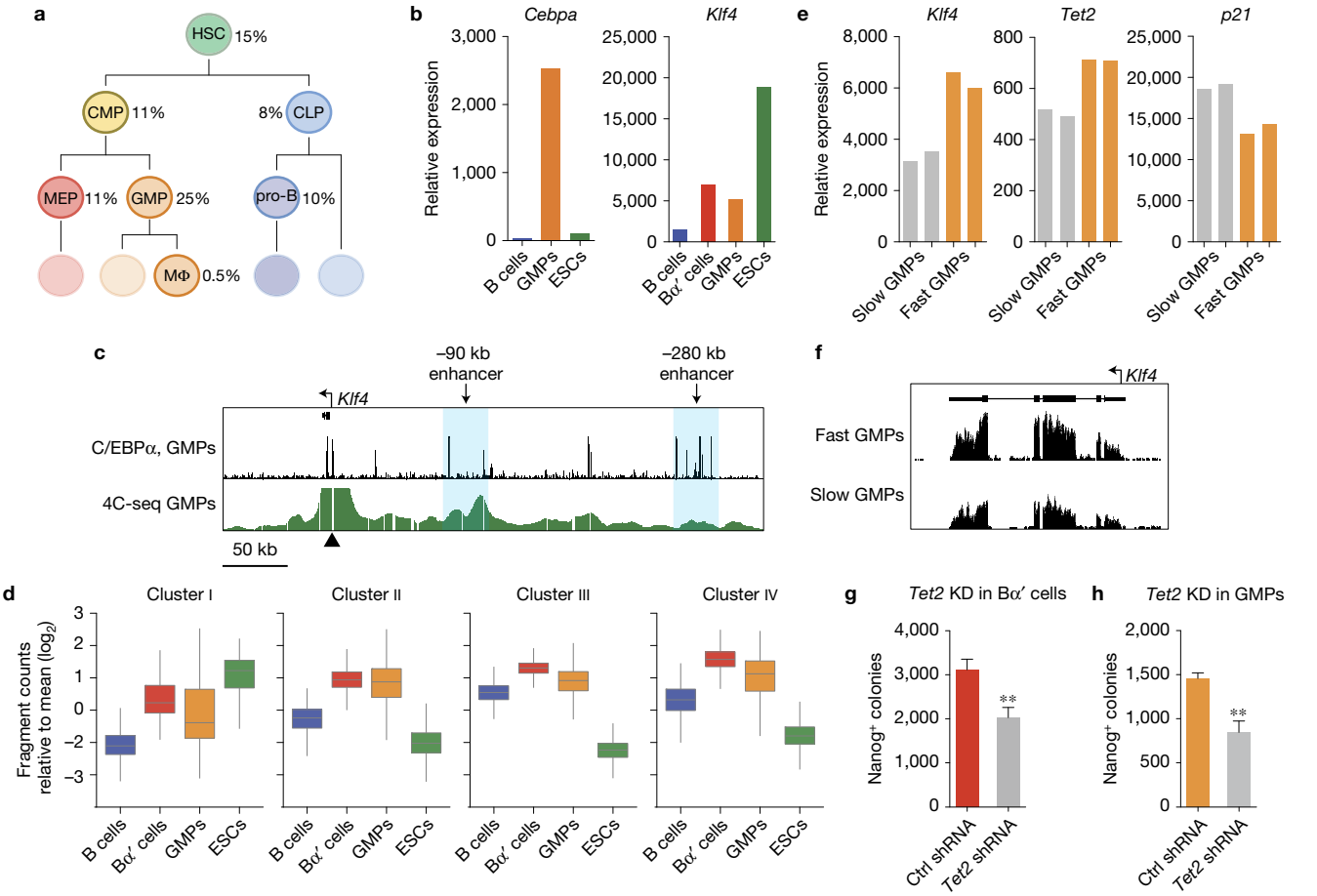


Figure 6 Comparison between molecular properties of Bα′ cells and GMPs. (**a**) A schematic of the haematopoietic lineage tree showing the percentage of single cells that could be reprogrammed into iPSCs as determined in ref. 27. (**b**) Comparison of *Cebpa* and *Klf4* expression in B cells, GMPs and ESCs (expression microarrays). The data represent the average from two biologically independent samples. (**c**) Genome browser screenshot of the *Klf4* locus, showing C/EBPα ChIP-seq in GMPs, and 4C using the *Klf4* promoter as the view point (black triangle, similar to Fig. 5e). (**d**) Boxplots of a comparison between the four clusters of chromatin accessible regions as determined by ATAC-seq (Fig. 5a), with ATAC-seq data obtained from GMPs. Central line represents the median, hinges represent the 25th and 75th percentiles, and whiskers represent the lowest and highest values within 5×IQR (the interquartile range of the hinges). Cluster I, $n=525$; cluster II, $n=1,497$; cluster III, $n=2,489$; and cluster IV, $n=1,808$. (**e**) Microarray gene expression values for *Klf4*, *Tet2* and *p21* in fast- and slow-cycling GMPs. The data are from two biologically independent samples. (**f**) Representative screenshots of *Klf4* gene expression by RNA-seq for fast- and slow-cycling GMPs. (**g**) Nanog$^+$ iPSC colony counts after reprogramming of B cells depleted of *Tet2*. Error bars indicate s.d. ($n=3$ biological independent experiments). Statistical significance was determined using a two-tailed unpaired Student's *t*-test ($^{***}P<0.001$, $^{**}P<0.01$). (**h**) Nanog$^+$ iPSC colony counts after reprogramming of GMPs with a knockdown of Tet2. Error bars indicate s.d. ($n=3$ biological independent experiments). Statistical significance was determined using a two-tailed unpaired Student's *t*-test ($^{**}P<0.01$).

by preventing their ubiquitin-mediated degradation. In support of this possibility, proteasome inhibition has indeed been shown to increase the levels of Lsd1 and Hdac1 (refs 46,47). The finding that Lsd1 is required for the inactivation of B-SEs and the silencing of the B cell program suggests that it has a similar role as in ESCs, where it is responsible for the decommissioning of ESC-SEs during differentiation[31]. The observation that the Brd4 inhibitor JQ1 blocks reprogramming supports recent reports demonstrating that the Brd4–Med1–Cdk9 protein complex is required for reprogramming of MEFs (ref. 13) and for the maintenance of pluripotency of ESCs (ref. 12). That JQ1 also impaired reprogramming when used to treat B cells during the C/EBPα pulse was unexpected, but could be explained by the ability of C/EBPα to recruit Brd4 to ESC-SEs required for activation of pluripotency genes. Together, these results suggest that Brd4 is generally required for cell fate conversions and Lsd1 acts in a more tissue-restricted manner. In line with this interpretation, we found that JQ1 impairs the TFIID-enhanced reprogramming of MEFs into iPSCs (ref. 48), whereas S2101 exhibited a slight acceleration (Supplementary Fig. 7E).

As summarized in Fig. 7b, several lines of evidence help to explain why GMPs, like Bα′ cells, behave as elite cells for reprogramming: (i) GMPs express high levels of C/EBPα and strictly require the gene for their formation[42]; (ii) Klf4 is more highly expressed in GMPs than in B cells. In addition, the rapidly cycling GMP subpopulation with higher reprogramming efficiency express elevated levels of Klf4 (Fig. 6e,f); (iii) Tet2 is upregulated by C/EBPα, highly expressed in GMPs and required for the efficient reprogramming of both cell types[23] (Fig. 6g,h); (iv) GMPs specifically express a large set of differentiation-related genes shared with Bα′ cells[49] (Supplementary Fig. 6C,D); (v) GMPs share a substantial cohort of

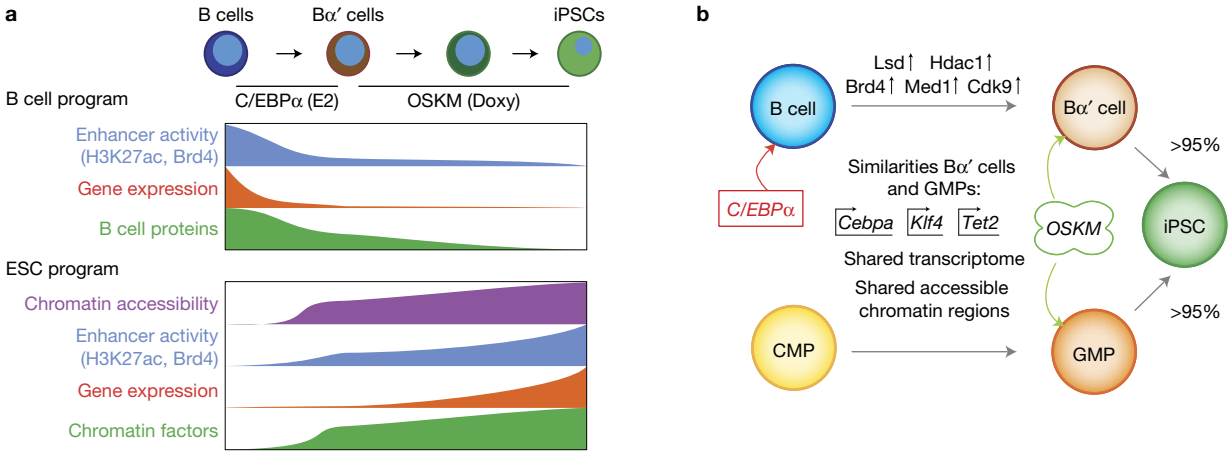


Figure 7 Summary diagrams. (**a**) A diagram summarizing the changes in enhancer activity, gene expression, proteome and chromatin accessibility described in this paper for both the B cell and the pluripotency gene expression program. (**b**) A diagram summarizing the main shared properties between Bα′ cells and GMPs, as elite cells for iPS reprogramming.

accessible chromatin regions (ATAC-seq cluster I) present in ESCs but not in B cells or MEFs.

Our data indicate that the interplay between C/EBPα and Klf4 induces chromatin accessibility to pluripotency factors as exemplified for the *Rarg* and *Lefty* enhancers where C/EBPα binds first followed by Klf4 (Fig. 5h). As described here and elsewhere[41], C/EBPα can act as a pioneer factor and Klf4 has also been shown to bind to closed chromatin[50]. However, it is possible that the predominant effect of C/EBPα is to activate *Klf4*, which subsequently binds to and activates other pluripotency genes (Fig. 5h).

Myeloid progenitors, including GMPs, also play an important role in blood cell malignancies because they have been shown to be key target cells in a mouse model for MLL–AF9-induced acute myeloid leukaemia[51] (AML). Strikingly, their oncogenic transformation strictly depends on the C/EBPα-driven differentiation to the GMP stage and/or on the transcriptional activity of C/EBPα (ref. 52), compatible with the notion that C/EBPα is responsible for chromatin remodelling events necessary for leukaemia[51–53]. It will now be interesting to determine whether the increase in the levels of chromatin-modifying factors described here also plays a role in AML formation, especially in view of the observation that human AML cells typically exhibit high levels of Lsd1 activity[54]. Our recent demonstration that C/EBPα can also poise human B cells for enhanced OSKM-induced reprogramming[55] underscores the factor's capacity to act across species.

Our findings have provided insights into the mechanism by which C/EBPα creates an elite cell state in B cells and the earliest events in reprogramming to iPSCs. It will now be interesting to determine whether upregulation of chromatin-related factors is a more general feature of somatic cell reprogramming, and whether this is also important for the GMP formation and their predisposition to malignancy. □

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Supplementary Information is available in the online version of the paper

ACKNOWLEDGEMENTS

We thank J. E. Bradner (Dana Farber Cancer Institute, Harvard Medical School, USA) for the JQ1 compound, J. Zuber (Research Institute of Molecular Pathology (IMP), Vienna Biocenter (VBC), Austria) for the Brd4 shRNA construct, R. Levine (Memorial Sloan Kettering Cancer Center, USA) for the Tet2 shRNA construct, L. de Andres for help with GMP isolations, L. Batlle for help with the chimaeric mice, the Genomics Facility and the Biomolecular Screening & Protein Technologies Facility of the CRG for technical assistance and members of the Graf laboratory for discussions. This work was supported by Ministerio de Educacion y Ciencia, SAF.2012-37167, Fundacio La Marató TV3 120410, AGAUR SGR 1136 and European Research Council Synergy Grant (4D-Genome). R.S. was supported by an EMBO Long-term Fellowship (ALTF 1201-2014) and a Marie Curie Individual Fellowship (H2020-MSCA-IF-2014). J.L.S. was supported by MINECO (IJCI-2014-21872). Work in the Porse laboratory was supported through a Centre grant from the NovoNordisk Foundation (Section for Stem Cell Biology in Human Disease and R179-A1513).

