Supplementary Materials (Phillips-Cremins et al.)

Supplementary Figure Captions

Supplementary Data

Supplementary Tables

Supplementary Figures

Supplementary Figure 7 (related to Figure 7): Hierarchy of constitutive and cell type-specific architectural features at the Olig1-Olig2 locus.

Extended Experimental Procedures

5C Enrichments

Interaction size distribution

Cluster analysis

Gene Ontology analysis

Gene Expression Analysis

Hidden Markov Model for Domain Calling

Author Contributions

Supplementary References

Supplementary Figure Captions

Supplementary Figure 1 (related to Figure 1): Stem cell differentiation and characterization. (**A-I**) Phase contrast images of graded steps of differentiation from pluripotent ES cells to multipotent NPCs. (**A**) V6.5 ES cells were expanded on feeder layers in the presence of LIF, (**B**) purified from feeders with 2 passages on gelatin, and (**C**) formed into embryoid bodies in the absence of LIF. (**D**) Four day-old embryoid bodies were plated on tissue culture plastic and (**E-G**) cultured in chemically defined ITSFn media for (**E**) 3 days, (**F**) 5 days, (**G**) 8 days. (**H-I**) Nestin-positive NPCs were expanded in serum-free media containing bFGF/lamin for (**H**) 2 days and (**I**) 4 days. Scale bar, 200 μm. (**J-M**) qRT-PCR analysis during stages of ES differentiation. (**J**) Oct4, (**K**) Nanog, (**L**) Nestin, and (**M**) Sox2 transcripts are normalized to transcripts of the housekeeping gene GAPDH. (**N-P**) Confocal imaging of pluripotent markers in V6.5 ES cells. **(N)** Oct4, **(O)** Nanog, **(P)** Sox2. Scale bar, 50 μm. (**Q-U**) Confocal imaging of pluripotent and neural markers in ES-derived NPCs. **(Q)** Oct4, **(R)** Nanog, **(S)** Sox2, **(T)** Nestin, **(U)** βIII-tubulin. Scale bar, 50 μm.

Supplementary Figure 2 (related to Figure 1): 3C and 5C template characterization. (**A**) Serial dilutions of 3C template (ES replicate 1) were resolved on a 0.8% agarose gel. The band at ~10kb, with low signal at the top of the well, indicates a quality 3C template with relatively high ligation efficiency and restriction digest efficiency. (**B**) Quantification of the 3C template titration using conventional PCR. Serial dilutions of 3C template were analyzed with one anchor and three test primers interrogating fragments at 4.3 kb, 12.7 kb, and 43 kb distance from the anchor along the linear genome. Templates show a concentration-dependence in signal and a general decrease in signal as primer distance from the anchor increases. These results confirm a quality 3C template. (**C**) Agarose gel analysis of the ES replicate 1 5C library displaying a strong 100 bp band. Negative controls (no ligase, no 5C primers, no 3C template, water) do not have a 100 bp band, indicating effective ligation of forward-reverse primer pairs across 3C junctions. (**D**) HindIII digestion of the 5C library demonstrating that the 100 bp band resolves to 50 bp band after digest. This further

confirms the specificity and purity of the 5C library. **(E)** Alternating 5C primer design scheme. **(F)** Raw count heatmaps for two replicates ES cells and two replicates NPCs are displayed for Nanog, Sox2, Klf4, Olig-Olig2, Nestin, Oct4, and gene desert regions. Deep red pixels represent highest read counts. Yellow pixels represent low read counts. Grey pixels represent zero counts.

Supplementary Figure 3 (related to Figure 1): Heatmaps comparing Hi-C and 5C data. (A-D) Normalized 5C and Hi-C interaction frequencies represented as two-dimensional heatmaps for ~1 Mb regions around **(A)** *Sox2*, **(B)** *Nestin*, **(C)** *Klf4*, and **(D)** *Olig1-Olig2*. Hi-C data (adapted from (Dixon et al., 2012)) are displayed in the left half of each panel for mouse E14 ES cells (top) and mouse cortex (bottom). TADs reported in (Dixon et al., 2012) are represented as tracks for domain calls (blue bars) and directionality index (downstream bias (green), upstream bias (red)). 5C data generated in this work are displayed in the right half of each panel for mouse V6.5 ES cells (top) and ES-derived NPCs (bottom). Constitutive and cell type-specific sub-topologies called with our HMM model are represented as black lines overlaid on 5C heatmaps and as a directionality index displayed as a hierarchy of black wiggle tracks.

Supplementary Figure 4 (related to Figure 2): Progression of 5C data through the pipeline at the Sox2 region in ES cells. (**A,D,G,J**) Heatmaps representing interaction frequencies for all forward-reverse primer pair combinations at the Sox2 locus in ES cells. Each pixel represents the counts for a specific forward-reverse primer pair combination. Deep red pixels represent highest read counts. Yellow pixels represent low read counts. Grey pixels represent zero counts. (**B,E,H,K**) Counts plotted as a function of the genomic distance between fragments represented by ligated primer pairs. (**C,F,I,L**) Box plots demonstrate the median interaction frequency between an individual forward or reverse primer and all of its potential interacting partners. Count distributions for each primer are ordered by median interaction level. (**A-C**) Raw counts. (**D-F**) Raw counts after correction for primer-specific artifacts. (**G-I)** Primer-corrected counts normalized for distance-dependence background. (**J-L**) Interaction scores. (**M-P**) Raw versus predicted counts for ES replicate 1 across all regions. (**M**) Raw counts

show a non-uniform variance across predicted counts, with lower predicted counts leading to increased relative variance. A normal-log normal compound distribution was fit with a predicted count-dependent variance parameter. (**N-P**) For a given predicted count, a distribution could be determined and used to predict p-values for each raw count at that predicted count level. (**N**) low counts, (**O**) intermediate counts, (**P**) high counts. (**Q**) Raw 5C counts vs. fragment size.

Supplementary Figure 5 (related to Figures 5 and 6): Lentiviral shRNA characterization. (**A-C**) qRT-PCR analysis of (**A**) Smc1, (**B**) CTCF and (**C**) Med12 mRNA levels after transduction of V6.5 ES cells with lentiviral shRNA and 3-4 days puromycin selection. Transcript levels are normalized to expression of the housekeeping gene GAPDH. (**D**) Phase contrast images of V6.5 ES cells to monitor phenotype and viability changes during lentiviral shRNA experiments.

Supplementary Figure 6 (related to Figures 5 and 6): Constitutive and cell type-specific interactions for DNA FISH analysis. (A-B) Interaction profiles displaying primer-corrected 5C signal for a specific anchor fragment (shown at 0 bp in the center of the plot) vs. all other fragments throughout the Olig1-Olig2 region. Primer-corrected 5C counts are represented as a discrete black vertical line at each fragment. Interaction profiles are shown for two replicates of ES cells (top) and two replicates of NPCs (bottom). Blue line shows a rough estimate of the expected background level of interactions as computed using Loess smoothing. Red spheres demarcate the specific fragments involved in 3-D looping interactions with the anchor fragment. (**A)** 5C anchor was chosen as the HindIII fragment represented by reverse primer '5C_325_Olig1-Olig2_REV_139' vs. all fragments in the Olig1-Olig2 region represented by forward primers. Red spheres mark fragments represented by forward primers '5C_325_Olig1-Olig2_FOR_225' and '5C_325_Olig1-Olig2_FOR_227' involved in constitutive 3-D looping interactions with anchor '5C_325_Olig1-Olig2_REV_139'. (**B)** 5C anchor was chosen as the HindIII fragment represented by forward primer '5C_325_Olig1-Olig2_FOR_201' vs. all fragments in the Olig1-Olig2 region represented by reverse primers. A red sphere marks the fragment represented by reverse primer

