Cell Stem Cell, Volume *18*

Supplemental Information

Cell-of-Origin-Specific 3D Genome Structure Acquired during Somatic Cell Reprogramming

Peter Hugo Lodewijk Krijger, Bruno Di Stefano, Elzo de Wit, Francesco Limone, Chris van Oevelen, Wouter de Laat, and Thomas Graf

Figure S1

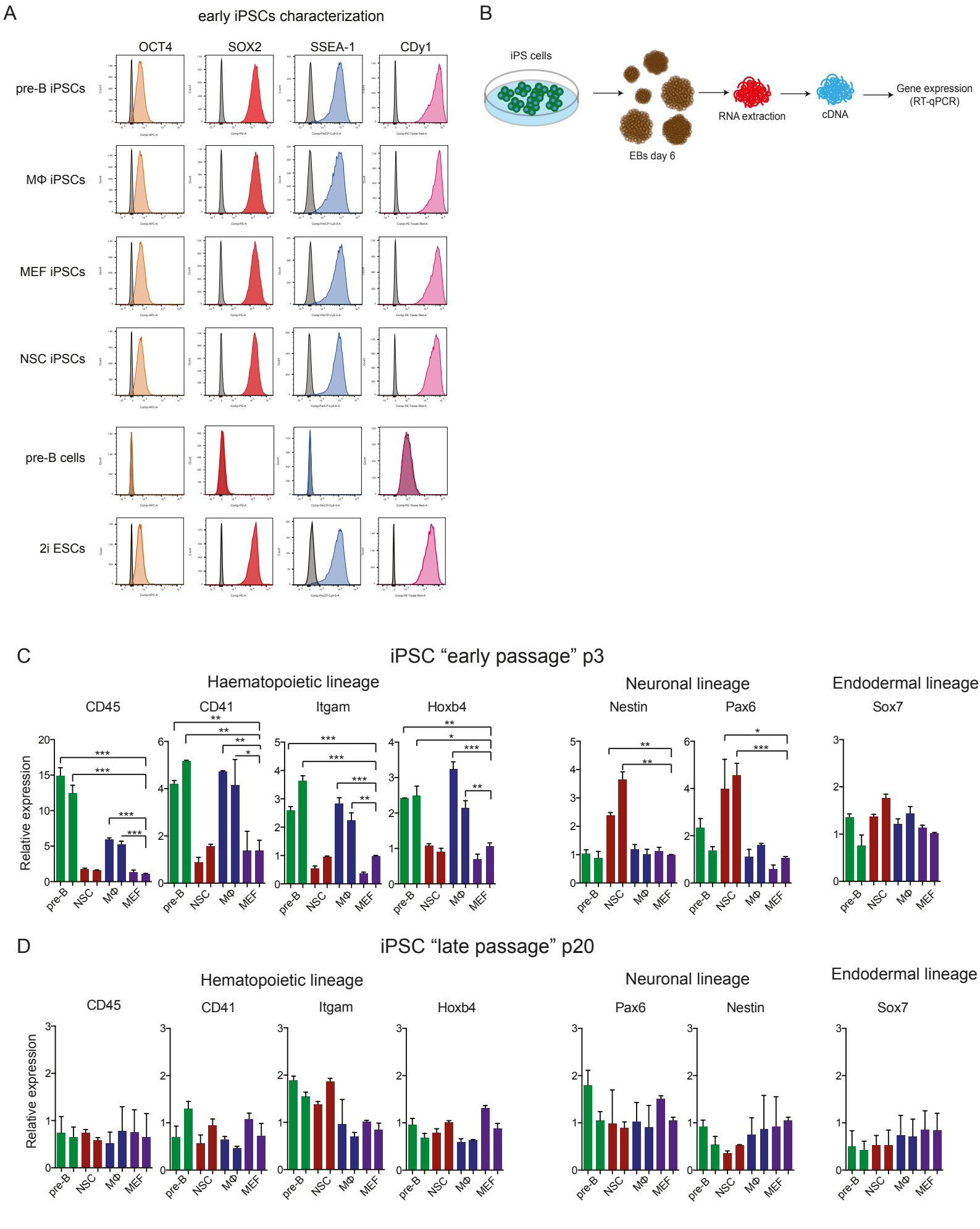
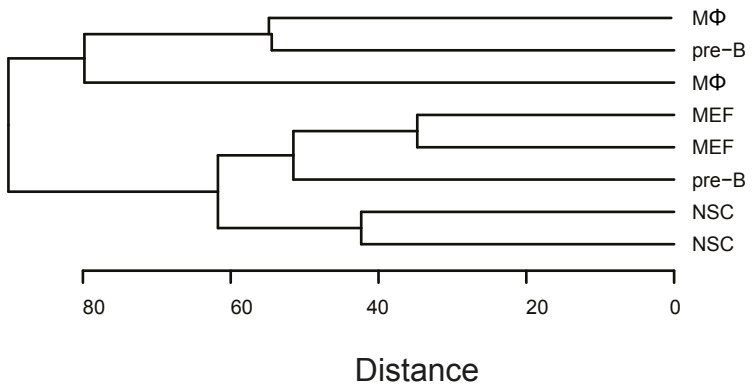
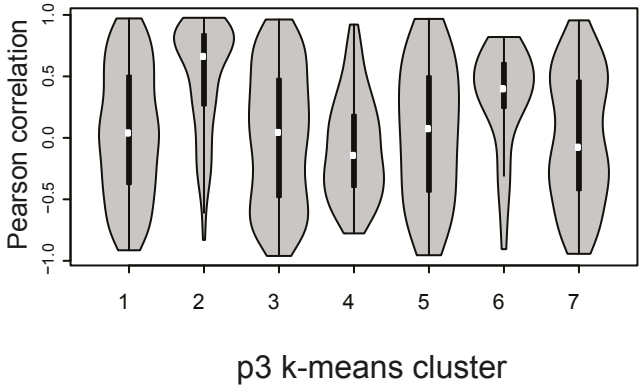

A
early iPSCs characterization
OCT4
SOX2
SSEA-1
CDy1
pre-B iPSCs
MΦ iPSCs
MEF iPSCs
NSC iPSCs
pre-B cells
2i ESCs
B
iPS cells
EBs day 6
RNA extraction
cDNA
Gene expression (RT-qPCR)
C
iPSC "early passage" p3
Haematopoietic lineage
Neuronal lineage
Endodermal lineage
CD45
CD41
Itgam
Hoxb4
Nestin
Pax6
Sox7
Relative expression
pre-B
NSC
MΦ
MEF
D
iPSC "late passage" p20
Hematopoietic lineage
Neuronal lineage
Endodermal lineage
CD45
CD41
Itgam
Hoxb4
Pax6
Nestin
Sox7
Relative expression
pre-B
NSC
MΦ
MEF

Figure S2

A



B



C

	GO class	GO ID	adjusted p-value	number of genes
cluster 1	regulation of response to stimulus	GO:0048583	0.0005	55
	regulation of cell proliferation	GO:0042127	0.0005	37
	cytokine production	GO:0001816	0.0005	20
	cellular response to interferon-beta	GO:0035458	0.0017	4
cluster 2	response to wounding	GO:0009611	1.16E-21	42
	immune system process	GO:0002376	1.74E-16	51
	fibrillar collagen	GO:0005583	1.25E-08	6
	collagen	GO:0005581	2.18E-12	14
	vasculature development	GO:0001944	1.81E-11	27
cluster 3	nucleus	GO:0005634	9.41E-06	80
cluster 4	NA			
cluster 5	zinc ion binding	GO:0008270	4.13E-07	76
	RNA biosynthetic process	GO:0032774	1.82E-05	116
cluster 6	NA			
cluster 7	vacuole	GO:0005773	0.0059	8