AUTHOR CONTRIBUTIONS

B.D.S. and T.G. conceived the study and wrote the manuscript. B.D.S. performed the cell culture, animal and molecular biology experiments and analysed the data. S.C., D.T. and M.F. performed the bioinformatics analysis, J.S.J. and B.P. the ChIP-seq experiments, J.L.S. and C.S.-M. the western blots and the protein immunoprecipitations, and A.L. the ATAC-seq experiments, M.W. and M.M. collected and analysed the proteomics data, R.S. performed and analysed the 4C-seq experiments; F.L. performed cell culture experiments.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://dx.doi.org/10.1038/ncb3326
Reprints and permissions information is available online at www.nature.com/reprints

1. Takahashi, K. *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131,** 861–872 (2007).
2. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126,** 663–676 (2006).
3. Okita, K., Ichisaka, T. & Yamanaka, S. Generation of germline-competent induced pluripotent stem cells. *Nature* **448,** 313–317 (2007).
4. Wernig, M. *et al. In vitro* reprogramming of fibroblasts into a pluripotent ES-cell-like state. *Nature* **448,** 318–324 (2007).
5. Maherali, N. *et al.* Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution. *Cell Stem Cell* **1,** 55–70 (2007).
6. Mikkelsen, T. S. *et al.* Dissecting direct reprogramming through integrative genomic analysis. *Nature* **454,** 49–55 (2008).
7. Brambrink, T. *et al.* Sequential expression of pluripotency markers during direct reprogramming of mouse somatic cells. *Cell Stem Cell* **2,** 151–159 (2008).
8. Stadtfeld, M., Maherali, N., Breault, D. T. & Hochedlinger, K. Defining molecular cornerstones during fibroblast to iPS cell reprogramming in mouse. *Cell Stem Cell* **2,** 230–240 (2008).

9. Li, R. *et al.* A mesenchymal-to-epithelial transition initiates and is required for the nuclear reprogramming of mouse fibroblasts. *Cell Stem Cell* **7,** 51–63 (2010).
10. Samavarchi-Tehrani, P. *et al.* Functional genomics reveals a BMP-driven mesenchymal-to-epithelial transition in the initiation of somatic cell reprogramming. *Cell Stem Cell* **7,** 64–77 (2010).
11. Hnisz, D. *et al.* Super-enhancers in the control of cell identity and disease. *Cell* **155,** 934–947 (2013).
12. Di Micco, R. *et al.* Control of embryonic stem cell identity by BRD4-dependent transcriptional elongation of super-enhancer-associated pluripotency genes. *Cell Rep.* **9,** 234–247 (2014).
13. Liu, L. *et al.* Transcriptional pause release is a rate-limiting step for somatic cell reprogramming. *Cell Stem Cell* **15,** 574–588 (2014).
14. Buganim, Y., Faddah, D. A. & Jaenisch, R. Mechanisms and models of somatic cell reprogramming. *Nat. Rev. Genet.* **14,** 427–439 (2013).
15. Wang, Y. *et al.* Embryonic stem cell-specific microRNAs regulate the G1-S transition and promote rapid proliferation. *Nat. Genet.* **40,** 1478–1483 (2008).
16. Krizhanovsky, V. & Lowe, S. W. Stem cells: the promises and perils of p53. *Nature* **460,** 1085–1086 (2009).
17. Chen, J. *et al.* H3K9 methylation is a barrier during somatic cell reprogramming into iPSCs. *Nat. Genet.* **45,** 34–42 (2013).
18. Liang, G., He, J. & Zhang, Y. Kdm2b promotes induced pluripotent stem cell generation by facilitating gene activation early in reprogramming. *Nat. Cell Biol.* **14,** 457–466 (2012).
19. Onder, T. T. *et al.* Chromatin-modifying enzymes as modulators of reprogramming. *Nature* **483,** 598–602 (2012).
20. Maherali, N. & Hochedlinger, K. Tgfβ signal inhibition cooperates in the induction of iPSCs and replaces Sox2 and cMyc. *Curr. Biol.* **19,** 1718–1723 (2009).
21. Korpal, M., Lee, E. S., Hu, G. & Kang, Y. The miR-200 family inhibits epithelial-mesenchymal transition and cancer cell migration by direct targeting of E-cadherin transcriptional repressors ZEB1 and ZEB2. *J. Biol. Chem.* **283,** 14910–14914 (2008).
22. Yamanaka, S. Elite and stochastic models for induced pluripotent stem cell generation. *Nature* **460,** 49–52 (2009).
23. Di Stefano, B. *et al.* C/EBPα poises B cells for rapid reprogramming into induced pluripotent stem cells. *Nature* **506,** 235–239 (2014).
24. Rais, Y. *et al.* Deterministic direct reprogramming of somatic cells to pluripotency. *Nature* **502,** 65–70 (2013).
25. Bar-Nur, O. *et al.* Small molecules facilitate rapid and synchronous iPSC generation. *Nat. Methods* **11,** 1170–1176 (2014).
26. Vidal, S. E., Amlani, B., Chen, T., Tsirigos, A. & Stadtfeld, M. Combinatorial modulation of signaling pathways reveals cell-type-specific requirements for highly efficient and synchronous iPSC reprogramming. *Stem Cell Rep.* **3,** 574–584 (2014).
27. Eminli, S. *et al.* Differentiation stage determines potential of hematopoietic cells for reprogramming into induced pluripotent stem cells. *Nat. Genet.* **41,** 968–976 (2009).
28. Guo, S. *et al.* Nonstochastic reprogramming from a privileged somatic cell state. *Cell* **156,** 649–662 (2014).
29. Ying, Q. L. *et al.* The ground state of embryonic stem cell self-renewal. *Nature* **453,** 519–523 (2008).
30. Shi, Y. *et al.* Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* **119,** 941–953 (2004).
31. Whyte, W. A. *et al.* Enhancer decommissioning by LSD1 during embryonic stem cell differentiation. *Nature* **482,** 221–225 (2012).
32. Wang, G. L. *et al.* HDAC1 cooperates with C/EBPα in the inhibition of liver proliferation in old mice. *J. Biol. Chem.* **283,** 26169–26178 (2008).
33. Grebien, F. *et al.* Pharmacological targeting of the Wdr5-MLL interaction in C/EBPα N-terminal leukemia. *Nat. Chem. Biol.* **11,** 571–578 (2015).
34. Adam, R. C. *et al.* Pioneer factors govern super-enhancer dynamics in stem cell plasticity and lineage choice. *Nature* **521,** 366–370 (2015).
35. Northrup, D. L. & Allman, D. Transcriptional regulation of early B cell development. *Immunol. Res.* **42,** 106–117 (2008).
36. Laurent, B. *et al.* A specific LSD1/KDM1A isoform regulates neuronal differentiation through H3K9 demethylation. *Mol. Cell* **57,** 957–970 (2015).
37. Loven, J. *et al.* Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell* **153,** 320–334 (2013).
38. Roe, J. S., Mercan, F., Rivera, K., Pappin, D. J. & Vakoc, C. R. BET bromodomain inhibition suppresses the function of hematopoietic transcription factors in acute myeloid leukemia. *Mol. Cell* **58,** 1028–1039 (2015).
39. Filippakopoulos, P. *et al.* Selective inhibition of BET bromodomains. *Nature* **468,** 1067–1073 (2010).
40. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10,** 1213–1218 (2013).
41. van Oevelen, C. *et al.* C/EBPa activates pre-existing and *de novo* macrophage enhancers during induced pre-B cell transdifferentiation and myelopoiesis. *Stem Cell Rep.* **5,** 232–247 (2015).
42. Zhang, D. E. *et al.* Absence of granulocyte colony-stimulating factor signaling and neutrophil development in CCAAT enhancer binding protein α-deficient mice. *Proc. Natl Acad. Sci. USA* **94,** 569–574 (1997).
43. Hasemann, M. S. *et al.* C/EBPα is required for long-term self-renewal and lineage priming of hematopoietic stem cells and for the maintenance of epigenetic configurations in multipotent progenitors. *PLoS Genet.* **10,** e1004079 (2014).
44. Maza, I. *et al.* Transient acquisition of pluripotency during somatic cell transdifferentiation with iPSC reprogramming factors. *Nat. Biotechnol.* **33,** 769–774 (2015).
45. Chen, J. *et al.* Vitamin C modulates TET1 function during somatic cell reprogramming. *Nat. Genet.* **45,** 1504–1509 (2013).
46. Canettieri, G. *et al.* Histone deacetylase and Cullin3-REN(KCTD11) ubiquitin ligase interplay regulates Hedgehog signalling through Gli acetylation. *Nat. Cell Biol.* **12,** 132–142 (2010).
47. Wu, Y. *et al.* The deubiquitinase USP28 stabilizes LSD1 and confers stem-cell-like traits to breast cancer cells. *Cell Rep.* **5,** 224–236 (2013).
48. Pijnappel, W. W. *et al.* A central role for TFIID in the pluripotent transcription circuitry. *Nature* **495,** 516–519 (2013).
49. Di Tullio, A. *et al.* CCAAT/enhancer binding protein α (C/EBP(α))-induced transdifferentiation of pre-B cells into macrophages involves no overt retrodifferentiation. *Proc. Natl Acad. Sci. USA* **108,** 17016–17021 (2011).
50. Soufi, A. *et al.* Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. *Cell* **161,** 555–568 (2015).
51. Ye, M. *et al.* Hematopoietic differentiation is required for initiation of acute myeloid leukemia. *Cell Stem Cell* **17,** 611–623 (2015).
52. Ohlsson, E. *et al.* Initiation of MLL-rearranged AML is dependent on C/EBPα. *J. Exp. Med.* **211,** 5–13 (2014).
53. Krivtsov, A. V. *et al.* Cell of origin determines clinically relevant subtypes of MLL-rearranged AML. *Leukemia* **27,** 852–860 (2013).
54. Harris, W. J. *et al.* The histone demethylase KDM1A sustains the oncogenic potential of MLL-AF9 leukemia stem cells. *Cancer Cell* **21,** 473–487 (2012).
55. Bueno, C. *et al.* Reprogramming human B-cells into induced pluripotent stem cells and its enhancement by C/EBPα. *Leukemia* http://dx.doi.org/10.1038/leu.2015.294 (2015).

DOI: 10.1038/ncb3326

METHODS

Mice. We used a cross between 'reprogrammable mice' containing a doxycycline-inducible OSKM cassette and tetracycline transactivator[56], and the Oct4–GFP reporter[57], as previously described[23] (Fig. 1a).

B cell isolation was routinely performed from 8- to 16-week-old male and female mice.

Mice were housed in standard cages under 12 h light–dark cycles and fed *ad libitum* with a standard chow diet. All experiments were approved by the Ethics Committee of the Barcelona Biomedical Research Park (PRBB) and performed according to Spanish and European legislation.