'5C_325_Olig1-Olig2_REV_304' involved in an ES-specific 3-D looping interaction with anchor '5C_325_Olig1-Olig2_FOR_201'. **(C)** Arcplot of looping interactions compared to epigenetic marks in a ~330 kb region downstream of the *Olig1* gene. Constitutive interactions anchored by constitutive CTCF+Smc1 sites are shown in black and cell type-specific interactions anchored by ES-specific Smc1 sites are shown in red. Shaded grey bars highlight windowed fragments represented by '5C_325_Olig1-Olig2_REV_304' and '5C_325_Olig1-Olig2_FOR_201' found at the base of the ES-specific looping interaction shown in **(B)** between the *Olig1* TSS and a putative downstream ES-specific enhancer. The *Olig1* gene is highlighted in orange, while other genes at this locus are highlighted in green. Fragments represented by '5C_325_Olig1-Olig2_REV_304' and '5C_325_Olig1-Olig2_FOR_201' were used to design probes for the DNA FISH analysis shown in Figure 6. **(D)** Venn diagram comparing binding patterns for high-confidence ($P<1x10^{-8}$) Oct4, Sox2, and Nanog occupied sites in 5C regions. **(E)** Gene ontology analysis of genes co-localized with ES-specific Smc1 occupied sites enriched at the base of ES-specific interactions compared to a background of all genes in ES-specific interactions.

Supplementary Figure 7 (related to Figure 7): Hierarchy of constitutive and cell type-specific architectural features at the Olig1-Olig2 locus. TADs computed from Hi-C data (Dixon et al., 2012) are presented as tracks for domain calls (blue bars) for a 9 Mb region around *Olig1-Olig2* in E14 ES cells and cells isolated from mouse cortex. Within this larger 9 Mb region, a high-resolution view for the 1 Mb region around *Olig1-Olig2* is provided by two-dimensional heatmaps of 5C data in mouse V6.5 ES cells (top) and ES-derived NPCs (bottom). Constitutive and cell type-specific sub-domains called with our HMM model in both cell types are indicated with black lines overlaid on 5C heatmaps. Within the 1 Mb-sized 5C region, arcplots provide a locus-wide view of specific looping interactions called significant in ES cells and NPCs with out probabilistic model. Interactions are represented as a mirror image, with significant interactions in ES cells and NPCs displayed above and below the gene track, respectively. Constitutive interactions mediated by constitutive CTCF+Smc1 sites are displayed in black, while ES-specific interactions bridged by ES-specific Smc1 sites are shown in red. *Olig1* and

Olig2 genes are highlighted in orange, while other genes at this locus are highlighted in green.

Phillips-Cremins_et_al_2013
Supplementary_Table_1

Size of region	**# of fragments**	**# of potential interactions**
Oct4 (2.1 Mb)	317	25,110
Olig1-Olig2 (1.15 Mb)	274	18,768
Sox2 (1.0 Mb)	265	17,556
Klf4 (1.0 Mb)	251	15,750
Nestin (1.1 Mb)	205	10,504
Nanog (1.15 Mb)	189	8,930
Gene Desert (0.56 Mb)	50	625
Trans		504,726

Phillips-Cremins_et_al_2013
Supplementary_Table_5

Library Code	Replicate	Instrument	Lane	Total Reads	PE1 Mapped Reads	PE2 Mapped Reads
ES 1	1	Illumina GA2	1	8603598	7423512	6657000
			2	10115195	8709373	7769788
ES 2	2	Illumina GA2	1	9816055	8486783	7704521
			2	9848146	8485678	7457185
NPC 1	1	Illumina Hi-Seq	1	70829342	25453780	24185269
NPC 2	2	Illumina Hi-Seq	1	95559621	59969460	58640003

Phillips-Cremins_et_al_2013
Supplementary_Table_6

Antibody	Cell Type	Mapped Test ChIP-Seq reads	p-value threshhold	# of occupied sites in 5C regions	Test ChIP Reference	Test Sample GEO ID	Control Samples	Mapped Control ChIP-Seq reads	Control Sample GEO ID
Med12	mES (V6.5)	22,938,650	1E-8	419	M.H. Kagey, et al[1]	GSM560345, GSM560346	mES Whole Cell Extract	27,002,573	GSM747546
Smc1	mES (V6.5)	24,028,494	1E-8	427	M.H. Kagey, et al[1]	GSM560341, GSM560342	mES Whole Cell Extract	27,002,573	GSM747546
CTCF	mES (159-2)	9,562,677	1E-8	391	Stadler, et al[2]	GSM747534	mES Whole Cell Extract	10,202,630	GSM747545
CTCF	ES-derived NPC	13,641735	1E-8	286	this study	Present work Paired-end 1	NPC Whole Cell Extract	14,041,323	Present work Paired-end 1
Smc1	ES-derived NPC	10,922433	1E-8	229	this study	Present work Paired-end 1	NPC Whole Cell Extract	14,041,323	Present work Paired-end 1
H3K4me1	mES (V6.5)	11,437,522	1E-8	309	M.P. Creyghton, et al[4]	GSM594577	V6.5 Whole Cell Extract	14,682,811	GSM307154, GSM307155, GSM594599
H3K4me1	ES-derived NPC	4,471,210	1E-8	150	A. Meissner, et al[5]	GSM281693	NPC Whole Cell Extract	4,369,951	GSM307617
H3K4me3	mES (V6.5)	6,809,878	1E-8	172	T.S. Mikkelsen, et al[3]	GSM307618	V6.5 Whole Cell Extract	6,008,440	GSM307154, GSM307155
H3K4me3	ES-derived NPC	3,397,613	1E-8	121	T.S. Mikkelsen, et al[3]	GSM307613	NPC Whole Cell Extract	4,369,951	GSM307617
H3K27ac	mES (V6.5)	11,128,384	1E-8	356	M.P. Creyghton, et al[4]	GSM594579 Rep2	V6.5 Whole Cell Extract	14,682,811	GSM307154, GSM307155, GSM594599
H3K27ac	ES-derived NPC	8,831,628	1E-8	187	M.P. Creyghton, et al[4]	GSM594585	NPC Whole Cell Extract	14,041,323	Present work Paired-end 1
Oct4	mES (V6.5)	3,951,875	1E-8	109	A. Marson, et al[6]	GSM307137 Rep1	V6.5 Whole Cell Extract	3,517,916	GSM560357
Sox2	mES (V6.5)	3,936,527	1E-8	68	A. Marson, et al[6]	GSM307138 Rep1	V6.5 Whole Cell Extract	3,517,916	GSM560357
Nanog	mES (V6.5)	3,644,219	1E-8	83	A. Marson, et al[6]	GSM307141 Rep2	V6.5 Whole Cell Extract	3,517,916	GSM560357
Med12	mES (V6.5)	22,938,650	1E-4	683	M.H. Kagey, et al[1]	GSM560345, GSM560346	mES Whole Cell Extract	27,002,573	GSM747546
Smc1	mES (V6.5)	24,028,494	1E-4	753	M.H. Kagey, et al[1]	GSM560341, GSM560342	mES Whole Cell Extract	27,002,573	GSM747546
CTCF	mES (159-2)	9,562,677	1E-4	532	Stadler, et al[2]	GSM747534	mES Whole Cell Extract	10,202,630	GSM747545
CTCF	ES-derived NPC	13,641735	1E-4	394	this study	Present work Paired-end 1	NPC Whole Cell Extract	14,041,323	Present work Paired-end 1
Smc1	ES-derived NPC	10,922433	1E-4	388	this study	Present work Paired-end 1	NPC Whole Cell Extract	14,041,323	Present work Paired-end 1
H3K4me1	mES (V6.5)	11,437,522	1E-4	567	M.P. Creyghton, et al[4]	GSM594577	V6.5 Whole Cell Extract	14,682,811	GSM307154, GSM307155, GSM594599