D

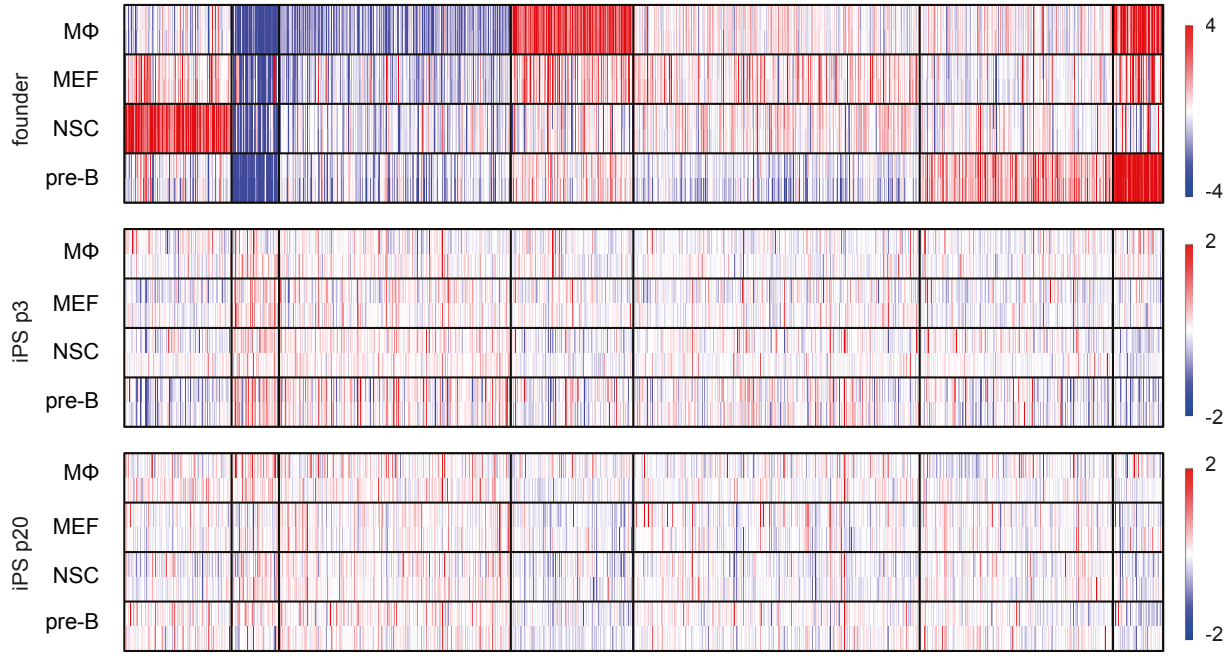


Figure S3

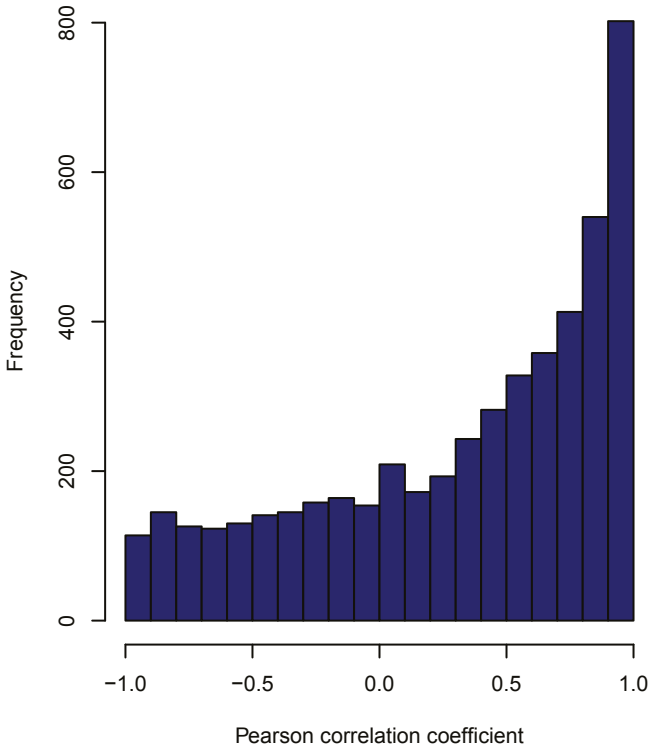
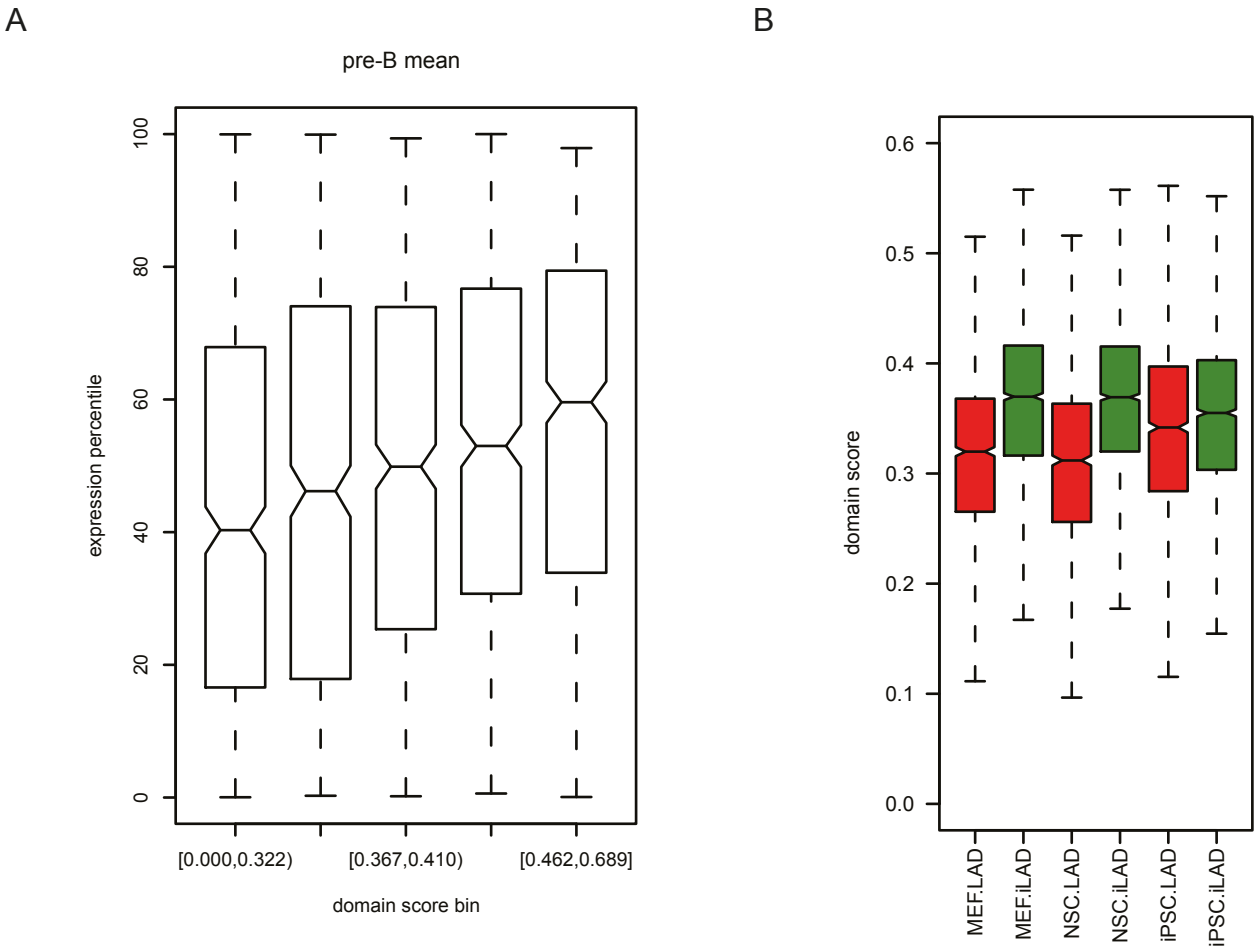