Cell cultures. ESCs (E14TG2) were cultured on gelatinized plates in N2B27 media (50% DMEM-F12, 50% Neurobasal medium, N2 (100×), B27 (50×)) supplemented with small-molecule inhibitors PD (1 μM, PD0325901) and CHIR (3 μM, CHIR99021), as well as LIF (10 ng ml^{-1}). CD19^{+} pre-B cells were obtained from bone marrow of reprogrammable mice with monoclonal antibody against CD19 (concentration of 0.2 μg per 10^6 cells; clone 1D3, BD Pharmingen, Cat. no. 553784) using MACS sorting (Miltenyi Biotech). Cell purity was confirmed by FACS using an LSR2 machine (BD) (>98%). After isolation, B cells were grown in RPMI medium supplemented with 10% FBS and 10 ng ml^{-1} IL-7 (Peprotech), 100× L-glutamine, 100× penicillin/streptomycin, 100× nonessential amino acids, 1,000× β-mercaptoethanol (Life Technologies). Lin^{-} c-Kit^{+} Sca-1^{-} CD16^{+}/CD32^{+} CD34^{+} GMP cells were isolated by FACS sorting using a BD INFLUX sorting machine and cultured in STEMSPAN medium (Stemcell Technologies) supplemented with 100 ng ml^{-1} SCF, 50 ng ml^{-1} IL3, 50 ng ml^{-1} Flt3L and 50 ng ml^{-1} mTPO (all from Peprotech). The following antibodies have been used for GMP isolations: eFluor780 anti-mouse c-Kit (0.5 μl for 5 × 10^6 cells; clone 2B8, eBioscience, Cat. no. 47-1171-80), PeCy7 anti-mouse Sca-1 (0.5 μl for 5 × 10^6 cells; clone D7, eBioscience, Cat. no. 25-5981-81), Alexa Fluor 647 rat anti-mouse CD34 (2 μl for 5 × 10^6 cells; clone RAM34, BD Pharmingen, Cat. no. 560230), anti-mouse CD16/CD32 FITC (1 μl for 5 × 10^6 cells; clone 93, eBioscience, Cat. no. 11-0161-82) and the lineage depletion kit (Miltenyi, Cat. no. 130-092-613). Mouse embryonic fibroblasts were isolated from E13.5 embryos of reprogrammable mice crossed with Oct4–GFP mice and cultured in DMEM supplemented with 10% FBS, 100× L-glutamine and 100× penicillin/streptomycin.

Reprogramming experiments. Pre-B cells were isolated from the bone marrow of reprogrammable mice crossed with Oct4–GFP mice, infected with C/EBPαER–hCD4 retrovirus, plated at 500 cells cm^{-2} in gelatinized plates (12 wells) on irradiated MEF feeders in RPMI medium. To activate C/EBPα, cultures were treated for 18 h with 100 nM β-oestradiol (E2), resulting in Bα′ cells.

After E2 washout, the cultures were switched to serum-free N2B27 medium supplemented with IL-4 10 ng ml^{-1}, IL-7 10 ng ml^{-1} and IL-15 2 ng ml^{-1}. To activate OSKM, the cultures were treated with 2 μg ml^{-1} of doxycycline. From day 2, onwards we supplemented the N2B27 medium with 20% KSR (Life Technologies), 3 μM CHIR99021 and 1 μM PD0325901.

OSKM Oct4–GFP MEFs were seeded in gelatinized plates and induced with 2 μg ml^{-1} of doxycycline in ES medium.

A step-by-step protocol describing the reprogramming procedure can be found at Nature Protocol Exchange[58]. All cell lines have been routinely tested for mycoplasma contamination.

Stable cell lines. The PlatE retroviral packaging cell line was obtained from Cell Biolabs (Cat. no. RV-101). The C10 pre-B cell line stably expressing C/EBPαER–GFP has been described previously[59].

Immunofluorescence. For Nanog staining, the cells were fixed with 4% paraformaldehyde, blocked and incubated with the primary antibody (1:500; Calbiochem, SC100) overnight at 4 °C. They were then stained with secondary anti-rabbit antibodies (1:1,000; Alexa Fluor 547, Life Technologies) at room temperature for one hour. Nuclear staining was performed with DAPI (Invitrogen).

RNA extraction. To remove the feeders, cells were trypsinized and pre-plated for 30 min before RNA isolation with the miRNeasy mini kit (Qiagen). RNA was eluted from the columns using RNase-free water and quantified by Nanodrop. cDNA was produced with the High Capacity RNA-to-cDNA kit (Applied Biosystems).

FACS analyses. Oct4–GFP expression was analysed with an LSR II FACS (BD Biosciences) using Diva v6.1.2 (BD Biosciences) and FlowJo software v10 (TreeStar).

Isolation of fast- and slow-cycling cells. CFSE (Life Technologies) was used to subdivide GMPs into fast- and slow-cycling cell populations. 3 × 10^5 cells were resuspended in 1 ml PBS with 0.2 μM of CSFE for 5 min and then rinsed twice with four volumes of PBS. After 24 h, cells were sorted on the basis of FITC brightness as indicated in Supplementary Fig. 8A.

Gene expression arrays. RNA samples with a RIN > 9 were subjected to transcriptional analyses using Agilent expression arrays. For hybridization, 500 ng of total RNAs were labelled using Agilent's QuickAmp labelling kit and hybridized to Agilent 8X60K expression arrays.

qRT-PCR analyses. qRT-PCR reactions were set up in triplicate with the SYBR Green QPCR Master Mix (Applied Biosystems) and primers as listed in Supplementary Table 3. Reactions were run on an AB7900HT PCR machine with 40 cycles of 30 s at 95 °C, 30 s at 60 °C and 30 s at 72 °C.

Small-molecule inhibitors. Lsd1 inhibition was achieved by treating the cells with 50 μM S2101 (Calbiochem). Brd4 was inhibited with 100 nM JQ1 obtained from J. E. Bradner. Hdac1 inhibition was achieved by culturing the cells in the presence of 1 mM of valproic acid.

BrdU staining. To check cell proliferation, BrdU was added to the culture medium at a concentration of 10 μM for 6 h. Staining was performed using the BrdU APC Flow Kit from BD Pharmingen.

Cell viability. Cell viability was assessed using the Pacific Blue Annexin V/SYTOX AADvanced Apoptosis Kit for flow cytometry from Thermo Fisher Scientific.

Luciferase assay. The −90 kb and −280 kb enhancers of *Klf4* were amplified by qPCR using the following primers: −90 kb forward 5′-GCCCTCGAGCGGGTCTGGCCTTCAGTGATA-3′ and reverse 5′-GCCATGCATCCGGACTCCCTTTTGCTAGTG-3′; −280 kb forward 5′-GCCCTCGAGCTGGTATATGCACACATGCAC-3′ and reverse 5′-GCCCTGCAGCTCTCCCTGCATTGGCTTAGT-3′.

The PCR fragments were digested using either XhoI and PstI or XhoI and NsiI and cloned in the CSI–LUC2–minP vector in the XhoI and PstI sites.

Twelve-well plates with 293T cells were transfected with 0.1 μg of a β-galactosidase control plasmid, 0.5 μg of the respective enhancer plasmid or the empty vector and 0.5 μg of C/EBPα, or 0.5 μg of C/EBPα and PU.1 plasmids or 0.5 μg of GFP plasmid using the TransIT transfection reagent (Mirus). Luciferase and β-galactosidase activities were measured using the Luciferase Assay System (Promega) and the β-Galactosidase Enzyme Assay System (Promega) following the manufacturer's protocol. All experiments were carried out in triplicates and normalized to β-galactosidase activity to account for differences in transfection efficiencies.

Vectors and virus production and infection. The retroviral vectors LMN–Ctrl-shRNA–PGK–GFP and LMN–Lsd1-shRNA–PGK–GFP were purchased from Transomic Technologies. The retroviral vector containing the Brd4 shRNA (no. 552) has been described previously[60]. The pMX vectors of TAF4, TAF5, TAF6 and TBP (ref. 48) were obtained from Addgene. The retroviral vector Tet2-shRNA–PGK–GFP has been described previously[61]. The C/EBPαER–hCD4 retroviral vector has been used before[62]. Viral production and B cell infection were performed as described previously[62]. For the infection of GMPs, freshly sorted cells were infected by centrifugation for 90 min on Retronectin (Clontech)-coated plates.

4C. 4C-seq was performed as described previously[63]. Briefly, 0.5–1.0 million crosslinked nuclei were digested with Csp6I followed by ligation under dilute conditions. After decrosslinking and DNA purification, samples were digested overnight with DpnII and once more ligated under dilute conditions. Column-purified DNA was directly used as input for inverse PCR using primers (Supplementary Table 3) with Illumina adapter sequences as overhangs. Several PCR reactions were pooled, purified and sequenced on an Illumina HiSeq 2500.

RNA-seq. Libraries were prepared using the SMARTer Universal Low Input RNA Kit (Clontech Laboratories) according to the manufacturer's protocol. The libraries were analysed using Agilent DNA 1000 chip to determine the quantity and size distribution and then quantified by qPCR using the KAPA Library Quantification Kit (ref. KK4835, KapaBiosystems) before amplification with Illumina's cBot. Libraries were loaded at a concentration of 10 pM onto the flow cell and sequenced on Illumina's HiSeq 2000.

Chromatin immunoprecipitation (ChIP). The H3K4me2 ChIPs were performed using the True MicroChIP kit from Diagenode following the manufacturer instructions. Low-cell-number ChIP-seq experiments were performed essentially as previously described[64], with varying RIPA-buffer washes. Briefly, cells were fixed in 1% formaldehyde in 50% PBS/50% culture medium, with rotation for 10 min at room temperature. For Brd4 ChIP, cells were fixed with 2 mM DSG in PBS for

20 min, followed by addition of 1% formaldehyde, 10 min, both steps with rotation at room temperature. Chromatin was sheared to 200-bp fragments using a Bioruptor sonicator and 0.5 ml siliconized tubes. H3K27ac-carrying histones were precipitated using four low-salt RIPA-buffer washes and protein A Sepharose beads, C/EBPα and Brd4 using six low-salt RIPA-buffer washes and protein G and protein A beads respectively, and Klf4 with two low- and two high-salt RIPA washes and protein G beads. See below for the references of the antibodies used.

Protein fractionation. Gel-filtration experiments were performed in the Biomolecular Screening & Protein Technologies Facility using a Superose 6 10/300 GL column (GE Healthcare). For analysis of endogenous Lsd1–Hdac1–C/EBPα complex, Bα′ cells were lysed in RIPA buffer, and 1 mg of cleared lysate was loaded onto a single Superose 6 10/300 GL column, calibrated using a mixture of molecular mass marker proteins (MWGF1000, Sigma-Aldrich). A 300 µl portion of lysate was loaded onto the column and collected into 250 µl fractions; fractions were processed for western blot analysis.

Co-immunoprecipitation experiments and western blots. Owing to the large number of cells required, the immunoprecipitation experiments were performed in an inducible, C/EBPαER-containing C10 pre-B cell line[59]; the western blots were performed with primary cells. Cell lysates were prepared using NET-2 buffer (200 mM NaCl, 50 mM Tris pH 7,5, 0.1% Triton and 1× protease inhibitors) for 20 min at 4 °C and centrifugated for 10 min at 16,000*g* to remove cellular debris. Supernatants were used for protein detection by western blot or for co-immunoprecipitation. For the latter 19/20 of the extract was pre-cleared with magnetic beads (Dynabeads, Invitrogen) by rotating for 2 h on a wheel at 4 °C. One-twentieth of the lysates was kept as the input. Separately, to crosslink the antibodies to beads, 50 µl of beads were incubated with 3 µg of antibody or IgG for 2 h at room temperature on a wheel and then washed once with PBS and twice with 0.2 M triethanolamine pH 8.2. The suspension was then incubated for 20 min at room temperature with 20 mM dimethyl pimelimidate, the reaction stopped with two 5 min washes with 50 mM Tris pH 8 followed by three washes with PBS and the beads incubated with 0.1 M citric acid pH 3 for 2 min to remove non-crosslinked antibodies followed by two washes with lysis buffer. Finally, the crosslinked antibodies were incubated with the pre-cleared protein extracts overnight at 4 °C on a wheel and the beads selected with a magnet, 1/20th of the supernatant kept as the unbound control fraction and the beads washed six times with lysis buffer. To elute, the beads were resuspended in 1× Laemmli buffer (without βME) and heated for 20 min at 60 °C. Magnetic beads were then separated with a magnet, the supernatant complemented with 5% βME and boiled before loading in an SDS–PAGE acrylamide gel to analyse bound proteins by western blotting. As controls, one lane was loaded with the original input, another lane with the unbound fraction and a third lane with beads coupled to either mouse or rabbit IgG. Finally, after transferring the proteins to a membrane, the blot was blocked with 5% milk in TBS–Tween, probed with an antibody to the homologous protein, and after developing stripped and probed with an antibody against the antigen tested for interaction.