H3K4me1	ES-derived NPC	4,471,210	1E-4	282	A. Meissner, et al[5]	GSM281693	NPC Whole Cell Extract	4,369,951	GSM307617
H3K4me3	mES (V6.5)	6,809,878	1E-4	209	T.S. Mikkelsen, et al[3]	GSM307618	V6.5 Whole Cell Extract	6,008,440	GSM307154, GSM307155
H3K4me3	ES-derived NPC	3,397,613	1E-4	150	T.S. Mikkelsen, et al[3]	GSM307613	NPC Whole Cell Extract	4,369,951	GSM307617
H3K27ac	mES (V6.5)	11,128,384	1E-4	516	M.P. Creyghton, et al[4]	GSM594579 Rep2	V6.5 Whole Cell Extract	14,682,811	GSM307154, GSM307155, GSM594599
H3K27ac	ES-derived NPC	8,831,628	1E-4	228	M.P. Creyghton, et al[4]	GSM594585	NPC Whole Cell Extract	14,041,323	This work Paired-end 1
Oct4	mES (V6.5)	3,951,875	1E-4	267	A. Marson, et al[6]	GSM307137 Rep1	V6.5 Whole Cell Extract	3,517,916	GSM560357
Sox2	mES (V6.5)	3,936,527	1E-4	136	A. Marson, et al[6]	GSM307138 Rep1	V6.5 Whole Cell Extract	3,517,916	GSM560357
Nanog	mES (V6.5)	3,644,219	1E-4	144	A. Marson, et al[6]	GSM307141 Rep2	V6.5 Whole Cell Extract	3,517,916	GSM560357

[1]Kagey, M.H. et al., Mediator and cohesin connect gene expression and chromatin architecture, Nature 467, 430–435 (2010).
[2] Stadeler, M.B. et al., DAN-binding factors shap the mouse methylome at distal regulatory regions. Nature 480, 490-495 (2011).
[3] Mikkelsen, T.S. et al., Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature 448, 553-560 (2007).
[4] Creyghton, M.P. et al., Histone H3K27ac separates active from poised enhancers and predicts developmental state. PNAS 107 (50), 21931-21936 (2010).
[5] Meissner, A. et al., Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature 454 (7205), 766-770 (2008).
[6] Marson, A. et al., Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. Cell 134 (3): 521-533, 2008.

Phillips-Cremins_et al_2013
Supplementary_Table_7

Olig1/2 Constitutive Loop		
Pimer Name	**Primer sequences**	**Coordinates**
Olig_CONST 1A FP	TCCCTGGTATTTGAACCTGTGGCT	chr16:91,068,272-91,068,296
Olig_CONST 1A RP	TCTCAGAGATCTGCTTGGCGTTGT	chr16:91,079,564 - 91,079,588
Olig_CONST 1B FP	TTAGGATCAGAATAAACTGTCACGGTAG	chr16:91,352,244 - 91,352,272
Olig_CONST 1B RP	GATACCTTTTGGGGCAGTGTAGTTTCAG	chr16:91,362,170 - 91,362,198
Olig1/2 ES-specific Loop		
Pimer Name	**Primer sequences**	**Coordinates**
Olig_ES 2A FP	GTCAGGGTAAGCATCAGGATAAG	chr16:91,257,516-91,257,539
Olig_ES 2A RP	CATCCGTCTGAAGAGCAGTATC	chr16:91,267,523-91,267,545
Olig_ES 2B FP	ACATGCACGTTCACATACGCACAC	chr16:91,556,201-91,556,225
Olig_ES 2B RP	TCGCCAACCAAGATGTCGGATACA	chr16:91,564,742-91,564,766

Extended Experimental Procedures

ES cell culture

Murine V6.5 ES cells (genotype *129SvJae x C57BL/6*; male) were procured from Novus Biologicals (Littleton, CO) at passage 18 and expanded at 37° in 5% CO2 on Mitomycin-C-inactivated MEF feeder layers. ES cell expansion media consisted of Dulbecco's Modified Eagle Medium (DMEM) supplemented with 15% fetal bovine serum (FBS, GIBCO), 10^3 U/ml leukemia inhibitory factor (ESGRO, Millipore), 1x nonessential amino acids (Lonza), 0.1 mM 2-mercaptoethanol, 4 mM L-glutamine (Cellgro), and 1x penicillin/streptomycin (Cellgro). Media was exchanged every 2 days and the cells were passaged at approximately 70% confluence. After initial expansion, ES cells were passaged 2-3 times on 0.1% gelatin to remove contaminating feeder cells. At the time of fixation for downstream assay, cells were ~70-75% confluent.

ES cell differentiation

V6.5 ES cells were differentiated into neural progenitor cells (NPCs) using established techniques (Meissner et al., 2008; Mikkelsen et al., 2007; Okabe et al., 1996). Briefly, after expansion, ES cells were trypsinized and cultured using rotary orbital motion (Carpenedo et al., 2007) in bacterial dishes for 4 days in the absence of LIF to promote embryoid body formation. Embryoid bodies were then plated on tissue-culture plastic dishes and after 24 hours media was changed to serum-free defined media consisting of DMEM/F12 (Life Technologies) supplemented with 5 ug/ml insulin (Sigma), 50 ug/ml human APO transferrin (Sigma), 30 nM sodium selenite (Sigma), 2.5 ug/ml human plasma fibronectin (Gibco), and 1x penicillin/streptomycin (Cellgro). After 5-7 days selection in ITSFn, adherent cells were trypsinized, triturated to a single cell suspension, and re-plated on tissue culture dishes coated with 15 ug/ml poly-L-ornithine (Sigma) and 1 ug/ml human plasma fibronectin (Gibco). Cells were then further propagated for 2-4 days in DMEM/F12 supplemented with 25 ug/ml insulin, 50 ug/ml human APO transferrin, 30 nM sodium selenite, 20 nM progesterone (Sigma), 100 nM putrescine (Sigma), 1 ug/ml laminin (Sigma), 10 ng/ml Fgf2 (R&D Systems), and 1x penicillin/streptomycin. FGF2 was added daily to promote NPC proliferation.