Frequency
0
200
400
600
800
−1.0
−0.5
0.0
0.5
1.0
Pearson correlation coefficient

Figure S4


A
pre-B mean
expression percentile
0
20
40
60
80
100
[0.000,0.322)
[0.367,0.410)
[0.462,0.689]
domain score bin
B
domain score
0.0
0.1
0.2
0.3
0.4
0.5
0.6
MEF.LAD
MEF.iLAD
NSC.LAD
NSC.iLAD
iPSC.LAD
iPSC.iLAD

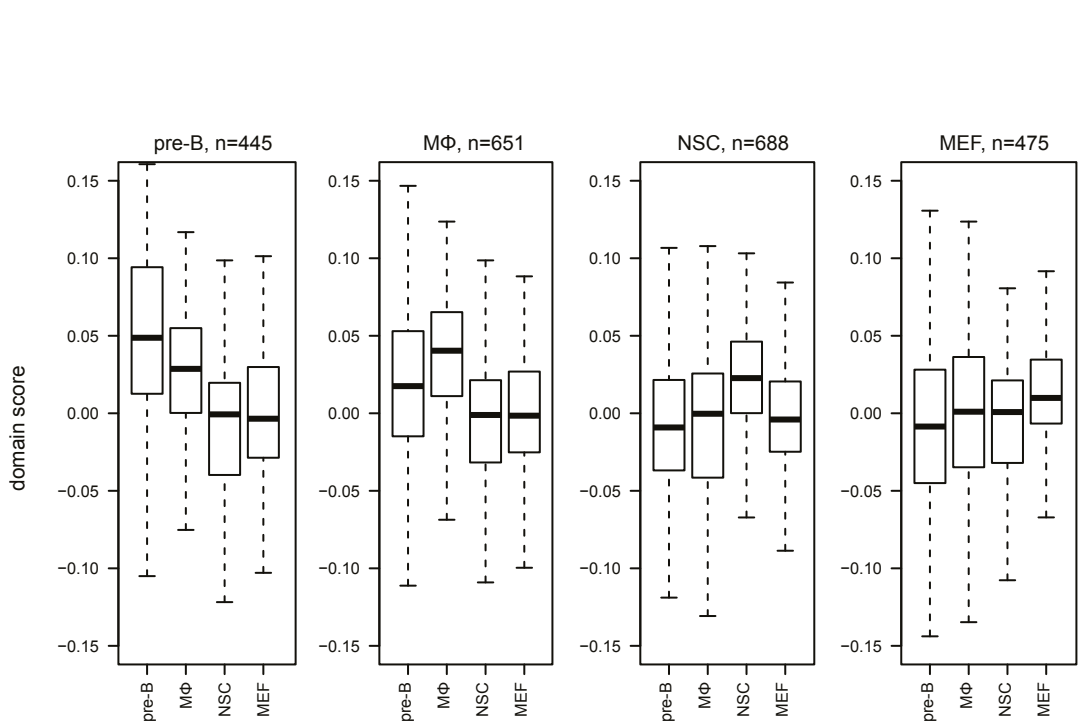

C
pre-B, n=445
MΦ, n=651
NSC, n=688
MEF, n=475
domain score
0.15
0.10
0.05
0.00
−0.05
−0.10
−0.15
pre-B
MΦ
NSC
MEF

Figure S5

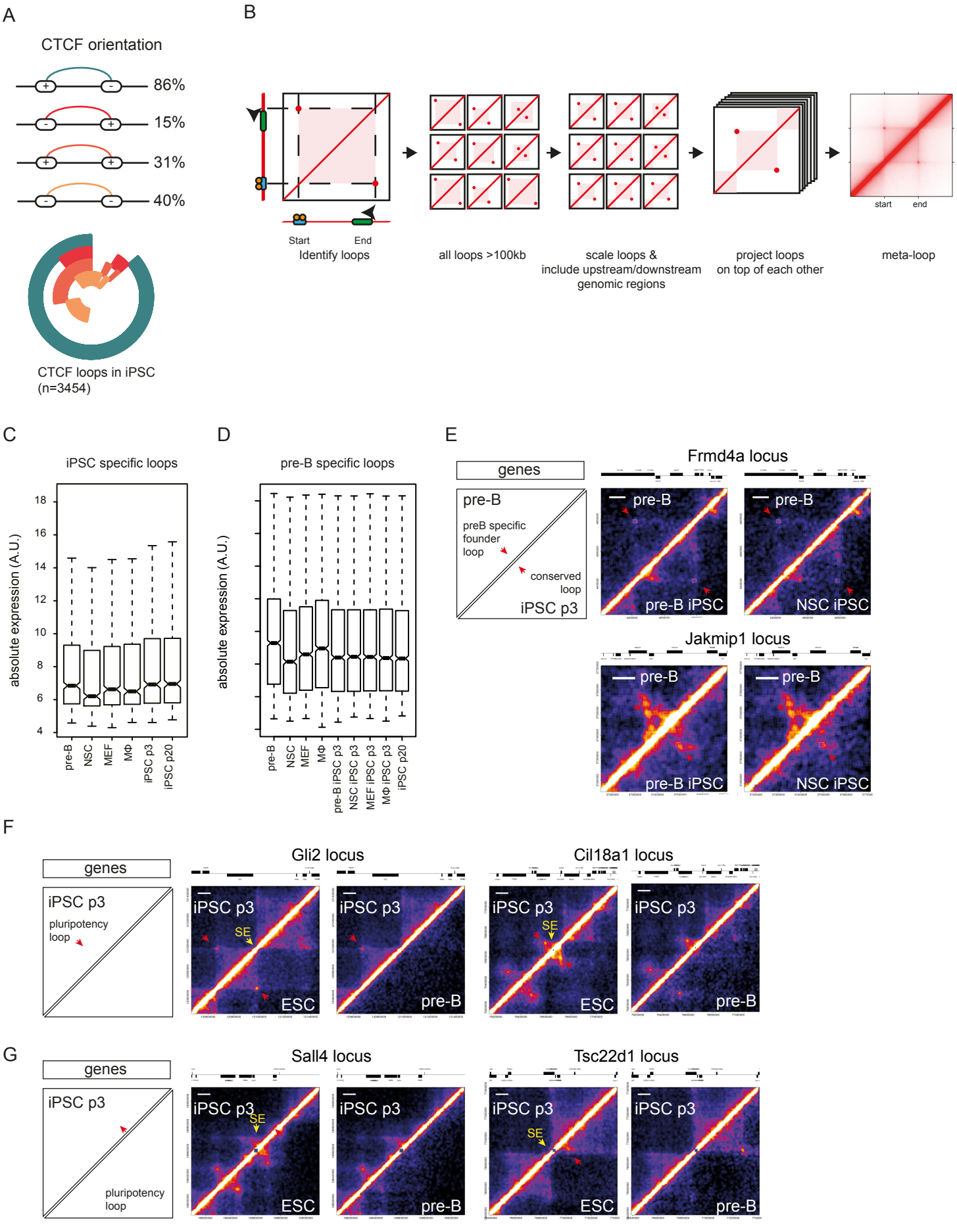
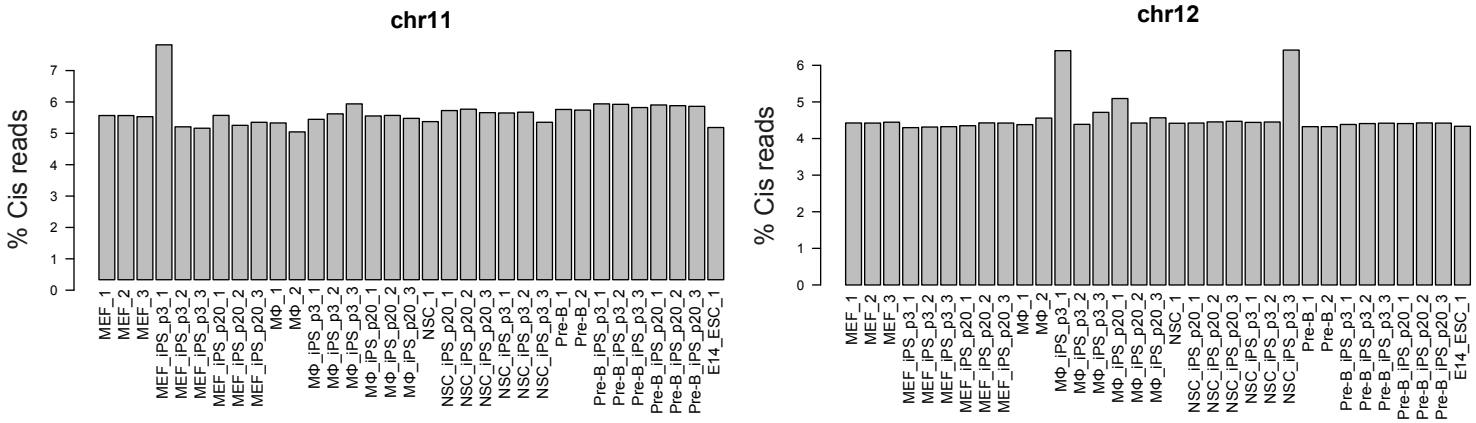