C/EBPα immunoprecipitation for mass spectrometry. For affinity purification experiments of C/EBPα, we used our inducible, C/EBPαER-containing C10 pre-B cell line[59]. Immunoprecipitations were performed using three separate samples with antibody (Santa Cruz, SC-61) and three samples with IgG essentially as above but without pre-clearing and without crosslinking the antibodies to the beads. After overnight incubation with extracts, beads were washed three times with NET-1.5 buffer (150 mM NaCl, 50 mM Tris pH 7.5, 0.1% Triton) and twice with TBS. Precipitated proteins were eluted with elution buffer (6 M urea, 2 M thiourea, 10 mM Hepes pH 8.0) containing 2 mM dithiothreitol for 30 min at room temperature, followed by a second elution with elution buffer. Both elutes were combined, alkylated with 5 mM CAA, diluted 5 times with 50 mM ammonium bicarbonate and digested with 1 µg trypsin overnight at room temperature. The resulting peptide mixtures were acidified with 1% TFA, desalted with StageTips containing three layers of C18 material and analysed by mass spectrometry as described below.

Antibodies for ChIPs, western blots and immunoprecipitations. The antibodies used in this study were H3K4me2 (2 µg for 25 µg of chromatin) (Abcam, ab7766), Lsd1 (1:1,000) (Kdm1a, Abcam, ab17721), Brd4 (ChIP: 2.5 µg ml^{-1}) (Bethyl Laboratories, A301-985A); (western blot: 1:1,000) (Abcam, ab128874), H3K27ac (1 µg ml^{-1}) (Abcam, ab4729), Klf4 (1:200) (R&D, AF3158), Hdac1 (1:2,000) (Abcam, ab7028), C/EBPα (western blot: 1:300) (Santa Cruz, SC-61); (ChIP: 0.2 µg ml^{-1}) (Santa Cruz, SC-61), PU.1 (0.5 µg ml^{-1}) (Santa Cruz, SC-352X), ERα antibody (1:500) (Santa Cruz, SC-543), H3 (1:2,000) (Abcam, ab1791), Parp (1:1,000) (no. 9542, Cell Signaling), β-tubulin (1:5,000) (Sigma, T7816), GAPDH (1:1,000) (Santa Cruz, SC-32233), PCNA (1:400) (Santa Cruz, SC-56 (PC10)), Cdk9 (1:1,000) (Cell Signaling, rabbit mAb no. 2316), Oct4 (1:200) (Santa Cruz, SC-8628), Gdf3 (1.5 µg ml^{-1}) (R&D, AF958), Lin28a (1:1,000) (Cell Signaling, no. 8641), Sall4 (1 µg ml^{-1}) (Abcam, ab57577), Nanog (1:500) (Calbiochem, SC1000), Sox2 (1:1,000) (Santa Cruz, SC-17320), Tcfcp2l1 (1 µg ml^{-1}) (R&D, AF5726), hCD4 (0.5 µl per 10^6 cells) (clone RPA-T4, eBioscience, 13-0049-80), β-catenin (1:1,000) (BD Biosciences, 610153).

Chimaeric mice. For the chimaera formation assay, 10–15 iPS cells (Agouti colour coat) were injected into 3.5 dpc blastocysts of C57BL/6-Albino mice (white coat colour). Chimaerism of the transplanted offspring was assessed by the presence of agouti coat colour derived from the iPS cells.

Whole-cell proteomic analysis. Sample preparation for mass spectrometry was performed essentially as described previously[65]. Cells were lysed in lysis buffer (6 M guanidinium chloride, 10 mM TCEP), incubated for 10 min at 95 °C and sonicated for 15 min using a Bioruptor. Proteins were alkylated with 40 mM 2-chloroacetamide, diluted tenfold with digestion buffer (20 mM Tris-HCl, pH 8.5 and 10% ACN) and digested with 1:50 (w/w) Lys-C (Wako) and 1:50 (w/w) trypsin (Promega) at 37 °C overnight. The resulting peptide mixture was acidified by addition of 1% TFA and desalted on StageTips with three layers of SDB-RPS. Peptides were separated on 50-cm columns of ReproSil-Pur C18-AQ 1.9 µm resin (Dr. Maisch GmbH) packed in-house. Liquid chromatography was performed on an EASY-nLC 1000 ultrahigh-pressure system coupled through a nanoelectrospray source to a Q-Exactive mass spectrometer (all from Thermo Fisher Scientific). Peptides were loaded in buffer A (0.1% formic acid) and separated applying a nonlinear gradient of 5–60% buffer B (0.1% formic acid, 80% acetonitrile) at a flow rate of 250 nl min^{-1} over 240 min. Data acquisition switched between a full scan and five data-dependent MS/MS scans. Multiple sequencing of peptides was minimized by excluding the selected peptide candidates for 45 s. For AP-MS experiments, a linear gradient of 5–30% buffer B was applied over 70 min, with dynamic exclusion set to 15 s.

ATAC-seq. ATAC-seq was performed as previously described[40]. A total of 100,000 cells were washed once with 100 µl PBS and resuspended in 50 µl lysis buffer (10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, 0.2% IGEPAL CA-630). The suspension of nuclei was then centrifuged for 10 min at 500*g* at 4 °C, followed by the addition of 50 µl transposition reaction mix (25 µl TD buffer, 2.5 µl Tn5 transposase and 22.5 µl nuclease-free H_2O) and incubation at 37 °C for 45 min. DNA was isolated using the MinElute Kit (Qiagen). Library amplification was done by two sequential PCR reactions (8 and 5 cycles, respectively). After the first PCR reaction, the library was selected for fragments below 700 bp with AmpureXP beads followed by a second PCR reaction. Libraries were purified with Qiaquick PCR (Qiagen) and integrity checked on a Bioanalyser before sequencing.

Bioinformatic analyses. All sequencing data were mapped onto the mouse genome assembly mm10 (Ensembl GRCm38.78) using STAR (-outFilterMultimapNmax 1 –outFilterMismatchNmax 999 –outFilterMismatchNoverLmax 0.06; for RNA-seq: -sjdbOverhang 100 –outFilterType BySJout -alignSJoverhangMin 8 -alignSJDBoverhangMin 1 -alignIntronMin 20 -alignIntronMax 1000000 -alignMatesGapMax 1000000; for genomic sequencing: -alignIntronMax 1 -alignEndsType EndToEnd -alignMatesGapMax 2000) and analysed with R (3.1.0) using packages from the bioconductor suite (v3.0). For peak calling, regions overlapping the 'Encode blacklist' were removed. All clusterings were performed using the Rpackage Mfuzz (2.26.0). All GO enrichment analyses were performed using the Rpackage ReactomePA (1.10.1).

For RNA-seq, mapping was performed using Ensembl annotation (GRCm38.78). Genes expression quantification was performed using the Rpackage Rsubread (1.16.1). Sample scaling and statistical analysis were performed using the Rpackage DESeq2 (1.6.3) and vsd counts were used for further analysis. Genes changing significantly at any time point were identified using the nbinomLRT test (FDR < 0.01) and for >2-fold change between at least two time points (average of replicates, vsd values). Genes upregulated between two time points were selected using the nbinomWaldTest (FDR < 0.01) and for >1.5-fold change.

For ATAC-seq, duplicates reads were removed using Picard (http://picard.sourceforge.net) (MarkDuplicates, REMOVE_DUPLICATES=true). Bigwig tracks were made using DeepTools BamCoverage (1.5.9.1). Peak calling was performed using macs2 (2.1.0.20140616) (-f BAMPE -g mm -p 0.001 -nolambda). For quantitative analyses, peaks from all time points were merged as one set of non-overlapping regions using Bedops. Reads were counted on merged regions for each time point, using the Rpackage csaw (1.0.7). Counts were scaled on genome-wide fragments using the Rpackage csaw (function WindowCounts on 10 kb windows), and size factors were calculated using DESeq2. Scaled counts were transformed to a $\log_2$ scale (with pseudo-count of 1). Regions becoming more accessible in Bα′ cells were identified as showing a >1.5-fold increase after pulse and a difference of more than 10 fragments, and a difference of >2-fold changes and >40 fragments between any of the time points. Association of peaks with genes was performed with the Rpackage ChipPeakAnno, using Ensembl transcripts (GRCm38.78).

DOI: 10.1038/ncb3326

Motif analyses were performed using RSAT. Repeats from the Ensembl annotation (GRCm38.75) were masked using bedtool maskfasta. Motif discovery was performed using peaks-motif and compared with the databases JASPAR (v.2015.03) and HOCOMOCO (v9). Motif enrichment in ATAC-seq clusters was performed with matrixQuality. We used as background a Markov model of order 1. The maximum-normalized-weight-difference score (MNWD) of matrix quality was used as enrichment score.

For the ChIP-seq, duplicate reads were removed using picard. Bigwig tracks were made using DeepTools BamCompare to subtract the input from the ChIP (-scaleFactorsMethod SES -ratio subtract -fragmentLength 200). H3K27ac peak calling was performed using macs2 on immunoprecipitation versus input (-broad -q 0.01, -broad-cutoff 0.01). Super-enhancers were called using ROSE (ref. 66) on H3K27ac data (-t 2500). Quantitative analyses were performed as for ATAC-seq. Average plots were obtained using DeepTools computeMatrix and profilers.

For proteomic, raw mass-spectrometry data were analysed with the MaxQuant workflow (1.5.1.6 and 1.5.3.29). Peak lists were searched against the Uniprot mouse FASTA database (2013_05) combined with 262 common contaminants by the integrated Andromeda search engine. FDR was set to 1% for both peptides (minimum length of 7 amino acids) and proteins. 'Match between runs' (MBR) with a maximum time difference of 0.7 min was enabled. Relative protein amounts were determined by the MaxLFQ algorithm, with a minimum ratio count of two peptides. Missing values were input from a normal distribution using the Perseus software package (width = 0.2, downshift = 1.8 s.d.).

For AP-MS, Student's *t*-test was applied for detection of differentially enriched proteins between triplicate immunoprecipitations of specific antibody and control IgG. For total protein analysis, proteins changing significantly at any time point were identified using an LRT test (FDR < 0.1) from the Rpackage msmsTests. Proteins changing between two time points were selected for 1.5>-fold change between each replicate (all four comparisons).

For the 4C analysis, the sequence of the reading primer was trimmed from the 5′ of reads using the demultiplex.py script from the R package fourCseq (version 1.0.0, allowing 4 mismatches). Reads in which this sequence could not be found were discarded. Reads were mapped using STAR and processed using fourCseq to filter out reads not located at the end of a valid fragment and to count reads per fragment. Tracks of signal were made after smoothing the RPKM counts per fragment with a running mean over five fragments.

Independent component analysis (ICA) and canonical correlation analysis (CCA) were performed on gene expression data using the R packages fastICA and CCA, respectively.

All codes for bioinformatic analysis are available on request.

Accession numbers. Microarray data obtained after reprogramming of B cells in FBS+LIF conditions[23]: GSE52397. H3K27ac ChIP-seq in ESC (ref. 67): GSE62380; Oct4, Nanog and Sox2 (ref. 66): GSE44286; Klf4 (ref. 68): GSE11431; Brd4: GSE36561. C/EBPα in GMPs: GSE43007 (ref. 43). ATAC-seq in MEFs (ref. 44): GSE67298. RNA-seq in different haematopoietic cell types[69]: GSE60101.