Immunofluorescence Staining

Cells were cross-linked in fresh 4% paraformaldehyde for 30 minutes and washed 1-2x with PBS. After permeabilization with 0.1% Triton X-100 for 10 minutes, cells were washed with 0.1% Tween20 in PBS for 10 minutes and then incubated in blocking buffer (PBS with 5% bovine serum albumin (BSA)) for 60 minutes. Next, cells were stained overnight at 4ºC with antibodies for Oct4 (Santa Cruz Biotechnology, polyclonal rabbit anti-Oct3/4, H-134, sc-9081), Nanog (Millipore, polyclonal rabbit anti-Nanog, AB5731), Sox2 (R&D Systems, monoclonal mouse anti-Sox2, MAB2018), Nestin (Millipore, monoclonal mouse anti-Nestin, MAB353), or βIII-Tubulin (Covance, monoclonal mouse anti-β-Tubulin, TUJ1, MMS-435P). Chamber slides were washed 3x for 10 minutes in 0.1% Tween20 in PBS and then incubated for 4 hours at room temperature with either goat anti-mouse-conjugated Alexa Fluor 488 (Invitrogen) or goat anti-rabbit-conjugated Alexa Fluor 555 (Invitrogen). Finally, cells were washed again 2x for 15 minutes in 0.1% Tween20 in PBS, 3x for 15 minutes in PBS, and then mounted in mounting media with DAPI.

qRT-PCR

mRNA transcripts were quantified using qRT-PCR according to standard methods (Carpenedo et al., 2007; Phillips et al., 2008; Phillips et al., 2006). Briefly, RNA was extracted from ES cells, stage 3 NPCs, and stage 4 NPCs with an RNeasy Mini kit (Qiagen Inc, Valencia, CA). Reverse transcription for complementary DNA synthesis was performed with 1 µg of RNA per sample using the SuperScript First Strand Synthesis System for RT-PCR (Invitrogen). Gene expression was assayed with quantitative RT-PCR using the MyIQ cycler (BioRad) as described (Carpenedo et al., 2007). Primers used to evaluate cell type-specific genes were designed with Beacon Designer software as follows: Oct4 (Forward primer, 5'-CCGTGTGAGGTGGAG-3'; Reverse primer, 5'-GCGATGTGAGTGATCTGC-3'), Nestin (Forward primer, 5'-GGAGAAGCAGGGTCTACA-3'; Reverse primer, 5'-AGCCACTTCCAGACTAAG-3'), Nanog (Forward primer, 5'-GAAATCCCTTCCCTCGCCAT-3'; Reverse primer, 5'-CTCAGTAGCAGACCCTTGTAAG-3'), Sox2 (Forward primer, 5'-CCGTGGTTACCTCTTCCTC-3'; Reverse primer, 5'-GCCCCAGGGATGATCTAAGC-

3'), and GAPDH (Forward primer, 5'- GCCTTCCGTGTTCCTAC-3'; Reverse primer, 5'-GCCTGCTTCACCACCTT-3'). Transcript concentrations were calculated using standard curves and normalized to GAPDH expression levels.

Lentiviral shRNA characterization

Lentiviral shRNA plasmids cloned into the pLKO.1 vector were purchased from Open Biosystems (Thermo Scientific). Specific clones used in this study include: CTCF shRNA (ID: TRCN0000039019, shRNA sequence: ATTACCAACTACTTTCTCTGC); Med12 shRNA (ID: TRCN0000096466, shRNA sequence: ATCCTGAAACATGAACAAGGC); and Smc1 shRNA (ID: TRCN0000109033, shRNA sequence: TTTATCTGTTCAAATGCCTGC). Lentiviral particles were produced using the TransLenti Viral Packaging Mix (TLP4615, Open Biosystems) and the Arrest-In Transfection Reagent in H293T cells as described in the kit manual. Virus was harvested by collecting H293T medium, spinning at 500 g for 10 minutes at room temperature, and transferring supernatant to a new 15 mL conical tube. One volume of Lenti-X Concentrator solution (Clontech) was added to three volumes of supernatant, incubated at 4° C overnight, and then centrifuged at 1500 g at 4° C for 45 minutes to pellet virus particles. Pellet was resuspended in 1 mL DMEM, aliquoted, and stored at -80° C until use. Viral titer was calculated as ~10^6 $^-10^7$ using a lentiviral titer kit from Mellgenlabs (http://mellgenlabs.com/Documents/lentititer_protocol.htm).

Lentiviral transductions

V6.5 Murine ES cells were expanded on 100 mm tissue culture polystyrene petri dishes (Corning, Corning, NY) containing confluent Mitomycin-C treated MEFs. After expansion, ES cells were passaged 1x onto gelatin-coated petri dishes to purify feeders and then seeded at a density of 2x10^6 cells per 100 mm plate. Two days after seeding, ES cell media with 6 µg/ml hexadimethrine bromide (polybrene, Sigma) was added to the cells. After 15 minutes of incubation at 37° C, lentiviral particles (viral titer of ~10^6 $^-10^7$ TU/µl, volume for MOI of 5:1) were gently added to the cells and incubated overnight. The media was then replaced with new ES cell medium for an additional 24

hours. The following day, and for each of the next 3 days, media was exchanged with ES cell media containing 2-3.5 µg/ml puromycin (Gibco).

Slides for DNA FISH

To prepare cells for DNA FISH, a single cell suspension of ES cells was obtained by dissociating V6.5 monolayers using 0.05% Trypsin/EDTA (Mediatech). After centrifugation, the cell pellet was resuspended in 5 ml of 4% paraformaldehyde to fix for 10 minutes. Cells were then pelleted again and quenched with 0.1M Tris-HCl (Sigma) for 10 minutes, followed by one wash in PBS. After counting with a hemacytometer, cell density was adjusted to $2x10^6$ cells/ml. Finally, 150 µL droplets of the cell suspension were added to poly-l-lysine coated slides (Sigma). Slides were incubated at room temperature until all droplets dried on the slides and all samples were stored in PBS containing 0.1% sodium azide (Sigma) until further analysis.

DNA FISH

Three-color DNA FISH was carried out according to procedures described previously (Guo et al., 2011) in wild type V6.5 ES cells, ES cells after lentiviral shRNA for CTCF, ES cells after lentiviral shRNA for Med12, and ES cells after lentiviral shRNA for Smc1. Position-specific 10 kb probes were amplified by long-range PCR using BAC templates with primers listed in **Table S7**. 10 kb FISH probes for fragment A and fragment B anchoring the base of a specific looping interaction were labeled with Alexa Fluor 594 (red) and 488 (green), respectively. BACs were used as anchors and labeled with Alexa Fluor 697 (blue). All probes and BACs were hybridized to ES cell slides prepped as described above. Signals were visualized by epifluorescence microscopy using a Nikon T2000 instrument. 20-40 0.2 µM Z-sections were recorded followed by deconvolution using NIS-Elements software. Distances between red and green probes were measured according to published procedures (Guo et al., 2011) and divided into 5 categories (<0.2, 0.2-0.5, 0.5-0.8, 0.8-1.0 µm) for ~100-130 alleles. The percentage of alleles in each category was quantified and compared for each genotype.

3C template generation and characterization

3C templates were generated for ES (n=2) and NPC (n=2) pellets using HindIII as previously described (Dekker et al., 2002; Gheldof et al., 2010; van Berkum and Dekker, 2009). Briefly, $5x10^7$-$1x10^8$ cells were cross-linked in PBS in presence of formaldehyde to a final concentration of 1% for 10 minutes. Cross-linking was terminated by the addition of 3M glycine to a final working concentration of 125 mM. Quenching was initiated for 5 minutes at room temperature followed by 15 minutes at 4° C. Cross-linked cells were harvested by scraping and then centrifuged for 10 minutes at 400g. Pellets were snap frozen and stored at -80°C.

At the time of the experiment, cross-linked pellets were re-suspended in lysis buffer (10mM Tris-HCl pH=8.0, 10mM NaCl, 0.2% Igepal CA-630) supplemented with protease inhibitors (Sigma, St. Louis, MO) and a final concentration of 2mM EDTA, 2mM EGTA, and 250 µM DNase g inhibitor 6-DTAF (Anaspec, Fremont, CA) and incubated for 30 minutes on ice. Cells were lysed using a 2mL dounce homogenizer, washed with 1X NEBuffer2 buffer (10mM Tris-HCl, 50mM NaCl, 10mM $MgCl_2$, 1mM DTT), and re-suspended in NEBuffer2. To solubilize chromatin, SDS was then added to a final concentration of 0.1% and lysates were incubated at 65° for 10 minutes. Triton X-100 was then added to a final concentration of 1% to quench SDS. To digest chromatin, each individual 444 uL aliquot of solubilized chromatin was then incubated with 400U HindIII (NEB) overnight at 37° C with shaking. HindIII was inactivated by incubating lysates at 65° for 30 minutes after addition of SDS to a final concentration of 1.56%.