A
CTCF orientation
86%
15%
31%
40%
CTCF loops in iPSC
(n=3454)
B
Start
End
Identify loops
all loops >100kb
scale loops &
include upstream/downstream
genomic regions
project loops
on top of each other
start
end
meta-loop
C
iPSC specific loops
absolute expression (A.U.)
pre-B
NSC
MEF
MΦ
iPSC p3
iPSC p20
D
pre-B specific loops
absolute expression (A.U.)
pre-B
NSC
MEF
MΦ
pre-B iPSC p3
NSC iPSC p3
MEF iPSC p3
MΦ iPSC p3
iPSC p20
E
genes
pre-B
preB specific
founder
loop
conserved
loop
iPSC p3
Frmd4a locus
pre-B
pre-B iPSC
pre-B
NSC iPSC
Jakmip1 locus
pre-B
pre-B iPSC
pre-B
NSC iPSC
F
genes
iPSC p3
pluripotency
loop
Gli2 locus
iPSC p3
SE
ESC
iPSC p3
pre-B
Cil18a1 locus
iPSC p3
SE
ESC
iPSC p3
pre-B
G
genes
iPSC p3
pluripotency
loop
Sall4 locus
iPSC p3
SE
ESC
iPSC p3
pre-B
Tsc22d1 locus
iPSC p3
SE
ESC
iPSC p3
pre-B

Figure S6

A



B

Shared CTCF peaks
iPSC p3

Shared CTCF peaks iPSC p3	MEF #1	MEF #2	Mo #1	Mo #2	NSC #1	NSC #2	pre-B #1	pre-B #2
MEF #1	69152	63001	53970	59026	66341	67832	65647	64723
MEF #2	63001	71707	54260	61396	70196	71250	69664	67580
Mo #1	53970	54260	61074	52518	56437	58149	56375	56507
Mo #2	59026	61396	52518	64721	63784	64410	63336	61943
NSC #1	66341	70196	56437	63784	90396	87422	80209	75848
NSC #2	67832	71250	58149	64410	87422	109839	85347	80581
pre-B #1	65647	69664	56375	63336	80209	85347	87678	76727
pre-B #2	64723	67580	56507	61943	75848	80581	76727	84428

C

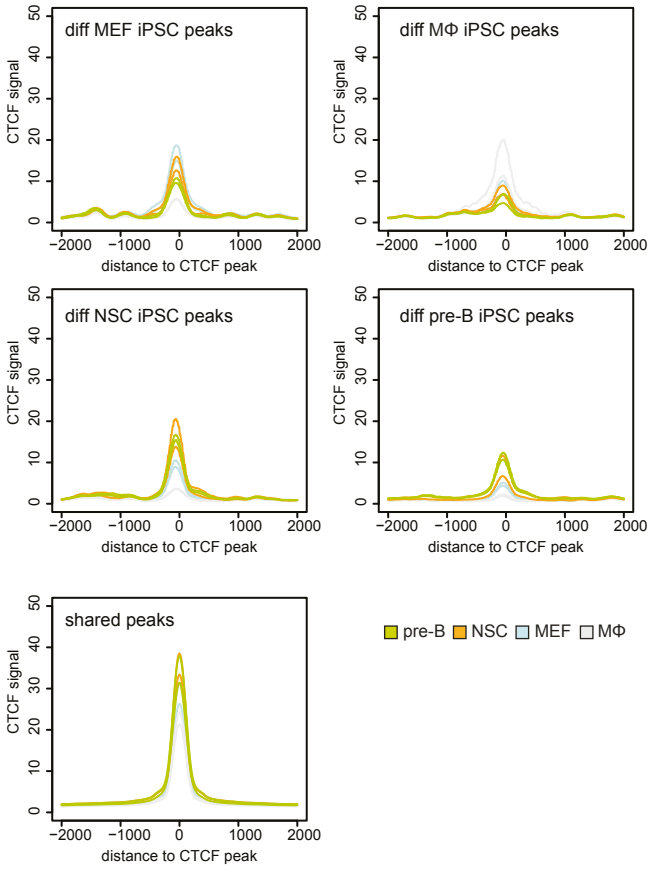


Figure S1. Differentiation bias in early passage iPSCs. Related to Figure 1.

(A) FACS analyses of early passage iPSCs, pre-B and naïve ESC for the expression of the pluripotency transcription factors OCT4 and SOX2, the surface marker SSEA-1 and for the stem cell specific dye CDy1. Essentially 100% of the different P3 iPS cell populations stained positive for these markers, similar to the ESCs.
(B) Schematic representation of the expression analysis of selected lineage specific genes in day 6 embryoid bodies (EBs) derived from iPSCs.
(C) Expression analysis of EBs derived from early passage iPSCs derived from pre-B (blue), NSC (green), MΦ (red) and MEF (purple). Normalized against PGK. Error bars indicate s.d (n=3). Student's t-test *p<0.05, **p<0.01, ***p<0.001
(D) Expression analysis of EBs derived from late passage iPSCs derived from pre-B (blue), NSC (green), MΦ (red) and MEF (purple). Normalized against PGK. Error bars indicate s.d (n=3). Student's t-test *p<0.05, **p<0.01, ***p<0.001

Figure S2. Cell-of-origin influences gene expression in early passage iPSCs in an indirect manner. Related to Figure 2.