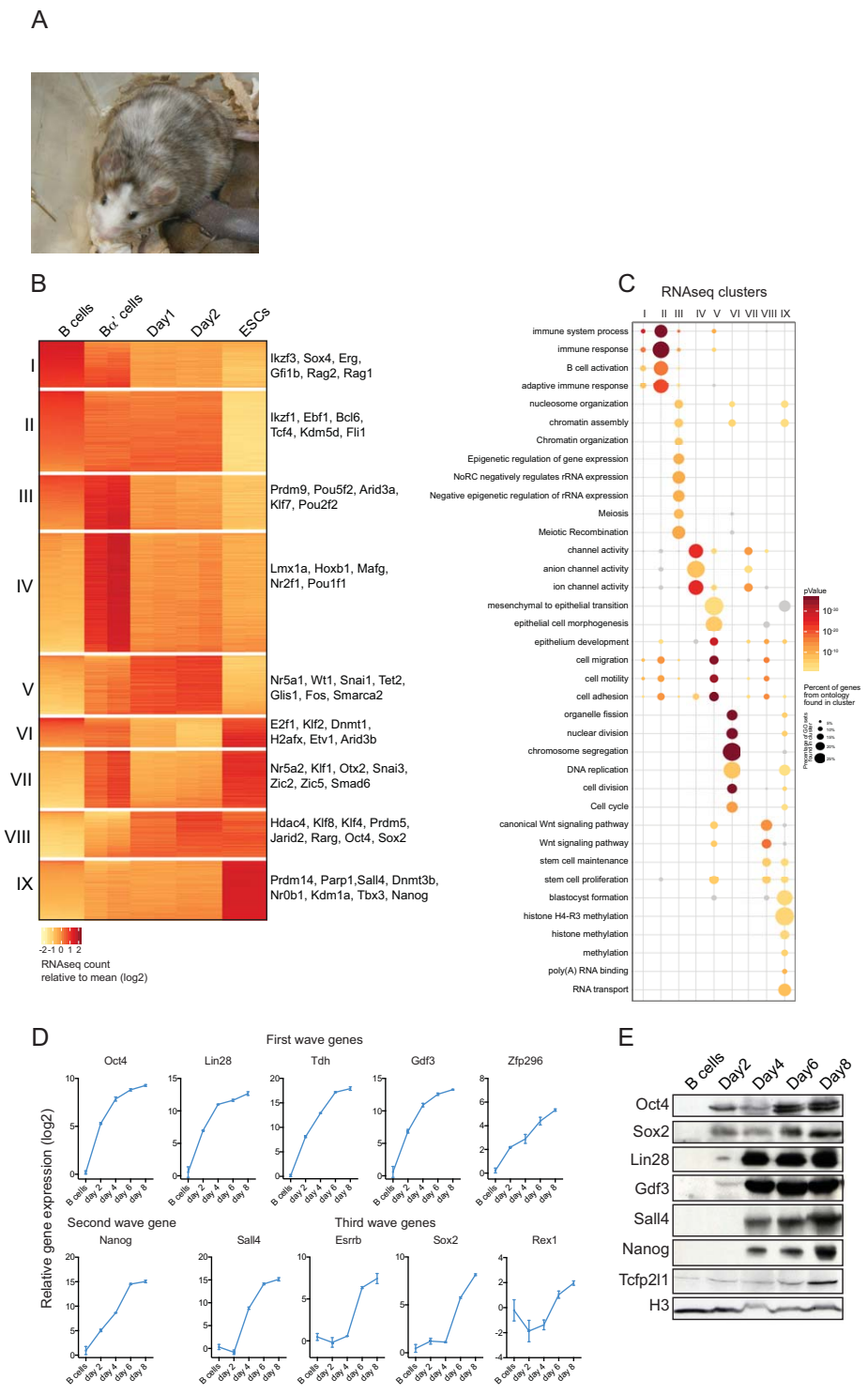
The RNA-seq, ChIP-seq and ATAC-seq data obtained for B cells, Bα′ cells, day 1 cells, day 2 cells and ESCs described in this paper have been deposited in GEO under the number GSE71218. Proteomic data have been deposited in proteome Xchange under the identifier PXD002769.

Statistics and reproducibility. Western blot, immunoprecipitation, ChIP, qRT-PCR, immunofluorescence, immunoprecipitation mass spectrometry, alkaline phosphate and FACS data presented are representative of at least three independent experiments that yielded similar results; microarray, RNA-seq and proteomic data were obtained from two independent experiments that yielded similar results. Statistical analyses were performed using Prism software (GraphPad). No statistical method was used to predetermine sample size and the experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

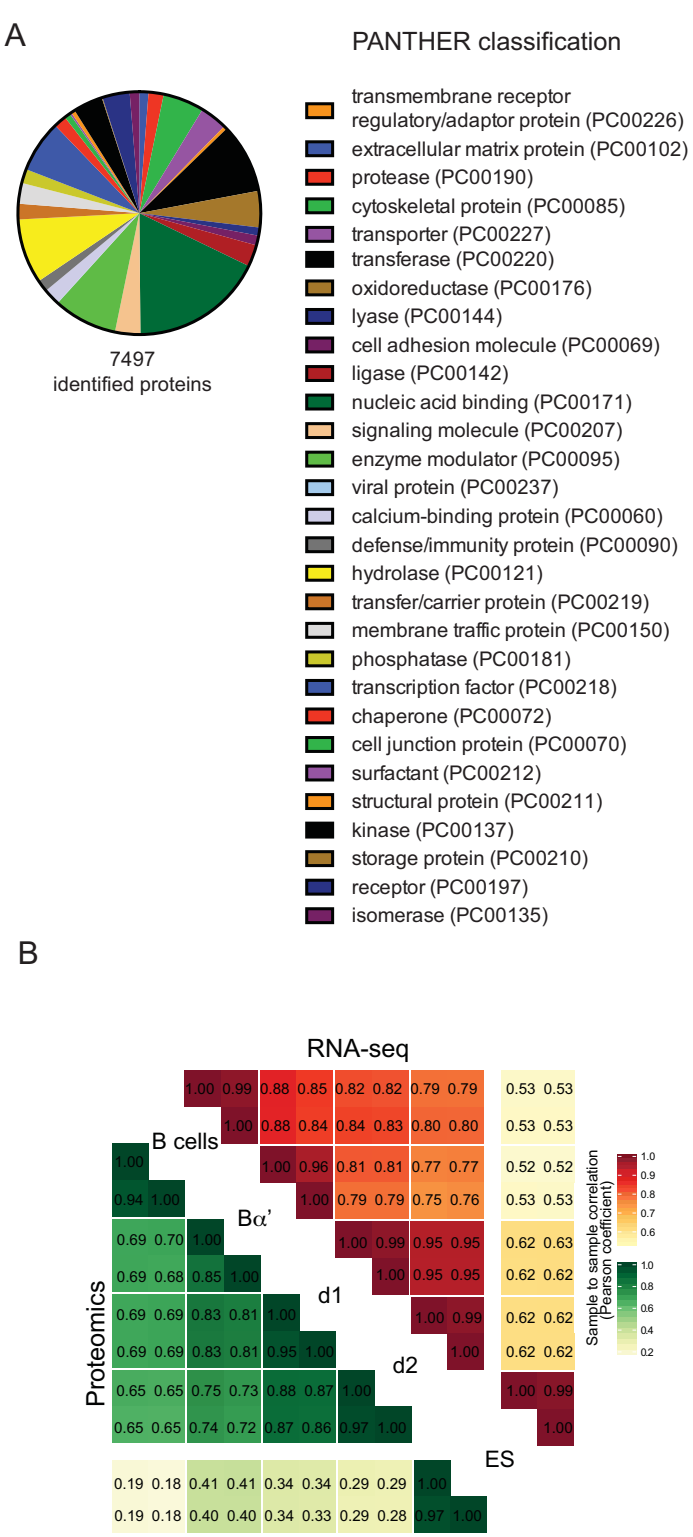
56. Carey, B. W., Markoulaki, S., Beard, C., Hanna, J. & Jaenisch, R. Single-gene transgenic mouse strains for reprogramming adult somatic cells. *Nat. Methods* **7,** 56–59 (2010).
57. Boiani, M., Eckardt, S., Scholer, H. R. & McLaughlin, K. J. Oct4 distribution and level in mouse clones: consequences for pluripotency. *Genes Dev.* **16,** 1209–1219 (2002).
58. Di Stefano, B. & Graf, T. Rapid generation of induced pluripotent stem cells from mouse pre-B cells. *Nat. Protoc. Exch.* http://dx.doi.org/10.1038/protex.2016.001 (2016).
59. Bussmann, L. H. *et al.* A robust and highly efficient immune cell reprogramming system. *Cell Stem Cell* **5,** 554–566 (2009).
60. Zuber, J. *et al.* RNAi screen identifies Brd4 as a therapeutic target in acute myeloid leukaemia. *Nature* **478,** 524–528 (2011).
61. Figueroa, M. E. *et al.* Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. *Cancer Cell* **18,** 553–567 (2010).
62. Di Stefano, B. & Graf, T. Very rapid and efficient generation of induced pluripotent stem cells from mouse pre-B cells. *Methods Mol. Biol.* **1357,** 45–56 (2014).
63. Stadhouders, R. *et al.* Multiplexed chromosome conformation capture sequencing for rapid genome-scale high-resolution detection of long-range chromatin interactions. *Nat. Protoc.* **8,** 509–524 (2013).
64. Jakobsen, J. S. *et al.* Amplification of pico-scale DNA mediated by bacterial carrier DNA for small-cell-number transcription factor ChIP-seq. *BMC Genomics* **16,** 46 (2015).
65. Kulak, N. A., Pichler, G., Paron, I., Nagaraj, N. & Mann, M. Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nat. Methods* **11,** 319–324 (2014).
66. Whyte, W. A. *et al.* Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153,** 307–319 (2013).
67. Ji, X. *et al.* Chromatin proteomic profiling reveals novel proteins associated with histone-marked genomic regions. *Proc. Natl Acad. Sci. USA* **112,** 3841–3846 (2015).
68. Chen, X., Xu, H., Yuan, P., Fang, F. *et al.* Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* **133,** 1106–1117 (2008).
69. Lara-Astiaso, D. *et al.* Immunogenetics. Chromatin state dynamics during blood formation. *Science* **345,** 943–949 (2014).