Ligation was performed under dilute conditions that promote intramolecular ligation at 16° C for 2 hour in ligation buffer (1% Triton X-100, 0.1mg/mL BSA, 1mM ATP, 50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 10 mM DTT) with 10 uL of T4 DNA ligase (Invitrogen). To reverse cross-links, samples were then treated with 63.5 µg/mL Proteinase K (Invitrogen) at 65° C. Four hours later, Proteinase K was added again to 127 µg/mL and then incubated overnight at 65° C. DNA was purified by subjecting samples to a series of phenol and phenol-chloroform extractions before precipitation with ethanol. Pellets were re-suspended in 1-2 mL TE Buffer and subjected to three further rounds of phenol-chloroform extraction before precipitation again with ethanol. Pellets were washed 8-12 times with 70% Ethanol before resuspension in 100-500 uL TE Buffer and subsequent treatment for 3 hour with 100 µg/mLRNase A at 37° C. 3C

template concentrations were calculated using LabWorks image analysis software (UVP, Upland, CA) (**Figure S2A**).

In order to confirm the quality of the 3C templates generated, conventional PCR was performed with an anchor primer (5'-GCACACAGCGCTTACCTTGGAGAGATTTTG-3') representing a fragment in the gene desert region (mm9; chr5: 133242078-133800000) and a series of four test primers representing fragments at 4.3 kb (5'-GGATGAGGACGCTTTAGACGTATTCTCCAG-3'), 12.7 kb (5'-AACAGAGCTAGACGTTTTGGCTGGAGTAGC-3'), and 41.5 kb (5'-TGTAGTGGAAGGACGCTTCCTCAGACCTT-3') distance on the linear genome from the anchor primer. Intensity of the PCR signal was dependent on the concentration of the 3C template and also inversely proportional to the genomic distance between the anchor and test primers (**Figure S2B**). These phenomena were observed for all four templates created in this study, indicating that the digestion and ligation procedures were successful.

5C primer design

5C primers were designed at HindIII restriction sites using the my5Csuite primer design tools (Lajoie et al., 2009). An alternating scheme was pursued in which reverse and forward primers were designed against every other fragment (**Figure S2E**). Thresholds used in primer design include: U-BLAST, 3; S-BLAST, 50; 15-MER: 800; MIN_FSIZE, 100; MAX_FSIZE, 50,000; OPT_TM, 65; OPT_PSIZE, 30. Primers were excluded if unique mapping could not be achieved for fragments spanning highly repetitive sequences. The universal T7 sequence was tethered to all forward primers (5'-TAATACGACTCACTATAGCC-3') and the reverse complement to the universal T3 sequence was tethered to all reverse primers (5'-TATTAACCCTCACTAAAGGGA-3'). In total, 768 forward primers and 783 reverse primers were designed, spanning 7 genomic regions ranging in size from 600kb to 2Mb (**Table S2, Table S3**). The 5' end of each reverse primer was modified with a phosphate group by incubating the reverse primer pool with T4 polynucleotide kinase. Finally, primers were pooled so that each 5C primer was present at a stock equimolar concentration of 5 fmol/uL.

5C library generation

5C experiments were performed as previously described (Bau et al., 2011; Dostie and Dekker, 2007; Dostie et al., 2006; van Berkum and Dekker, 2009) to amplify from the 3C template the subset of chromatin interactions that occur within only pre-selected regions of interest. First, primers were annealed to the 3C template at 48° C for 16 hours or overnight. Each multiplexed annealing reaction contained 1x NEBuffer4, 3C template (1500 ng ES cell DNA, 6000 ng NPC DNA), and 1 fmol of each 5C primer. Next, pairs of forward and reverse annealed primers were nick-ligated in 1x Taq ligase buffer with 10 U *Taq* ligase for 60 minutes at 48° C. Forward primers were designed to bind directly upstream and reverse primers to bind directly downstream of a given restriction site in the 3C ligation product. Therefore, the annealing and ligation steps result in a library of ligated primer-pairs that represent, in principle, a carbon-copy of only 3-D interactions between genomic loci within the regions of interest.

5C libraries were then selectively amplified by leveraging the principles of ligation-mediated amplification. Forward primers contain a universal T7 sequence and reverse primers contain the reverse complement to the universal T3 sequence. Thus, amplification of the carbon-copy library in a high-throughput and massively parallel manner was achieved by PCR with 30 cycles and use of universal T7 (5'-TAATACGACTCACTATAGCC-3') and T3 (5'–TATTAACCCTCACTAAAGGGA-3') primers. At least 4-5 independent amplification reactions were performed for each annealing reaction. PCR products for each biological replicate were pooled and then concentrated with the Qiaquick PCR purification kit (Qiagen, Germany). Negative controls confirmed the specificity of the ligated 5C library, including: no 5C primers, no ligase, no 3C template (3000 ng of salmon sperm DNA in place of template), and water only (**Figure S2C**). Finally, >90% digestion of the 100 bp ligated library with HindIII confirmed the presence of a newly established restriction site after ligation (**Figure S2D**). These results confirm the quality and purity of the newly purified 5C library.

Sequencing and Initial Data Processing

To prepare libraries for annealing to the sequencing flow cell, the 100 bp band representing the 5C library was size selected from a 2% agarose gel and purified with a

QIAquick gel purification kit (Qiagen). For ES cell libraries, a 3'-adenine was added using dATP and Taq DNA polymerase, followed by subsequent ligation to Illumina paired-end adaptor oligonucleotides (Illumina, San Diego, CA). Adaptor-modified libraries were then linkered with Illumina PCR primers PE 1.0 and 2.0 using 18 cycles of PCR. DNA was size selected on a 2% agarose gel, purified with a QIAquick gel purification kit (Qiagen) and then submitted for paired-end sequencing using the Illumina GA2 platform at Emory University. For NPCs, 5C libraries were sent directly to Hudson Alpha Institute for Biotechnology (Huntsville, AL) for linkering and sequencing on the Illumina Hi-seq platform.

A summary of sequencing details for each biological replicate is provided in **Table S5**. Reads were aligned to a pseudo-genome consisting of all 5C primers (**Table S2, Table S3**) using Bowtie (http://bowtie-bio.sourceforge.net/index.shtml) (Langmead, 2010). To account for poor quality reads, sequences were required to have only one unique alignment and 5 and 3 bases were trimmed from the 5' and 3' ends of the read, respectively. After mapping, interactions were counted when both paired end reads could be uniquely mapped to the 5C primer pseudo-genome. Only interactions between forward-reverse primer pairs were tallied as a true count because forward-forward or reverse-reverse primer pairs represent an artifact in the 5C procedure. Overall, a total of 97,243 cis and 504,726 trans interactions were queried with an alternating primer design spanning the selected regions of interest (**Table S1**). Primers showing counts >100,00 total reads or <100 total reads were deemed outliers and removed from subsequent analyses. Removed primers are listed in **Table S4**.