(A) Unsupervised hierarchical clustering of the transcription profiles of late passage iPSCs.
(B) Violin plots showing the Pearson correlation scores comparison of expression profiles in founder cells and early passage iPSCs for all differentially expressed genes in the early passage iPSCs. Correlation scores are stratified based on the clusters from the k-means clustering in p3 iPSCs,
(C) Gene Ontology analysis of genes located in k-means clusters indentified in early passage iPSCs. Enrichment anlysis was done in WebGestalt (Zhang et al., 2005).
(D) k-means clustering of differentially expressed genes between pre-B, NSC, MΦ and MEF. Expression change of each differentially expressed gene is indicated for the founder cells, early passage iPSC and late passage iPSC (n=2).

Figure S3. Pearson correlation coefficients of the eigenvector with differential expression. Related to Figure 3.

A histogram shows the distribution of the correlation coefficient between the differential expression in the founders and the eigenvector score from the Hi-C in the genomic region that contained them. The analysis shows an intersection of regions that switch from A to B compartment or vice versa and genes that are significantly differentially expressed.

Figure S4. Higher domain scores are associated with increased expression and internal nuclear position. Related to Figure 4.

(A) Average expression of genes in pre-B plotted as the expression percentile of all TADs binned for the domain score in pre-B.
(B) Domain scores of MEF, NSC and iPS for LAD and iLAD identified in MEF, NSC and ESC (Peric-Hupkes et al., 2010).
(C) Domain scores of TADs containing tissue specific genes (having four-fold higher expression in both replicates of a given founder compared to the replicates of all other somatic cells) in pre-B, MΦ, NSC and MEF. From left to right tissue specific genes in pre-B, MΦ, NSC and MEF.

Figure S5. Characterization and validation of iPSC loops. Related to Figure 5.

(A) CTCF motif orientation of all iPSC loops identified between two ESC CTCF binding sites.
(B) Schematic representation of the meta-loop analysis.

(C) Absolute expression of genes (A) located in pre-B specific loops in pre-B, NSC, MEF, MΦ, iPSC p3 derived from pre-B, NSC, MEF, MΦ and iPSC p20.
(D) Absolute expression of genes located in iPS specific loops in pre-B, NSC, MEF, MΦ, iPSC p3 and iPSC p20.
(E) Hi-C interaction heatmap showing chromatin loops in pre-B cells (upper triangle) and iPSC derived from pre-B and NSC (lower triangle) at the *Frmd4a* (top) and *Jakmip1* (bottom) locus. pre-B specific founder loops conserved in iPSC are indicated with a red arrowhead. Scale bar indicates 100kb.
(F) Hi-C interaction heatmap showing chromatin loops in p3 iPSC (upper triangle) and ESC (lower triangle) at the *Gli2* (left) and *Cil18a1* (bottom) locus. Pluripotency loops are indicated with a red arrowhead. Super enhancers identified in ESC are indicated with a yellow arrowhead. Scale bar indicates 100kb.
(G) Hi-C interaction heatmap showing chromatin loops in p3 iPSC (upper triangle) and ESC (lower triangle) at the *Sall4* (left) and *Tsc22d1* (bottom) locus. Pluripotency loops are indicated with a red arrowhead. Super enhancers identified in ESC are indicated with a yellow arrowhead. Scale bar indicates 100kb.

Figure S6. 3D genome of iPSCs memorizes its cell of origin. Related to Figure 6.

(A) %cis reads for chromosome 11 (top) and chromosome 12 (bottom) for each Hi-C library
(B) Number of CTCF peaks shared between early passage iPSCs.
(C) CTCF ChIPseq scores for the cell-of-origin dependent differential and shared CTCF binding sites in early passage iPSCs.

Table S1. Number of Valid Hi-C Reads. Related to Figure 3.

Experiment	Valid Hi-C Reads	Valid Cis Reads	% Valid Cis Reads
MEF_1	14420095	10884256	75.5
MEF_2	13533085	9928328	73.4
MEF_3	11263329	8954602	79.5
Total MEF	39216509	29767186	75.9
MEF_iPS_p3_1	27702764	21645276	78.1
MEF_iPS_p3_2	16075757	11056408	68.8
MEF_iPS_p3_3	14370735	10619560	73.9
Total MEF iPS p3	58149256	43321244	74.5
MEF_iPS_p20_1	19167057	16217187	84.6
MEF_iPS_p20_2	9760872	8213675	84.1
MEF_iPS_p20_3	14731426	12561757	85.3
Total MEF iPS p20	43659355	36992619	84.7
MΦ_1	23243455	15315162	65.9
MΦ_2	32445564	21033790	64.8
Total MΦ	55689019	36348952	65.3
MΦ_iPS_p3_1	14524297	11065086	76.2
MΦ_iPS_p3_2	17716188	12314282	69.5
MΦ_iPS_p3_3	14942672	10996057	73.6
Total MΦ iPS p3	47183157	34375425	72.9
MΦ_iPS_p20_1	14942764	12207532	81.7
MΦ_iPS_p20_2	17926833	13980168	78.0
MΦ_iPS_p20_3	15112059	12472334	82.5
Total MΦ iPS p20	47981656	38660034	80.6
NSC_1	52575404	37706443	71.7
NSC_iPS_p20_1	15709417	13087382	83.3
NSC_iPS_p20_2	16485267	13955056	84.7
NSC_iPS_p20_3	18140176	15071380	83.1
Total NSC iPS p3	50334860	42113818	83.7
NSC_iPS_p3_1	18054935	15309925	84.8
NSC_iPS_p3_2	15899555	12904538	81.2
NSC_iPS_p3_3	17064723	12475987	73.1
Total NSC iPS p20	51019213	40690450	79.8
Pre-B_1	32448132	24918448	76.8
Pre-B_2	37414976	28567354	76.4
Total pre-B	69863108	53485802	76.6
Pre-B_iPS_p3_1	22309511	17535333	78.6
Pre-B_iPS_p3_2	33129155	25614782	77.3
Pre-B_iPS_p3_3	25033777	18835842	75.2
Total pre-B_iPS_p3	80472443	61985957	77.0
Pre-B_iPS_p20_1	37891651	32772065	86.5
Pre-B_iPS_p20_2	15601950	12973052	83.2
Pre-B_iPS_p20_3	18931024	14959635	79.0
Total pre-B_iPS_p20	72424625	60704752	83.8
E14_ESC_1	29065859	22671989	78.0

Table S2. Eigenvector values. Related to Figure 3.

Table S3. Domain segmentation in iPSC p3. Related to Figure 4.

Table S4. List with genomic coordinates of all the called chromatin loops. Related to Figure 5.

Table S5. List of primers used for RT-qPCR. Related to Figure 1.