nature cell biology

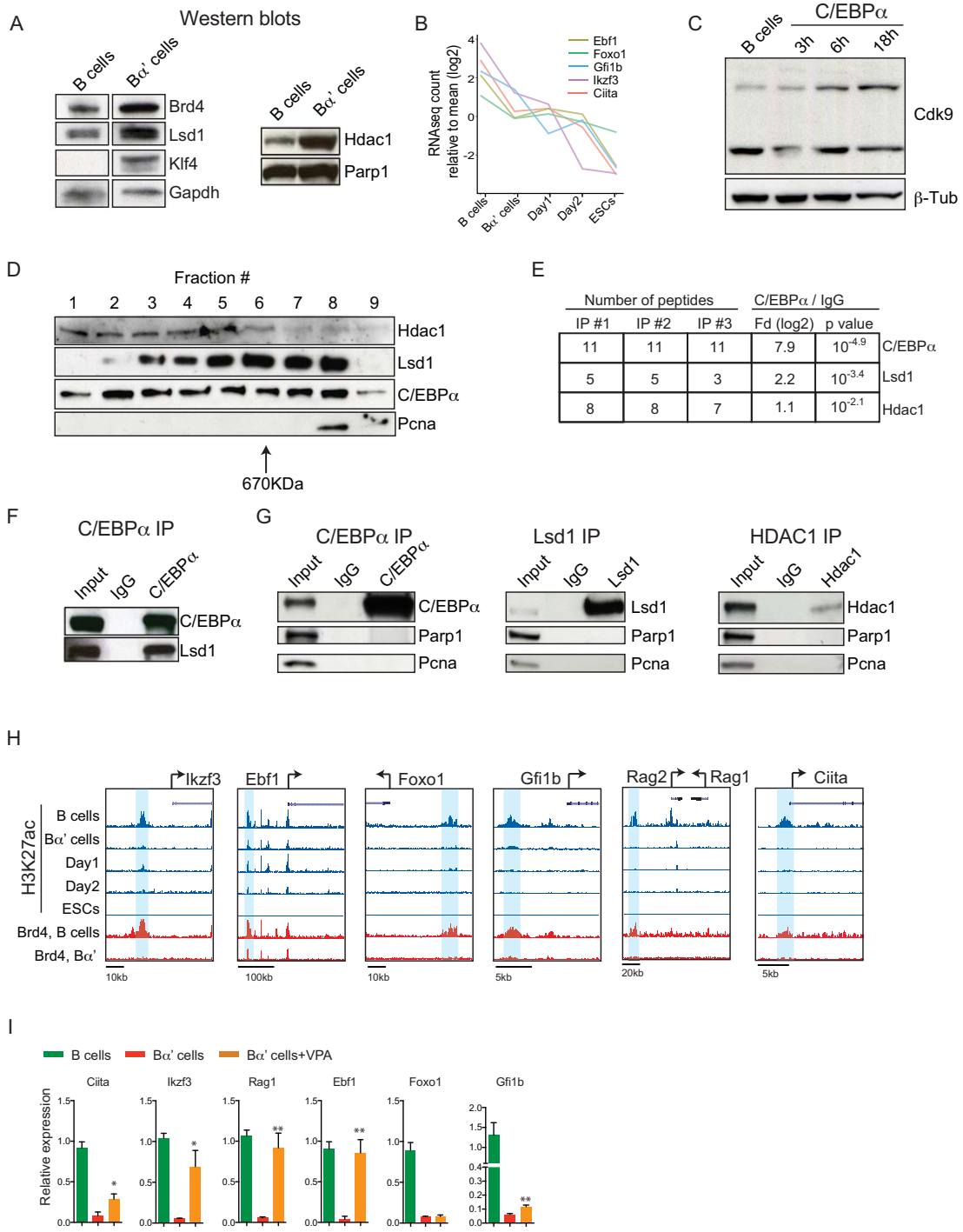
DOI: 10.1038/ncb3326



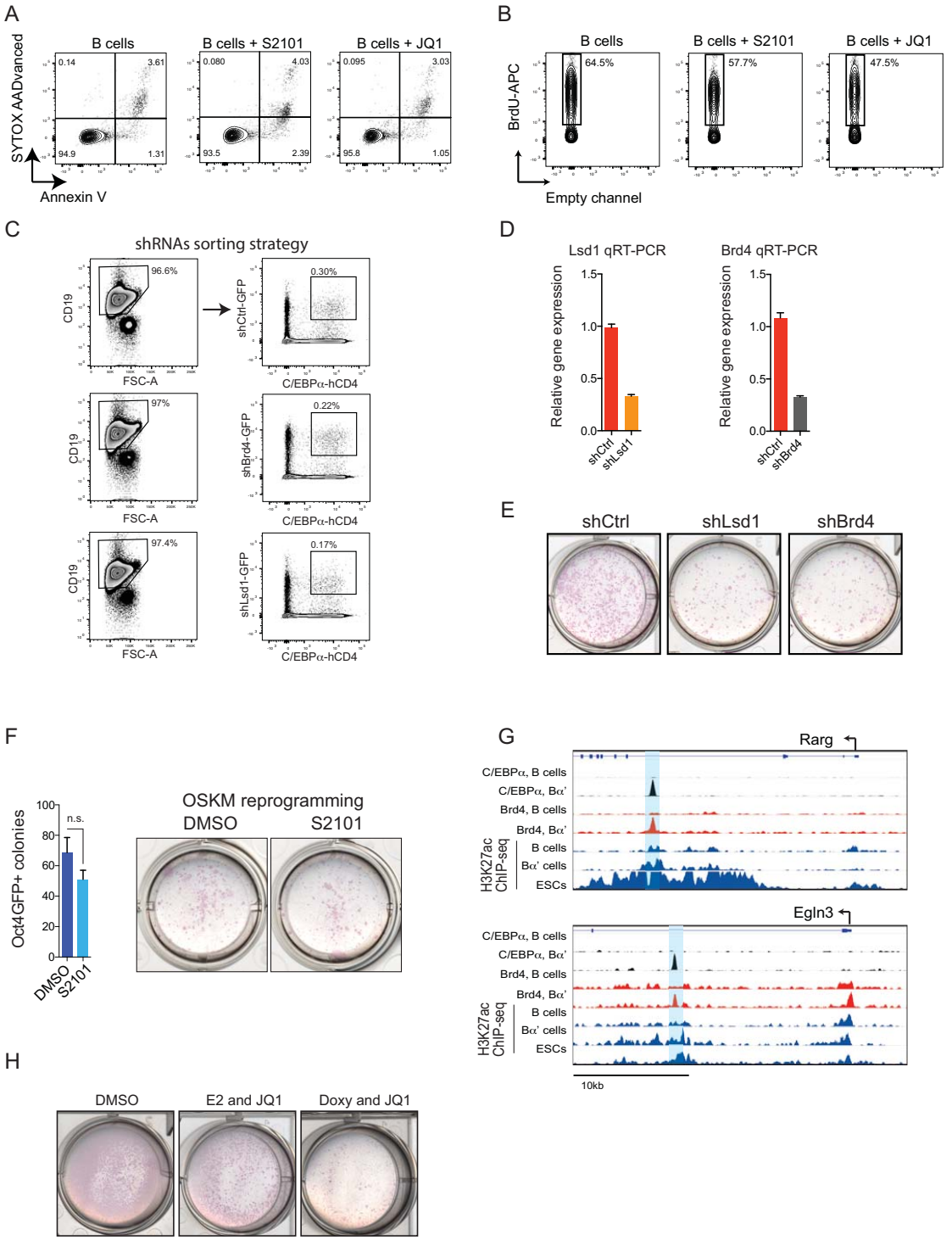
Supplementary Figure 1 Characterization of Bα' cell reprogramming into iPS cells. (A) Representative chimeric mouse obtained after blastocyst injection of aiPS clone. (B) Heatmap of RNA-seq data showing genes changing >2fold during reprogramming (FDR<1%, LRT test). (C) Gene Ontology (GO) analysis of protein clusters shown in panel A. The size of each circle represents the proportion of GO sets found in each cluster; the intensity of the color represents the *P*-value, determined by a hypergeometric test. (D) Gene expression (qRT-PCR) of selected pluripotency genes. Values were normalized against *Pgk* expression. Error bars indicate s.d. (n=3 biologically independent samples). (E) Representative western blots for selected pluripotency transcription factors. See Suppl. Fig. 8 for uncut gel images.



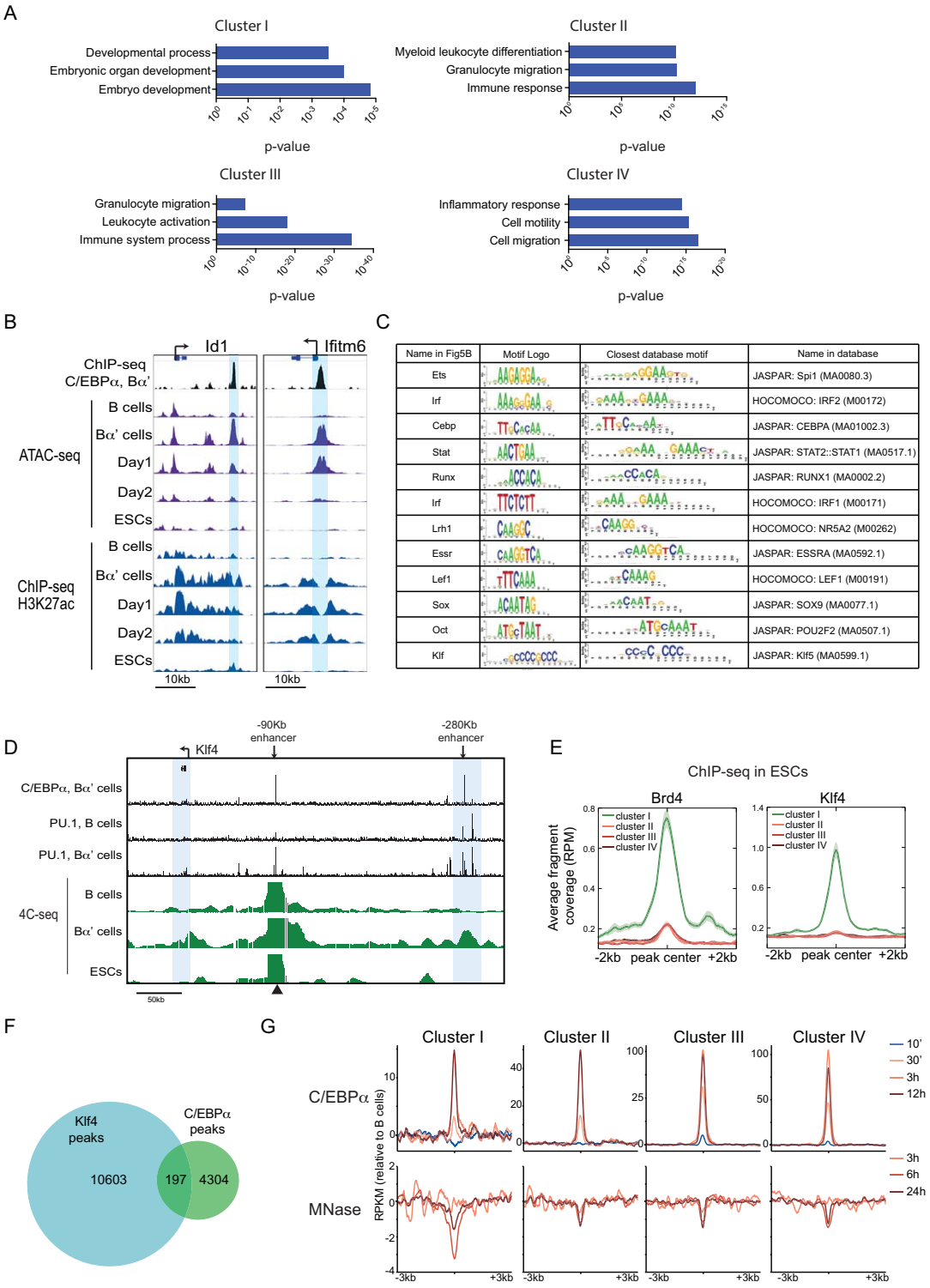
Supplementary Figure 2 Protein dynamics during reprogramming. (A) PANTHER classification for all the proteins identified by mass spectrometry in the samples tested. (B) Correlation between biological duplicates of RNA-seq and proteomic data. (C) C-means clustering of proteins changing >2 fold at any time points during reprogramming.