Sources of systematic bias in 5C data

The 3C-derived Hi-C technique has several systematic biases that must be accounted for to accurately assess interactions (Imakaev et al., 2012; Yaffe and Tanay, 2011). Similarly, we find that 5C also exhibits systematic biases that will ultimately yield false positive and false negative interaction frequencies not due to true biological signal (Sanyal et al., 2012). Bias could be introduced at any stage in the 5C experimental procedure. Differences in cross-linking or restriction digest efficiency can occur between technical and/or biological replicates. Furthermore, the size of the fragment has shown

a non-linear correlation with contact probability, likely due to non-specific effects of fragment size on ligation efficiency (Yaffe and Tanay, 2011). Here we see very similar digestion efficiencies between biological replicates, suggesting that this bias is not dominant in our procedure. Furthermore, to address bias related to ligation efficiency, we only used primers that map to fragments with a size range previously proven to be optimal for the 5C procedure (100bp – 50,000bp) (Gheldof et al., 2010; Lajoie et al., 2009). We also focused our analysis only on cis interactions that may not show the same bias (Sexton et al., 2012; Yaffe and Tanay, 2011). Finally, we do not observe a marked correlation between fragment size and counts in the current experiments (unpublished data). Overall, these sources of error are common to other molecular approaches and have been adequately addressed through experimental optimization and through use of technical replicates and negative controls.

Error unique to the 5C technique is introduced during the ligation-mediated amplification process. For example, primer-specific artifacts or variation in the concentration of genome copies in the 5C reaction can lead to amplification bias. Primer-specific artifacts could be due to differential primer annealing affinity, non-specific binding to alternative sites in the genome, differential primer ligation efficiency, or sequence-specific differences in polymerase tracking. Furthermore, bias related to the nucleotide composition of each fragment can be introduced during addition of Illumina adaptors and/or sequencing (Aird et al., 2011). In testing for these possible sources of error, we have observed a strong relationship between each set of primers interrogating a given interaction and the observed count. Evidence for this effect is easily observed in raw count heatmaps. Primer-specific effects appear as bands (or stripes), with an entire row or column showing increased or decreased counts (**Figure S4A**). Box plots demonstrate a continuous 50- to 100-fold variation in the median interaction frequency between a single primer and all possible mates within each region (**Figure S4C**), suggesting that sequence-specific artifacts independent from the biology have a dominant effect on 5C signal.

Probabilistic modeling of 5C interaction maps

Primer-Effects

To account for primer-specific artifacts, we have constructed a probabilistic model for the observed counts in a 5C experiment. The genomic distance between two loci and the probe-specific effects directly influence the observed count and are best modeled simultaneously. The expected *log* count $E(t_{ab})$ between probes a and b is modeled as:

$$E(t_{ab}) = e^{f_a + f_b + H(d_{ab}) + \mu_n}$$

where f_a and f_b are the two probe effects, $H(d_{ab})$ describes the expected background interaction of two fragments whose distance between midpoints is d_{ab}, and μ_n is the mean log-count for the region containing the two probes. While it is possible that some interaction effect may exist between different sets of probes, we found that an additive effect on the expected log-count yielded the best fit to the data without introduction of additional model complexity.

Distance-Dependence

The distance dependent contribution to the expected interaction level is modeled using a Weibull distribution:

$$H(d_{ab}) = \beta_n ln(e^{\lambda_n} Weibull(d_{jk}, \gamma_n, e^{\lambda_n})$$

where γ_n and λ_n are shape parameters and β_n is a scaling factor. Although an inverse power-law relationship has been previously proposed between distance and interaction (Rousseau et al., 2011), the tiling primer design used here interrogates a large number of short-range interactions, for which we find the power-law relationship does not hold.

Count Variance

For assessing the significance of interactions, it is important to accurately capture sources of variance in the observed count. Each step of the 5C procedure potentially

contributes error; and error introduced prior to amplification contributes exponentially to variability. The shape of the distribution and magnitude of its variance also varies greatly depending on the strength of the signal, possibly because of differences in sequencing efficiency for very infrequent interactions (**Figure S4N-P**). This results in a distribution composed on Normal and Log-Normal components:

$$t_{ab} \sim Normal(E(t_{ab}), \alpha E(t_{ab})^2 + \nu)\, e^{Normal(0,\rho)}$$

where ρ is the variance introduced prior to amplification, α is the post-amplification variance, and ν is the amplification-associated variance.

Model Optimization

For each region, model parameters describing the expected value and signal-dependent variance of each interaction under a normal-lognormal distribution were learned using stochastic gradient descent. Using the learned parameters, empirical p-values were computed under this distribution via Monte Carlo simulation for each observed interaction count. Interaction scores were derived using the inverse cumulative density function for a standard normal distribution. The resulting p-values and derived interaction scores are directly comparable across regions and datasets.

5C Peak-Calling Code

Code for 5C peak calling pipeline can be found at: https://bitbucket.org/bxlab/phillips-cremins_cell_2013.

Progression of 5C data through the pipeline

Using the Sox2 locus as a case study, we display count data in ES cells to illustrate the progression of data through each stage of the pipeline: raw reads (**Figure S4A-C**), raw reads corrected for primer-specific artifacts (**Figure S4D-F)**, primer-corrected data after removal of the distance-dependent effect (**Figure S4G-I**), and finally normalized data converted to interaction scores (**Figure S4J-L**). For all

downstream analyses, we restricted our attention to the strongest fragment-to-fragment interactions that are reproducible.

Thresholds for significant interactions

Only chromatin interactions with reproducibly high interaction scores in both replicates were subjected to further analysis. Reproducible interaction scores ≥ 1.751 in both replicates of ES cells and NPCs were parsed into a "constitutive looping interactions" group. Interaction scores reproducibly ≥ 2.576 in both ES cell replicates and < 1.644 in both NPC replicates were parsed into an "ES only looping interactions" group, while interaction scores reproducibly ≥ 2.409 in both NPC replicates and < 1.644 in both ES cell replicates were parsed into a "NPC only looping interaction" group. All queried chromatin contacts with interaction scores < 0.524 in all replicates were parsed into the "non-looping background" category.

By randomly permuting data (ES1 with NPC1 vs. ES2 with NPC2), the large majority of cell-type specific chromatin interactions were lost, suggesting that thresholds were sufficiently rigorous to minimize false positives. For ES-specific interactions, an empirical false discovery rate (eFDR) was computed to be 9.6% by taking the ratio of the number of loops in Figure 2A (with interaction scores reproducibly ≥ 2.576 in both ES rep1 and ES rep2 and reproducibly < 1.644 in both NPC rep1 and NPC rep2) to the number of loops in Figure 2B (with interaction scores reproducibly ≥ 2.576 in ES rep1 and NPC rep1 and reproducibly < 1.644 in ES rep2 and NPC rep2). For NPC-specific interactions, an empirical false discovery rate (eFDR) was computed to be 5.5% by taking the ratio of the number of loops in Figure 2A (with interaction scores reproducibly ≥ 2.409 in both NPC rep1 and NPC rep2 and reproducibly < 1.644 in both ES rep1 and ES rep2) to the number of loops in permuted Figure 2B (with interaction scores reproducibly ≥ 2.409 in ES rep1 and NPC rep1 and reproducibly < 1.644 in ES rep2 and NPC rep2).

Finally, we also note that no significant looping interactions above the expected background signal were detected in the gene desert negative control region, suggesting that the thresholds used in this analysis were sufficiently rigorous in removing non-specific background signal. By applying these stringent thresholds, only cell type-

specific loops corresponding to the top 0.096% and 0.190% of all queried long-range intra-chromosomal interactions in ES and NPC libraries, respectively, were considered for downstream analysis.