Gene	Forward	Reverse
Mac1	TCGTATGTGAGGTCTAAGACA	CAGCAGTGATGAGAGCCAAG
Sox1	GCAGCGTTTCCGTGACTTT	GGCAGAACCACAGGAAAGA
CD45	TTGTCACAGGGCAAACACCT	TTGGGGGTGTGGATTCAGTG
CD41	AAGGACAAACATGGAAGCGTG	CTCTTGACTTGCGTTTAGGGC
Hoxb4	CCTGGATGCGCAAAGTTCA	CGTCAGGTAGCGATTGTAGTGA
Nestin	CTCTTGGCTTTCCTGACCCC	AGGCTGTCACAGGAGTCTCA
Pax6	TTCCCGAATTCTGCAGACCC	GTCGCCACTCTTGGCTTACT
Sox7	TCAGGGGACAAGAGTTCGGA	CCTTCCATGACTTTCCCAGCA
Pgk	ATGTCGCTTTCCAACAAGCTG	GCTCCATTGTCCAAGCAGAAT

Supplemental Experimental Procedures

Mice

We used a 'reprogrammable mouse' line containing a doxycycline-inducible OSKM cassette, the reverse tetracycline transactivator (rtTA), (Carey et al., 2010) and an Oct4-GFP reporter transgene (Boiani et al., 2002) as described (Di Stefano et al., 2014). Mice were housed in standard cages under 12 hour 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 and iPSCs were cultured on Mitomycin-C treated MEFs in KO-DMEM (Invitrogen) supplemented with 1% nonessential amino acids (Invitrogen), 0.1 mM β-mercaptoethanol (Invitrogen), 1,000 U/ml LIF (Millipore) and 15% fetal bovine serum (FBS, Invitrogen) (ESC medium). Naïve ESCs were cultured in serum-free N2B27 medium on gelatin-coated dishes. N2B27 medium (500ml) was generated by inclusion of the following: 240ml DMEM/F12 (Invitrogen), 240ml Neurobasal (Invitrogen), 5ml N2 supplement (Invitrogen), 10ml B27 supplement (Invitrogen), 1,000 U/ml LIF (Millipore), 1% nonessential amino acids (Invitrogen), 0.1mM β-mercaptoethanol (Invitrogen), and the small molecules PD0325901 (Stemgent, 1μM) and CHIR (Stemgent, 3μM). $CD19^+$ pre-B cells and $Mac1^+$ macrophages were isolated from bone marrow with monoclonal antibodies to CD19 and Mac-1 (BD Pharmingen) respectively, using MACS sorting (Miltenyi Biotech). Pre-B cells were grown in RPMI medium with 10% FBS and 10ng/ml IL-7 (Peprotech); macrophages in DMEM with 10% FBS and 10ng/ml each of CSF1 and IL-3 (Peprotech). MEFs were established from day 13.5 mouse embryos and cultured in DMEM containing 10% FBS. NSC were isolated and cultured as previously described (Di Stefano et al., 2009). All media were supplemented with L-glutamine and penicillin/streptomycin (GIBCO).

Reprogramming

Reprogramming experiments with pre-B cells were performed as previously described (Di Stefano et al., 2014); with MEFs, macrophages and NSCs were conducted by plating 100.000 cells/well on gelatinized plates seeded with irradiated MEFs, using ESC medium supplemented with 2 μg/ml of doxycycline. For the isolation of iPSC lines, doxycycline was washed out after 15 days of reprogramming and colonies with ESC-like morphology were picked at 20 days before further passaging. iPSC lines were expanded for an additional 9 days (3 passages) to obtain P3 iPS cell lines or for 20 passages to obtain P20 iPS cell lines.

Immunofluorescence assays

The cells were fixed with 4% paraformaldehyde, blocked with 5% of goat serum solution in PBS and incubated with primary antibodies overnight at 4°C. On the next day, the cells were exposed to secondary antibodies (all Alexa Fluor from Invitrogen) at RT for one hour. The primary antibodies used were Nanog (Calbiochem, SC1000) and SSEA-1 (Santa Cruz, SC-21702). Nuclear staining was performed with DAPI (Invitrogen).

Flow cytometry

For FACS analysis of the pluripotency markers Oct4, Sox2, and SSEA-1, P3 iPSCs and naïve ESCs were fixed and stained using the Multicolor Flow Cytometry kit (R&D Systems, Minneapolis, MN), according to the manufacturer instructions. Live staining of the P3 iPSCs and naïve ESCs with the Stem Cell CDy1 Dye (Active Motif) was performed by incubating the cells with the CDy1 Dye for 1 hour at 37C, followed by washout of the dye and FACS analysis. For Oct4GFP expression, iPSCs were resuspended in PBS and analyzed by FACS. Cells were analyzed with an LSR Fortessa FACS machine (BD Biosciences) using Diva v6.1.2 (BD Biosciences) and FlowJo software v10 (TreeStar).

Chimeric mice

For the chimera formation assay, 10 to 15 iPS cells (Agouti color coat) were injected into 3.5dpc blastocysts of C57BL/6J mice (black coat color) and transferred into pseudo-pregnant CD1 females. Chimerism of the transplanted offspring was assessed by the presence of agouti coat color derived from the iPS cells.

Differentiation of iPSCs

Embryoid bodies were derived by plating iPSCs at a concentration of $1.3x10^6$ cells/ml in bacterial grade dishes in ES medium without LIF. After 6 days in culture they were harvested for total RNA extraction.

Expression analysis

To remove the feeder cells, iPS cells were cultured on gelatinized plates for 2 passages before RNA extraction. RNA was isolated with the miRNeasy mini kit (Qiagen), eluted with RNase-free water and quantified by Nanodrop. cDNA was produced with the High Capacity RNA-to-cDNA kit (Applied Biosystem). RNA samples with a RIN greater than 9 were analyzed by expression arrays. 500ng of total RNA per sample were labeled using Agilent's QuickAmp labeling kit and hybridized to Agilent 8X60K expression arrays. RT-qPCR reactions were set up in triplicate with the SYBR Green QPCR Master Mix (Applied Biosystem), using primers listed in Supplementary Table 2. Reactions were run on an AB7900HT PCR machine with 40 cycles of 30s at 95 °C, 30s at 58 °C and 30s at 72 °C. Raw array data was processed using limma (Ritchie et al., 2015). We perform "normexp" background correction with an offset of 16. We normalize between arrays using quantile normalization. To identify differently expressed genes we create a contrast matrix in which all pairwise comparisons are made. For each probe an empirical Bayes moderated t-statistic test is performed. An FDR correction is applied to the nominal p-values. All statistical analyses were performed in R. (R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.)

ChIPseq analysis

ChIP experiments were performed as described previously (van Oevelen et al., 2008) using an antibody against H3K27ac (ab4729, Abcam) and CTCF (Millipore, 07-729). DNA libraries were prepared using Illumina's reagents and instructions. All libraries were sequenced on the Hiseq2000 sequencer. Mapping of the raw fastq files was performed with bwa aln to the mm9 genome. Peak calling was performed with macs v. 1.4.2. (Zhang et al., 2008). Coverage scores for the creation of heatmaps were calculated with the compEpiTools package in R. To determine tissue-specific peaks in the founder tissues we calculated the coverage 600bp surrounding (300bp +/-) the summit of the peaks for a given tissue. We required that the coverage score was above 1 for the tissue of interest and below 1 for the three other tissues.

Differential peak calling

For the differential CTCF peaks between iPS lines we made use of THOR (Allhoff et al., 2014), which allows differential peak calling using replicate experiments using standard settings. After differential peak calling we performed a stringent selection of regions that were differentially bound. We selected regions that had a p-value below 10e-30 and had a fold change of at least 3.