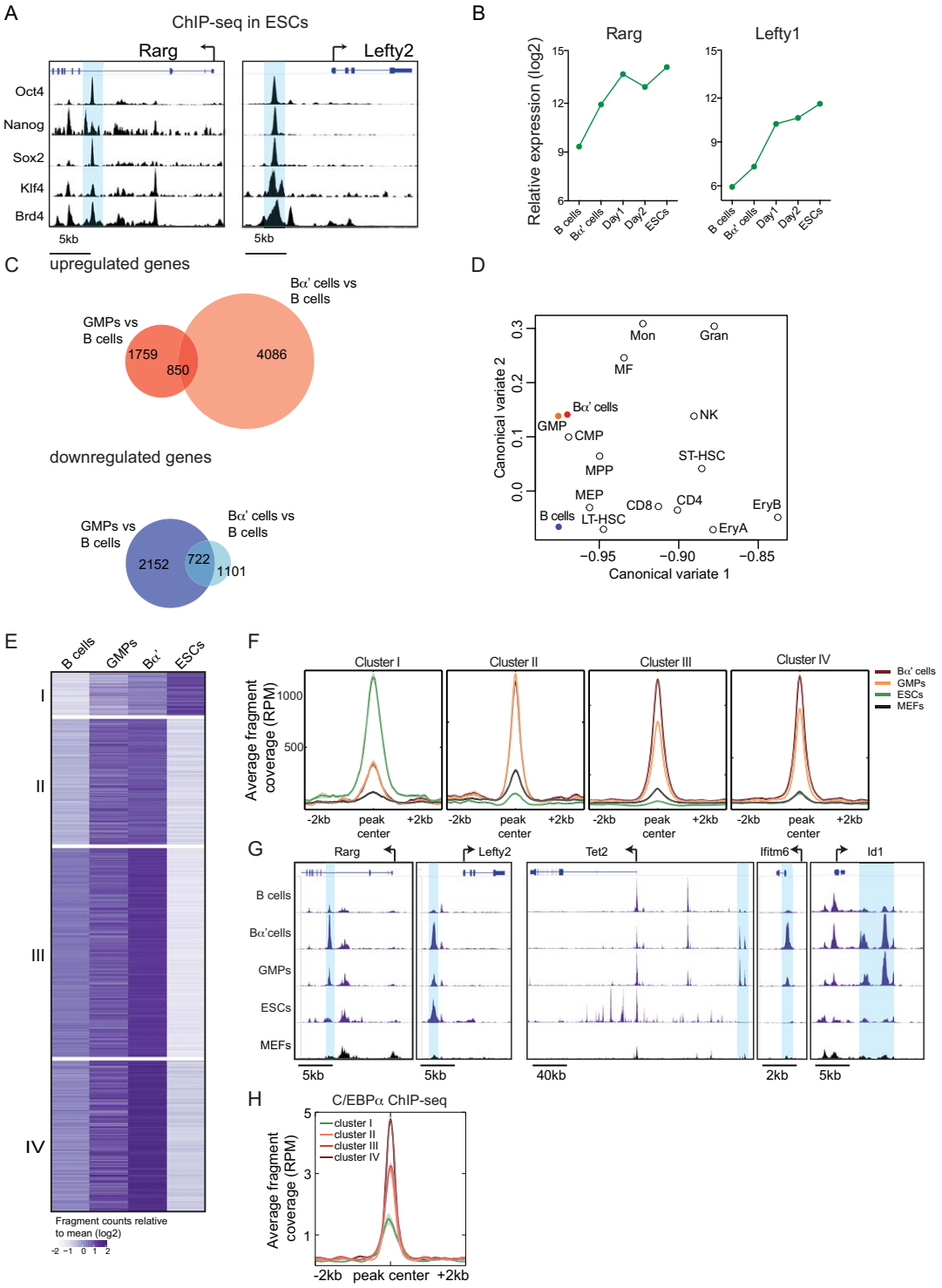
Supplementary Figure 3 Gene silencing induced by C/EBPα, protein interactions and B cell specific gene enhancer activities during reprogramming. (A) Representative western blots of Brd4, Lsd1, Klf4 and Hdac1 in B and Bα' cells. See Suppl. Fig. 8 for uncut gel images. (B) RNA-seq expression values for selected B cell specific genes. The data represent the average from two biologically independent samples. (C) Western blots of Cdk9 after induction of C/EBPα in B cells. See Suppl. Fig. 8 for uncut gel images. (D) Bα' cell extracts were fractionated on Superose 6 10/300 GL column and Hdac1, Lsd1 and C/EBPα were probed by western blot. See Suppl. Fig. 8 for uncut gel images. (E) Peptide counts, P-value and enrichment over IgG of C/EBPα, Hdac1 and Lsd1, for the IP-mass spectrometry shown in Fig. 3B. (F) C/EBPα co-immunoprecipitation experiment. Lsd1 or C/EBPα were probed by western blot. See Suppl. Fig. 8 for uncut gel images. (G) Co-immunoprecipitation of C/EBPα, Lsd1 and Hdac1. Parp1 and Pcna (negative controls) were probed by western blot. See Suppl. Fig. 8 for uncut gel images. (H) Screenshots of H3K27ac histone decoration and Brd4 binding by ChIP-seq at enhancers of selected B cell transcription factors. (I) Gene expression of selected B cell genes as measured by qRT-PCR in B cells (data from Fig. 3F), B cells treated for 18h with E2 (Bα' cells) and B cells treated for 18h with both E2 and the Hdac1 inhibitor VPA. Error bars indicate s.d. (n=3 biologically independent samples). Statistical significance was determined using a two-tailed unpaired Student's t-test (*P<0.05, **P<0.01).



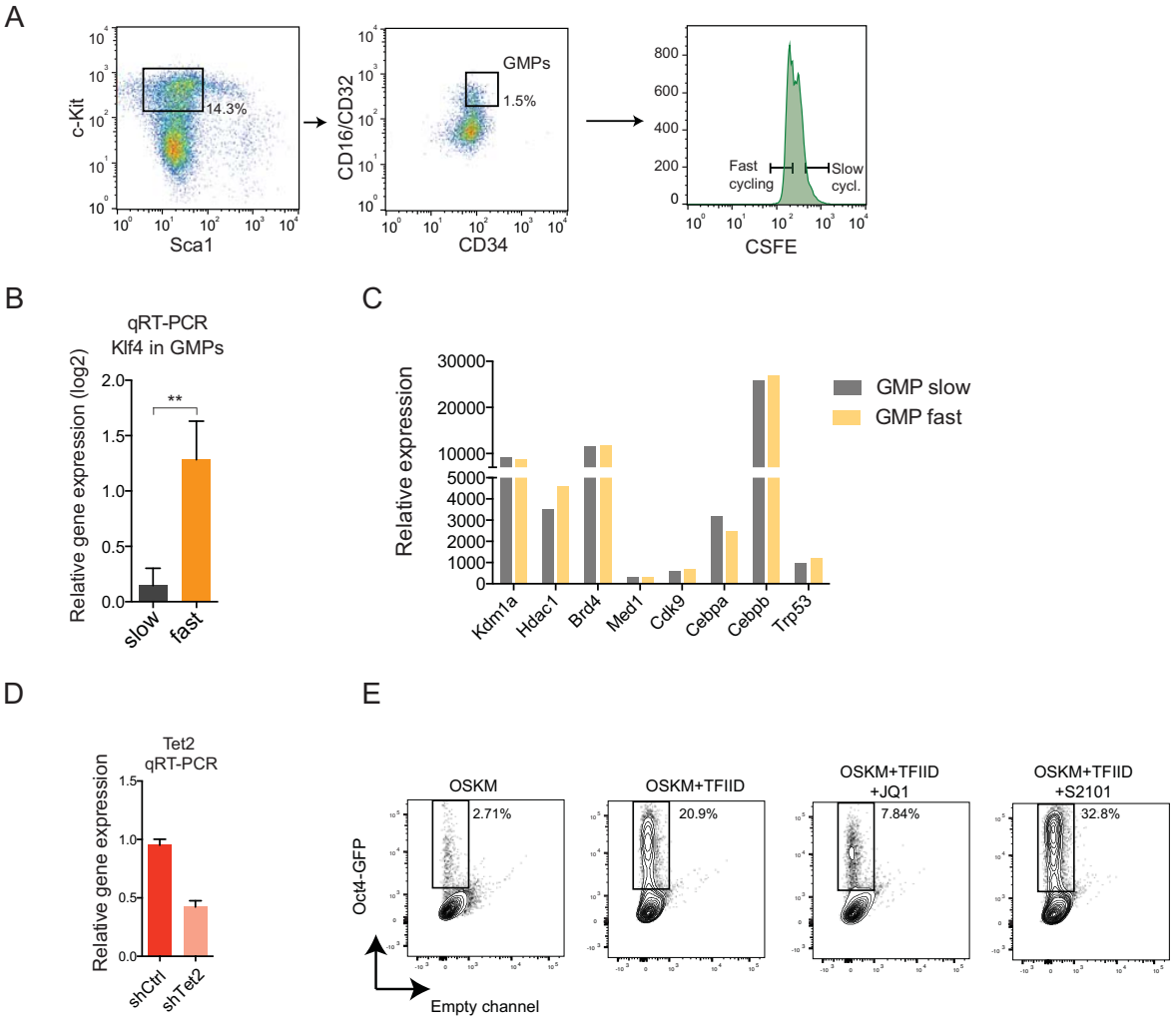
Supplementary Figure 4 Effect of Lsd1 and Brd4 inhibitions on iPS reprogramming. (A) Representative flow cytometry analysis of B cells treated with JQ1 or S2101 for 24 hours using Pacific Blue™ Annexin V/SYTOX® AADvanced™ Apoptosis Kit. (B) Representative BrdU (6h pulse) FACS staining of B cells treated with JQ1 or S2101 or DMSO as a control. (C) shRNA sorting strategy. (D) Gene expression by qRT-PCR of *Lsd1* and *Brd4* after specific knockdown in B cells. Error bars indicate s.d. (n=3 biologically independent samples). (E) Representative alkaline phosphatase positive iPS colonies obtained from reprogramming of B cells after Lsd1 and Brd4 knockdown. (F) Oct4-GFP and alkaline phosphatase positive iPS colonies obtained from reprogramming of B cells (OSKM alone without C/EBPα pulse) treated with S2101 or DMSO as a control. Error bars indicate s.d. (n=3 biologically independent samples). Statistical significance was determined using a two-tailed unpaired Student's t-test (n.s. P>0.05). (G) Genome browser screenshots of *Rarg* and *Egln3* loci showing C/EBPα, Brd4 and H3K27ac ChIP-seq data. (H) Representative alkaline phosphatase positive iPS colonies obtained from reprogramming of Bα' cells induced with OSKM and treated with JQ1 during C/EBPα (E2) or OSKM (Doxy) induction.



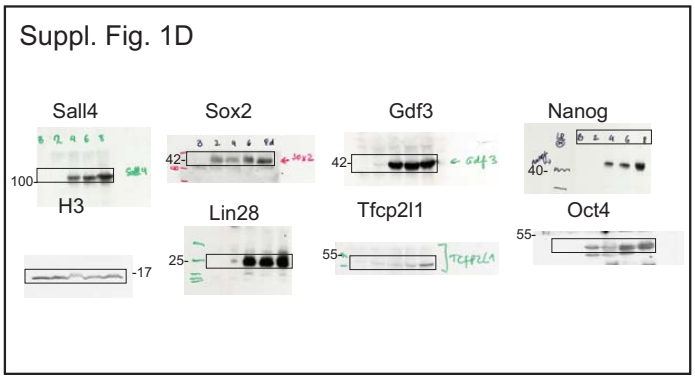
Supplementary Figure 5 ATAC-seq cluster analysis. (A) Gene ontology enrichment for genes associated with ATAC-seq peaks in each cluster shown in Figure 5A (nearest gene relative to the peak). P-values were determined by a hypergeometric test. (B) Genome browser screenshots of *Id1* and *Ifitm6* loci showing C/EBPα and H3K27ac ChIP-seq, as well as ATAC-seq data. (C) Selected over-represented DNA motifs shown in Figure 5B discovered (*de novo*) in ATAC-seq peaks, and similar motifs found in the JASPAR or HOCOMOCO database. (D) Genome browser screenshot of the *Klf4* locus showing C/EBPα and PU.1 ChIP-seq data, and 4C data using the newly discovered -90kb enhancer as view point (black triangle at the bottom). The second highlighted region (right) correspond to the second -280kb enhancer, as shown in Figure 5E. (E) Comparison of our ATAC-seq data (Fig. 5A), with Brd4 (GSE36561) and Klf4 (ref. 56) ChIP-seq data in ES cells. (F) Venn diagram showing the overlap between C/EBPα ChIP-seq peaks in Bα' cells and Klf4 ChIP-seq peaks in ES cells. (G) Average plots of C/EBPα ChIP-seq (top) and MNAse-seq signal (bottom) in the C10 pre-B cell line at different timepoints after induction of C/EBPα, for each ATAC-seq cluster (Fig. 5A). Profiles were normalized to B cells and centered on the median.