Chromatin Immunoprecipitation

Chromatin immunoprecipitation (ChIP) was performed as previously described with minor modifications (Kagey et al., 2010). $50x10^6$ NPCs were lysed in 5ml of LB1 (50 mM Hepes-KOH, pH 7.5, 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% NP-40, 0.25% Triton X-100), followed by LB2 (10 mM Tris-HCl, pH 8.0, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA). Sonication was carried out in LB3 (10 mM Tris-HCl, pH 8.0, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% Na-Deoxycholate, 0.5% N-lauroylsarcosine) with Branson Sonifier 250 for 30 x 12 sec pulses (output set between level 3 and 4). The resulting whole cell extract was incubated with Protein A Sepharose for 8 hours at 4°C. Pre-cleared extract was then incubated with 160 µl (50% v/v) of Protein A sepharose that had been pre-incubated with approximately 10 µg of the appropriate antibody overnight at 4°C. For Smc1a ChIP using Bethyl Laboratories (A300-055A) affinity purified rabbit polyclonal antibody, beads were washed 1X with LB3, 1X with 20mM Tris-HCl pH8, 500mM NaCl, 2mM EDTA, 0.1% SDS, 1%Triton X-100, 1X with 10mM Tris-HCl pH8, 250nM LiCl, 2mM EDTA, 1% NP40 and 1X with TE containing 50 mM NaCl. For CTCF ChIP using an Upstate 07-729 rabbit polyclonal antibody, beads were washed 7x with RIPA buffer and 1X with TE containing 50 mM NaCl. Bound complexes were eluted from the beads (50 mM Tris-HCl, pH 8.0, 10 mM EDTA and 1% SDS) by heating at 65°C for 15 min with occasional vortexing. Cross-links were reversed by overnight incubation at 65° C. Whole cell extract DNA reserved from the sonication step was also treated for crosslink reversal. Immunoprecipitated DNA and whole cell extract DNA were treated with RNaseA and Proteinase K. DNA was purified by phenol:chloroform:isoamyl alcohol extraction. ChIP-seq libraries were prepared for sequencing as previously described (Wood et al., 2011).

ChIP-seq Data Processing

A summary of all ChIP-seq data sets analyzed in this study is provided in **Table S6**. Data was downloaded from GEO (http://www.ncbi.nlm.nih.gov/geo/) and reanalyzed according to the following methodology: Sequences were aligned to NCBI Build 37 (UCSC mm9) using default parameters (-v1 -m1) in Bowtie. Only sequences that mapped uniquely to the genome were used for further analysis. Model-based Analysis for ChIP-Sequencing (MACS) was used for peak calling (http://liulab.dfci.harvard.edu/MACS/00README.html) (Zhang et al., 2008). For transcription factor or architectural protein ChIP-seq (e.g. CTCF, Smc1, Med12, Oct4, Nanog, and Sox2), default parameters were used with a p-value cutoff of P<1E-8 or P<1E-4. For histone modification ChIP-seq (e.g. H3K4me1, H3K27ac, H3K4me3), the model building step was skipped (by calling the parameter -- no model), but the local background estimation step was kept in the analysis performed at p-value cutoffs of P<1E-8 or P<1E-4. Background control files used to assess significance are listed in **Table S6**. After peak calling, only statistically significant occupied sites in the 5C regions-of-interest were considered further. Genomic coordinates for parsing ChIP-seq data include: Olig1-Olig2 (chr16: 90611386-91761386), Nestin (chr3: 87277995-88377995), Oct4 (chr17: 34428606-36528892), Nanog (chr6: 122184399-123334399), Sox2 (chr3: 34107373-35107373), Klf4 (chr4: 54891772-55891772), and gene desert control (chr5: 133242078-133800000).

Parsing Architectural Protein Subclasses

Stringent criteria were applied to provide a conservative estimate of architectural protein subclasses. Mediator (Med12) binding sites were merged if they fell within 500 bp end-to-end distance of each other. The CTCF+Med12+Smc1 subclass was defined by overlap between high-confidence binding sites (P<1E-8) for CTCF, Med12, and Smc1. The CTCF+Med12, CTCF+Smc1, and Med12+Smc1 subclasses were defined by overlap between high-confidence binding sites (P<1E-8) for two of the proteins and absence of the other protein (via subtraction of all low-confidence binding sites (P<1E-4)). The CTCF Alone, Smc1 Alone, and Med12 Alone subclasses were defined by subtraction of low-confidence binding sites (P<1E-4) for two of the proteins from high-confidence binding sites (P<1E-8) for the primary protein. Heatmaps illustrating

architectural protein subclasses genome-wide were created using the HOMER annotatePeaks.pl algorithm with the following parameters: -size 4000 -hist 25 -ghist (http://biowhat.ucsd.edu/homer/ngs/index.html). Data matrices were visualized using R (http://www.r-project.org/).

Parsing Pioneer Transcription Factor Subclasses

Stringent criteria were applied to provide a conservative estimate of pioneer transcription factor subclasses. The Oct+Nanog+Sox2 subclass was defined by overlap between high-confidence binding sites (P<1E-8) for Oct4, Sox2, and Nanog. The Oct4+Sox2, Nanog+Sox2, and Oct4+Nanog subclasses were defined by overlap between high-confidence binding sites (P<1E-8) for two of the proteins and absence of the other protein (via subtraction of all low-confidence binding sites (P<1E-4)). The Oct4 Alone, Nanog Alone, and Sox2 Alone subclasses were defined by subtraction of low-confidence binding sites (P<1E-4) for two of the proteins from high-confidence binding sites (P<1E-8) for the primary protein. Heatmaps illustrating pioneer factor subclasses genome-wide were created using the HOMER annotatePeaks.pl algorithm (parameters: -size 4000 -hist 25 –ghist) (http://biowhat.ucsd.edu/homer/ngs/index.html). Data matrices were visualized using R (http://www.r-project.org/).

Parsing ES-specific and NPC-specific Enhancers

Stringent criteria were applied to provide a conservative estimate of NPC-specific and ES-specific enhancers in 5C regions. First, H3K4me1 peaks were merged if they fell within 5 kb end-to-end distance of each other and H3K27ac peaks were merged if they fell within 5 kb end-to-end distance of each other. ES-specific enhancers were defined by overlap between high-confidence binding sites (P<1E-8) for H3K4me1 and H3K27ac marks in ES cells and absence of H3K4me1 and H3K27ac marks in NPCs (defined by subtraction of low-confidence NPC binding sites for H3K4me1 and H3K27ac (P<1E-4). NPC-specific enhancers were defined by overlap between high-confidence binding sites (P<1E-8) for H3K4me1 and H3K27ac marks in NPCs and absence of the H3K27ac mark in ES cells (defined by subtraction of low-confidence ES binding sites for H3K27ac (P<1E-4). To ensure subtraction of all potential genes, it was required that

parsed putative ES-specific and NPC-specific enhancers did not fall within 2 kb of a TSSs. Heatmaps illustrating cell type-specific enhancers genome-wide were created using the HOMER annotatePeaks.pl algorithm with parameters -size 16000 -hist 100 –ghist (http://biowhat.ucsd.edu/homer/ngs/index.html). Data matrices were visualized using R (http://www.r-project.org/).

Co-localization of Architectural Proteins with Transcription Factors or Enhancers

Stringent criteria were applied to provide a conservative estimate of the overlap between pioneer transcription factors and architectural protein subclasses. All occupied sites for high-confidence (P<1E-8) Oct, Nanog, or Sox2 were concatenated into a single list. The “OSN with architectural proteins” subclasses were defined by overlap between the OSN occupied sites and the architectural protein subclasses computed as described above. The “OSN without architectural proteins” subclass was defined by absence of CTCF, Smc1, and Med12 (via subtraction of all low-confidence binding sites (P<1E-4)) from the OSN binding sites list.