Hi-C template generation & mapping

pre-B cells, MEF, NSC, iPSC and E14 ESC were cross-linked and further processed as DpnII 3C template as previously described (Splinter et al., 2012). Hi-C libraries from macrophages were incubated at 65°C during the lysis and subsequent SDS step to inactivate nucleases present in lysates from these cells. Ligation of crosslinked DNA fragments was performed without incorporating biotin, as described previously (Sexton et al., 2012). Libraries for paired-end sequencing were generated from sonicated, ~500-800bp size selected, 3C templates using the TruSeq DNA LT Sample Prep Kit (Illumina). Sequencing of the Hi-C libraries was performed on an Illumina HiSeq 2000. ESC Hi-C data used for the visual inspection of super enhancer loci were obtained from (Geeven et al., 2015).

Paired-end FASTQ files are mapped independently to the mouse genome (mm9) using bwa mem. Reads mapping to repetitive regions in the genome are filtered from the dataset. The resulting files are further filtered and deduplicated using HiCUP (http://www.bioinformatics.babraham.ac.uk/projects/hicup/). Reads that are within 1kb of each other are filtered from the dataset as they likely represent uncut genomic fragments. Reads mapped to the X chromosome were excluded in all analyses.

Eigenvector analysis

To segment the genome into A and B compartments we use the method described by Lieberman-Aiden et al. (Lieberman-Aiden et al., 2009). We divided the chromosomes into bins of 300kb and created contact frequency matrices for every chromosome. Next we performed vanilla coverage normalization (Rao et al., 2014) on the contact matrix. To this end we calculate a vector V_c representing the total intrachromosomal coverage for every 300kb window. A normalization matrix M_{norm} is calculated by taking the outer product of V_c . A normalized contact matrix is calculated by dividing the raw contact matrix by the square root of M_{norm}. The normalized contact matrix is transformed to an observed/expected matrix, which is used to calculate a spearman rank correlation matrix. The first eigenvector of the correlation matrix is used as the A/B segmentation.

Domain segmentation

To segment the genome into topologically associating domains (TADs) we used the R package HiCseg (Levy-Leduc et al., 2014). The software package requires a contact matrix as input. To this end we the combined the data for all the early iPS cell lines into an early dataset. The tag positions were transferred into fragment space. We binned the fragments into windows of 50 fragments, corresponding to a median genomic size of 20kb and calculated a fragment window matrix and performed vanilla coverage normalization (see above). The resulting matrix was used as input for the HiCseg algorithm.

Domain score

Given the TADs identified in the domain segmentation, we calculate an intradomain contact score (simplified to domain score). The domains score is the ratio of the number of contacts that occur between regions within the same TAD (intraTAD contacts) over the total number of intrachromosomal contacts for a TAD (intraTAD + interTAD). To identify TADs that have a differential domain score we used the bioconductor package limma (Ritchie et al., 2015). We perform quantile normalization (using the normalizeQuantiles function in the limma package) on the domain scores. To identify TADs with a differential domain score between the iPS lines we create a contrast matrix in which all pairwise comparisons are made. For each TAD an empirical Bayes moderated t-statistic test is performed. Nominal p-values are corrected using the FDR method. TADs with an FDR value < 0.05 were selected for subsequent analysis.

Loop calling

For the loop calling we initially follow a similar procedure as for the domain segmentation, i.e. mapping Hi-C tags to fragments and creating a coverage normalized contact frequency matrix in fragment space. For the contact matrix we use overlapping windows with a size of 50 fragments and a step size of 5 fragments. The resulting matrix is used as the input for the looping calling procedure.

A chromatin loop in a 4C or Hi-C experiment is appreciated as an increase in signal over the background. For these data the background is not distributed randomly, rather it follows a monotonically decreasing pattern, which is a function of the distance to the site of interest. To identify chromatin loops we explicitly model this background distribution by performing monotonic regression on the columns of the contact matrix. For this we use the R package isotone (Mair, 2009). We use the gpava function with the solver weighted.fractile to calculate the background distribution. Values that reach above a certain threshold value over the background and with a minimal amount of reads are selected as loops. Loops represent a combination of two windows in the genome, effectively forming a rectangle. Due to the nature of the input matrix and the fact that loops can span multiple windows, we often find overlapping rectangles for a single loop. We collapse overlapping rectangles and the select the rectangle with the highest coverage as the chromatin loop.

Loops were called in every founder tissue separately and in a combined dataset containing all early iPS datasets. To call tissue specific loops in a given founder tissue we first identified the loops in the selected founder dataset. Next we calculated the coverage for all the founders over the loops called in the selected founder. Coverage scores were normalized to total amount of valid ditags per experiment. Tissue specific loops were loops that have at least a 2-fold higher coverage in the selected tissue compared to the other tissue. The procedure for the iPS specific loops is similar, except that the coverage of the iPS loops has to be 2-fold higher compared to all founder tissues.

Contact frequency plots

To visualize the interaction between loci we create 2D heatmaps that show the contact frequency along a color scale. For the heatmap we generate a windowed interaction matrix with the pairwise coverage between genomic windows of 20kb within a given region. We use a sliding window approach with a step size of 2kb. The contact matrix is normalized using vanilla coverage (VC) normalization (see above). The VC normalization matrix is divided by the median, so that the scale of the resulting normalized contact matrix is preserved. Because different Hi-C experiments had different numbers of reads we needed to scale them in order to directly compare them in our visualizations of chromatin loops. Our scaling method is similar to the calculation of the RPKM score in RNAseq experiments (Mortazavi et al., 2008). We scale our dataset to 1 million intrachromosomal contacts per 100 Mb of sequence.

Meta-loop analysis

In order to study the distribution of Hi-C tags around a set of loops we performed created a meta-loop. In essence, this is a 2D variation of a meta-gene analysis, commonly performed for ChIPseq data. In the meta-loop analysis we align Hi-C tags on a set of loops. Because these are of varying size we scale the loops to have the same size. In addition to the loop itself we also analyse 50% of the genome upstream of the loop and 50% of the genome downstream of the loop. The genomic regions that interact in the loop are used as anchor points for the alignment. A frequency matrix is subsequently created in which the tag frequency in the loop regions and the flanking regions is stored. The frequency matrix is represented as a heatmap.

Alignment of Hi-C data to ChIP peaks (PE-SCAn)

Intrachromosomal Hi-C captures located more than 5MB of each other were aligned to ChIP data as previously described (de Wit et al., 2013). Briefly, one of the paired Hi-C reads was aligned to the ChIP data. Only reads that mapped within 500 kb up- or downstream of the ChIP peaks were selected for further analysis. Of this reduced set the corresponding read was also aligned to the ChIP peaks within 500 kb, resulting in a set of two distances (dx, dy) to all the Hi-C di-tags that are found within 500 kb of these peaks, for every intrachromosomal pair of ChIP peak. From the distribution of dx and dy a frequency matrix was calculated with a bin size of 50 kb. To calculate whether the binding sites of a given factor show preferential spatial contacts anenrichment score was calculated over a randomized data set. The randomized data set is calculated by aligning the Hi-C data to a circularly permuted ChIPseq data set, that is, the ChIP peaks are linearly shifted 10 Mb along the chromosome.