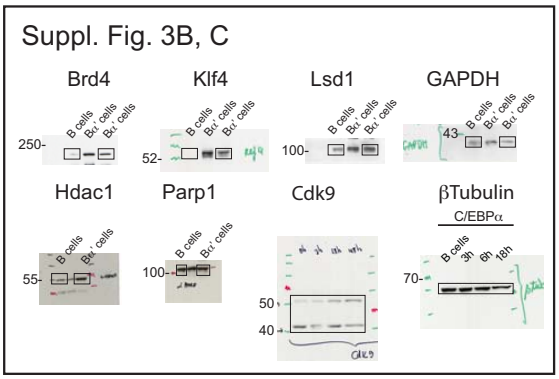


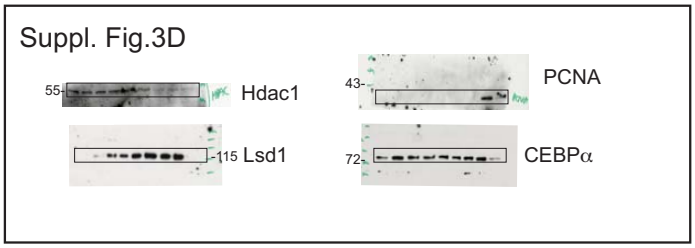
Supplementary Figure 6 C/EBPα induced changes in chromatin accessibility at myeloid and ES cell loci. (A) Genome browser screenshots of the *Rarg* and *Lefty2* loci showing ChIP-seq data for Oct4, Nanog, Klf4 and Brd4 in ESCs (ref. 56, 57 and GSE36561). (B) Gene expression profile by RNA-seq for *Rarg* and *Lefty1* during iPS reprogramming. The data represent the average from two biologically independent samples. (C) Comparison of GMPs and Bα' cells for the number of upregulated and downregulated genes (>2fold) between B and Bα' cells as well as between B cells and GMP, indicating the number of genes that overlap. (D) Canonical component analysis (CCA) of RNA-seq from B cells and Bα' cells, together with RNA-seq from different hematopoietic cell populations (ref. 58). (E) Heatmaps of ATAC-seq data from clusters I to IV of B cells, GMPs, Bα' cells, ESCs. (F) Average peak intensities of ATAC-seq data from clusters I to IV of GMPs, Bα' cells, ESCs and MEFs (ref. 44). (G) Genome browser screenshots of selected genomic loci displaying ATAC-seq data. (H) Average plot of C/EBPα ChIP-seq signal in GMPs for each ATAC-seq cluster.

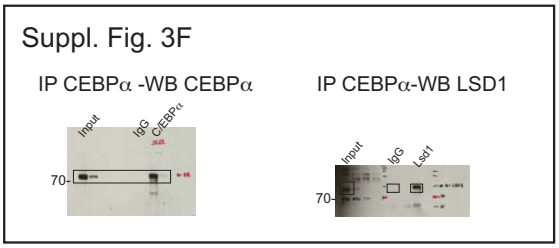


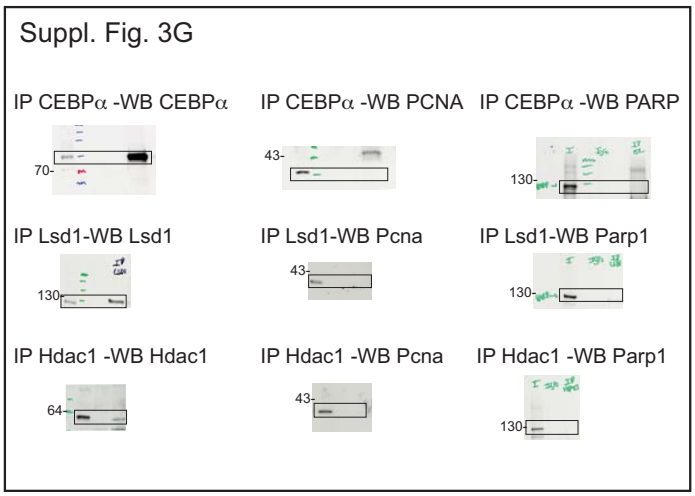
Supplementary Figure 7 Comparison of fast and slow cycling GMPs. (A) FACS plots showing sorting strategy to obtain GMPs and their separation into fast and slow cycling fractions after CSFE treatment. (B) *Klf4* expression as determined by qRT-PCR in fast and slow cycling GMPs. Error bars indicate s.d. (n=3 biologically independent samples). Statistical significance was determined using a two-tailed unpaired Student's t-test (**P<0.01). (C) Array expression values for selected genes in fast and slow cycling GMPs. The data represent the average from two biologically independent samples. (D) Tet2 knockdown efficiency tested by qRT-PCR. Error bars indicate s.d. (n=3 biologically independent samples). (E) Representative Oct4-GFP FACS analysis of OSKM-induced MEFs overexpressing TFIID and treated with JQ1 or S2101.

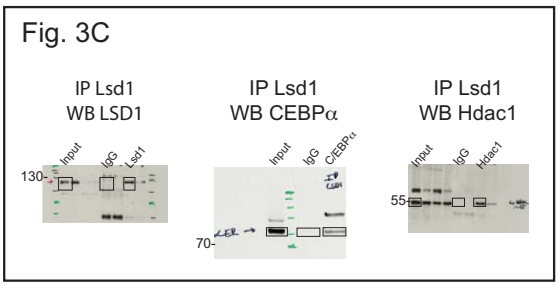


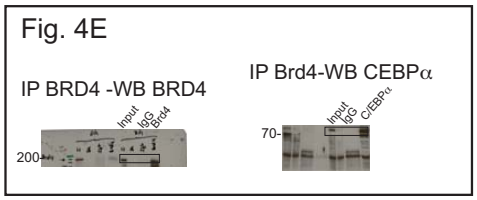












Supplementary Figure 8 Uncut gels.

Di Stefano et al_Suppl. Table 2

Gene	B cells	Ba' cells	Day1	Day2	ESCs
Sirt1	24.70275	27.23866	25.46961	26.54176	28.35141
Stx8	26.8133	25.78702	25.43324	27.02448	25.54576
Dcaf13	25.02057	28.36722	28.79546	29.19405	30.42817
Mkrn2	25.2734	24.22111	24.44819	25.5344	26.42498
Atg3	28.54859	27.55685	28.53241	28.94388	27.5876
Huwe1	27.74928	29.57995	29.38354	30.64386	31.68534
Xrcc5	26.16631	27.92961	27.53529	28.66051	30.04875
Uhrf1	29.13824	30.59902	30.22031	29.73002	31.70755
Trim33	25.54421	27.52574	26.53638	27.20658	28.10385
Atg7	28.70343	27.14042	28.4678	29.6404	26.30808
Arrb1	28.12383	25.44546	26.19578	28.05901	24.23697
Plk1	24.74554	27.64426	27.57963	26.2345	29.81607
Nedd4	25.4834	29.35289	30.33005	30.18005	32.64503
Fbxo22	27.85522	25.60917	27.27744	28.01333	29.16994
Ube2q1	24.48448	27.64395	27.82395	27.28807	27.78896
Ltn1	24.97531	27.54043	27.92525	28.42498	28.48326
Bid	27.77276	25.6609	27.3468	28.1997	27.91658
Senp3	25.23601	27.31095	27.73338	28.47647	30.40767
Trim28	32.05651	32.93272	32.59655	32.56633	34.43291
Trim30a	26.36436	25.46102	24.4115	26.1353	24.21135
Rnf213	33.08924	29.53486	31.33844	33.43363	31.60491
Psmd4	28.5357	30.09909	30.02147	30.02265	31.02364
Ubox5	25.54592	24.13976	25.1692	25.20078	25.37965
Ubr5	25.97307	28.40239	29.09079	29.83672	29.27233
Eif4e2	24.71424	27.16535	27.44242	27.68783	28.28967
Mycbp2	25.65503	24.56877	25.71953	27.95303	24.91644
Sash1	27.54677	25.3172	27.50339	27.36829	24.05951
Casc3	24.90638	26.14075	26.24807	26.37039	27.42977

Supplementary Table 1 List of genes in the independent component analysis shown in Fig. 1F.

Di Stefano et al _Supplementary Table 3

Primer Table

qPCR primers

Genes	Forward primer	Reverse primer
Tdh	CAGACTGAAGATAAAAGGCAG	GCATCTGTTCTTCTGATACC
Zfp296	CCATCTCAGAATCCAAAGAG	TATCTAGGTGTTGTGTGTCTGG
Nanog	CAGTTTTTCATCCCGAGAAC	CTTTTGTTTGGGACTGGTAG
Lin28a	TGTTCTGTATTGGGAGTGAG	CCATATGGTTGATGCTTTGG
Sall4	AAGAACTTCTCGTCTGCC	AGTGTACCTTCAGGTTGC
Gdf3	CGTCTTAAGGAAAATCATCCG	GGCAGACAAGTTAAAATAGAGG
Pou5f1	GTCCCTAGGTGAGCCGTCTTT	AGTCTGAAGCCAGGTGTCCAG
Sox2	ATGAGAGATCTTGGGACTTC	TCTATACATGGTCCGATTCC
Zfp42	GTTCGTCCATCTAAAAAGGG	TAGTCCATTTCTCTAATGCCC
Esrrb	AAAGCCATTGACTAAGATCG	AATTCACAGAGAGTGGTCAG
Pgk	ATGTCGCTTTCCAACAAGCTG	GCTCCATTGTCCAAGCAGAAT
18s	AACCCGTTGAACCCCATT	CCATCCAATCGGTAGTAGCG
Cdh1	CATGTTCACTGTCAATAGGG	GTGTATGTAGGGTAACTCTCTC
Ebf1	ACACAATTCATTCCCCGAAA	AAGTCAACGGTTTTGCATCC
Foxo1	AAGAGCGTGCCCTACTTCAA	CTCCCTCTGGATTGAGCATC
Gfi1b	TAATTCCTGGGCAAAAGAG	TGTTTGATTGTGTTCCAGCT
Ikzf3	CTTTTCTTCAGAACCCTGAC	CAATTGCTTGCTAATCTGTCC
Rag1	GAAGCTTCTGGCTCAGTCTACATCT	ACCTCATAGCGCTGCAGGTT
Ciita	CTGGACAAGAATGTCATCTG	TTGACTCTTATGGGCTATGG
Klf4	CATTAATTGTGTCGGAGGAAG	CCGTTTGGTACCTTTAGAAC
Lefty1	TGTGTGCTCTTTGCTTCC	GGGGATTCTGTCCTTGGTTT
Lsd1	TCATTCAGCTGCAAGAAAAG	TCCTCCTGAGTTTTCACTATC
Brd4	CTGATGTCCGATTGATGTTC	AGAGGACACTGTAACAACTG

ChIP primers

Genes	Forward primer	Reverse primer
Ebf1	CAGCAACCAAAACCTAGCAA	TCCCACTATTTATTCCCACA
Ciita	ACCTTGGGAGTATGCACTGG	AATTGGGTGACCACAGAAGC
Rag1	TCTCGCTCTCCTGTCAGTCA	CCGAGCAGAGACGTTAGCTT
Gfi1b	TCCCCAGAAATCATGTCAGA	GCTATTTCTGCCAAGGGTGA
Foxo1	CTGGTCAAGCTCTTGCCTGT	GGATTGCAAGTTCTCCTCCA
Ikzf3	GCCAAAGAAACACAGGCAAT	CCTCAAGAGCTGCTCACCTT
control (gene desert)	TCAGAAAGGAATCAATCAATCAAA	ATGCCCTCTTCTGGTGTGTC

4C primers

Genes	Reading primer	Non-reading primer
Klf4 promoter	AATGATACGGCGACCACCGAACACTC TTTCCCTACACGACGCTCTTCCGATCT GACAGGACAAGCGCGTAC	CAAGCAGAAGACGGCATACGAGAGATA CCTTTCACCAGGGAT
Klf4 enhancer	AATGATACGGCGACCACCGAACACTC TTTCCCTACACGACGCTCTTCCGATCT CGCTTTATGTTCTGCCAGTAC	CAAGCAGAAGACGGCATACGATGTCAC AGCCCCAGTAGTG

Supplementary Table 2 List of proteins annotated with the GO term "protein degradation".

Supplementary Table 3 List of primers used.