Data sources

H3K27ac data BMDM, H3K27ac data MEF from ENCODE (Yue et al., 2014), gene annotation from GENCODE (Harrow et al., 2012), H3K27ac data pre-B cells from (Lane et al., 2014), H3K27ac data NSC from (Creyghton et al., 2010), Sox2 data NPC and ESC from (Lodato et al., 2013), Pu.1 B cells from (Heinz et al., 2010), Oct4 data ESC from (Marson et al., 2008). LAD data from (Peric-Hupkes et al., 2010). CTCF data NPC (Phillips-Cremins et al., 2013), CTCF data MEF (Tedeschi et al., 2013), CTCF data macrophages (Daniel et al., 2014), CTCF data pre-B cells (Ribeiro de Almeida et al., 2011).

Supplemental References

Allhoff, M., Sere, K., Chauvistre, H., Lin, Q., Zenke, M., and Costa, I.G. (2014). Detecting differential peaks in ChIP-seq signals with ODIN. Bioinformatics *30*, 3467-3475.

Boiani, M., Eckardt, S., Scholer, H.R., and McLaughlin, K.J. (2002). Oct4 distribution and level in mouse clones: consequences for pluripotency. Genes & development *16*, 1209-1219.

Carey, B.W., Markoulaki, S., Beard, C., Hanna, J., and Jaenisch, R. (2010). Single-gene transgenic mouse strains for reprogramming adult somatic cells. Nature methods *7*, 56-59.

Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., *et al.* (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proceedings of the National Academy of Sciences of the United States of America *107*, 21931-21936.

Daniel, B., Nagy, G., Hah, N., Horvath, A., Czimmerer, Z., Poliska, S., Gyuris, T., Keirsse, J., Gysemans, C., Van Ginderachter, J.A., *et al.* (2014). The active enhancer network operated by liganded RXR supports angiogenic activity in macrophages. Genes & development *28*, 1562-1577.
de Wit, E., Bouwman, B.A., Zhu, Y., Klous, P., Splinter, E., Verstegen, M.J., Krijger, P.H., Festuccia, N., Nora, E.P., Welling, M., *et al.* (2013). The pluripotent genome in three dimensions is shaped around pluripotency factors. Nature *501*, 227-231.
Di Stefano, B., Prigione, A., and Broccoli, V. (2009). Efficient genetic reprogramming of unmodified somatic neural progenitors uncovers the essential requirement of Oct4 and Klf4. Stem cells and development *18*, 707-716.
Di Stefano, B., Sardina, J.L., van Oevelen, C., Collombet, S., Kallin, E.M., Vicent, G.P., Lu, J., Thieffry, D., Beato, M., and Graf, T. (2014). C/EBPalpha poises B cells for rapid reprogramming into induced pluripotent stem cells. Nature *506*, 235-239.
Geeven, G., Zhu, Y., Kim, B.J., Bartholdy, B.A., Yang, S.M., Macfarlan, T.S., Gifford, W.D., Pfaff, S.L., Verstegen, M.J., Pinto, H., *et al.* (2015). Local compartment changes and regulatory landscape alterations in histone H1-depleted cells. Genome Biol *16*, 289.
Harrow, J., Frankish, A., Gonzalez, J.M., Tapanari, E., Diekhans, M., Kokocinski, F., Aken, B.L., Barrell, D., Zadissa, A., Searle, S., *et al.* (2012). GENCODE: the reference human genome annotation for The ENCODE Project. Genome research *22*, 1760-1774.
Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Molecular cell *38*, 576-589.
Lane, A.A., Chapuy, B., Lin, C.Y., Tivey, T., Li, H., Townsend, E.C., van Bodegom, D., Day, T.A., Wu, S.C., Liu, H., *et al.* (2014). Triplication of a 21q22 region contributes to B cell transformation through HMGN1 overexpression and loss of histone H3 Lys27 trimethylation. Nature genetics *46*, 618-623.
Levy-Leduc, C., Delattre, M., Mary-Huard, T., and Robin, S. (2014). Two-dimensional segmentation for analyzing Hi-C data. Bioinformatics *30*, i386-392.
Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., *et al.* (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science *326*, 289-293.
Lodato, M.A., Ng, C.W., Wamstad, J.A., Cheng, A.W., Thai, K.K., Fraenkel, E., Jaenisch, R., and Boyer, L.A. (2013). SOX2 co-occupies distal enhancer elements with distinct POU factors in ESCs and NPCs to specify cell state. PLoS genetics *9*, e1003288.
Mair, P., Hornik, K. & de Leeuw, J. (2009). Isotone optimization in R: pool-adjacent-violators algorithm (PAVA) and active set methods. Journal of statistical software *32*, 1-24.
Marson, A., Levine, S.S., Cole, M.F., Frampton, G.M., Brambrink, T., Johnstone, S., Guenther, M.G., Johnston, W.K., Wernig, M., Newman, J., *et al.* (2008). Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. Cell *134*, 521-533.
Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nature methods *5*, 621-628.
Peric-Hupkes, D., Meuleman, W., Pagie, L., Bruggeman, S.W., Solovei, I., Brugman, W., Graf, S., Flicek, P., Kerkhoven, R.M., van Lohuizen, M., *et al.* (2010). Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. Molecular cell *38*, 603-613.
Phillips-Cremins, J.E., Sauria, M.E., Sanyal, A., Gerasimova, T.I., Lajoie, B.R., Bell, J.S., Ong, C.T., Hookway, T.A., Guo, C., Sun, Y., *et al.* (2013). Architectural Protein Subclasses Shape 3D Organization of Genomes during Lineage Commitment. Cell *153*, 1281-1295.
Rao, S.S., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., *et al.* (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell *159*, 1665-1680.
Ribeiro de Almeida, C., Stadhouders, R., de Bruijn, M.J., Bergen, I.M., Thongjuea, S., Lenhard, B., van Ijcken, W., Grosveld, F., Galjart, N., Soler, E., *et al.* (2011). The DNA-binding protein CTCF limits proximal Vkappa recombination and restricts kappa enhancer interactions to the immunoglobulin kappa light chain locus. Immunity *35*, 501-513.
Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic acids research *43*, e47.
Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A., and Cavalli, G. (2012). Three-dimensional folding and functional organization principles of the Drosophila genome. Cell *148*, 458-472.

Splinter, E., de Wit, E., van de Werken, H.J., Klous, P., and de Laat, W. (2012). Determining long-range chromatin interactions for selected genomic sites using 4C-seq technology: from fixation to computation. Methods *58*, 221-230.
Tedeschi, A., Wutz, G., Huet, S., Jaritz, M., Wuensche, A., Schirghuber, E., Davidson, I.F., Tang, W., Cisneros, D.A., Bhaskara, V., *et al.* (2013). Wapl is an essential regulator of chromatin structure and chromosome segregation. Nature *501*, 564-568.
van Oevelen, C., Wang, J., Asp, P., Yan, Q., Kaelin, W.G., Jr., Kluger, Y., and Dynlacht, B.D. (2008). A role for mammalian Sin3 in permanent gene silencing. Molecular cell *32*, 359-370.
Yue, F., Cheng, Y., Breschi, A., Vierstra, J., Wu, W., Ryba, T., Sandstrom, R., Ma, Z., Davis, C., Pope, B.D., *et al.* (2014). A comparative encyclopedia of DNA elements in the mouse genome. Nature *515*, 355-364.
Zhang, B., Kirov, S., and Snoddy, J. (2005). WebGestalt: an integrated system for exploring gene sets in various biological contexts. Nucleic acids research *33*, W741-748.
Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., *et al.* (2008). Model-based analysis of ChIP-Seq (MACS). Genome biology *9*, R137.