Stringent criteria were also applied to provide a conservative estimate of the overlap between cell type-specific enhancers and architectural protein subclasses. ES-specific and NPC-specific enhancers were parsed as detailed above. The “ES-specific enhancer with architectural proteins” subclasses were defined by overlap between the ES-specific enhancers and the architectural protein subclasses computed as described above. The “ES-specific enhancers without architectural proteins” subclass was defined by absence of CTCF and Smc1 sites in ES cells (via subtraction of all low-confidence binding sites (P<1E-4)) from the ES-specific enhancer list. The “NPC-specific enhancer with architectural proteins” subclasses were defined by overlap between the NPC-specific enhancers and the architectural protein subclasses computed as described above. The “NPC-specific enhancers without architectural proteins” subclass was defined by absence of CTCF and Smc1 sites in NPCs (via subtraction of all low-confidence binding sites (P<1E-4)) from the NPC-specific enhancer list.

5C Enrichments

Genomic coordinates for each fragment were defined by windowing around a given fragment and the nearest adjacent fragments marked by primers on the opposite strand. For example, the fragment queried by a forward primer was windowed to encompass the two adjacent reverse fragments (R-F-R). Similarly, the fragment queried by a reverse primer was windowed to encompass the two adjacent fragments queried by forward primers (F-R-F). A minimum window size was set at 8 kb for any given F-R-F or R-F-R set of fragments.

Enrichments for a particular ChIP-seq signal in the windowed fragments (F-R-F or R-F-R) at the base of an interaction were calculated in a series of steps. First, intersections were computed between a given protein subclass (defined as described above) and all windowed fragments within 5C regions. Next, a sum of the number of occupied sites in both windowed fragments anchoring the interaction base was computed for the full range of potential interactions queried. Finally, the fraction of ES-specific, NPC-specific, and constitutive interactions containing the protein subclass in at least one of the two windowed fragments anchoring each loop was computed and compared to the fraction of all queried interactions in the "non-looping background". Fisher's Exact Test was used to assess statistical significance of loops with and without co-occupied sites vs. non-looping background with and without co-occupied sites. In the current study, we did not distinguish between proteins binding in one windowed fragment or both windowed fragments anchoring the base of a particular loop.

Interaction Size Distribution

All significant interactions were grouped according to size into 4 bins ranging from 0-2 Mb. Interaction size was calculated as the mid-to-mid distance between the genomic coordinates bounding the two fragments anchoring the base of a specific point-to-point interaction. Two sub-groups were then parsed and compared. The test group consisted of only ES-specific or constitutive looping interactions enriched for a specific protein in the windowed fragments anchoring the loop base. The test distribution was compared to the background distribution of interaction sizes for only ES-specific or constitutive looping interactions depleted of a specific protein in the windowed fragments anchoring the interaction base. Data is presented as an enrichment in each

bin (computed as a ratio of the fraction of interactions containing a specific protein vs. the fraction of interactions depleted of the specific protein). Fisher's Exact Test was used to assess significance of enrichment in each length scale bin.

Cluster Analysis

Unsupervised *k*-means clustering was performed with R using Euclidean distance as the distance measure. Heatmaps were visualized using R (http://www.r-project.org/).

Gene Ontology

Gene ontology analysis was conducted with DAVID (http://david.abcc.ncifcrf.gov/) (Huang da et al., 2009). UCSC annotations for genes only within 5C regions of interest were used as background. The subclass of parsed genes co-localized with ES-specific Smc1 occupied sites and also intersecting the fragments at the base of ES-specific chromatin interactions were then evaluated against this background.

Gene Expression Analysis

Occupied sites for a particular protein were first parsed into a subclass of only those sites found anchoring the base of significant looping interactions. These sites were then further parsed into occupied sites anchoring the base of significant looping interactions that also co-localized with genes in the 5C regions. Two controls were parsed as follows: (1) all genes within fragments anchoring significant chromatin interactions and (2) all genes found in non-looping background interactions co-localized with specific protein co-occupied sites. Expression was then evaluated as the log2 ratio of expression in V6.5 ES cells vs. NPCs or the log2 ratio of expression in siRNA-treatment vs. wild type ES cells. Microarray data were downloaded from GEO: (1) Series GSE22557 (GSM559811 Med12_KD_Day5_Rep1; GSM559812 Med12_KD_Day5_Rep2; GSM559813 Smc1_KD_Day5_Rep1; GSM559814 Smc1_KD_Day5_Rep2) and (2) Series GSE24165 (GSM589518 V6.5_ES_cell_Rep1; GSM589519 V6.5_ES_cell_Rep2; GSM589520 V6.5_Neural_Progenitor_Cell_Rep1;

GSM589520 V6.5_Neural_Progenitor_Cell_Rep2) (Creyghton et al., 2010; Kagey et al., 2010). Significance between distributions was assessed using the Kolmogorov-Smirnov test.

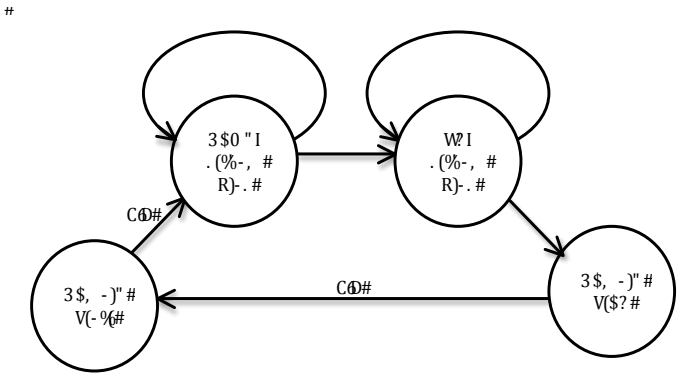
Hidden Markov Model for Domain Calling

In order to identify domains, we used a modified version of the previously published directionality index (DI) (Dixon et al., 2012). For each continuous block of a specified number of fragments, sets of interactions upstream (U) and downstream (D) of the primer block were found for several different DI tracks corresponding to different length scales. The distance up and downstream used for calculating the DI ranged from 100 – 650 kb and was divided into bins. The number of bins was determined for each region with the first bin containing interactions within the first 100 kb of the fragment block and each subsequent bin containing interactions from equally-sized non-overlapping bins spanning from 100-650 kb. Parameters used for each region include: Sox2 (fragments = 20 and bins = 2), Olig1-Olig2 (fragments = 6 and bins = 3), Nanog (fragments = 12 and bins = 2), Nestin (fragments = 20 and bins = 2), Klf4 (fragments = 6 and bins = 3), Oct4 (fragments = 15 and bins = 3).

The modified DI was calculated as a t-statistic within unequal sample size and variance.

$$s_{\bar{X}_1 - \bar{X}_2} = \left(\frac{\sum (U_i - E(U))^2}{|U|^2 - |U|} + \frac{\sum (D_j - E(D))^2}{|D|^2 - |D|} \right)^{0.5}$$

$$DI = \frac{E(U) - E(D)}{s_{\bar{X}_1 - \bar{X}_2}}$$



Domain calls were made using a 4-state mixture model HMM. Two states represented domain starts and stops while the remaining two states were downstream biased and upstream biased. Each state emitted from a three-distribution Gaussian mixture. The model was trained on data from all DI tracks using the Baum-Welch algorithm. In order to incorporate information across DI tracks and assess the hierarchical nature of interactions, domains were called in a greedy fashion from the largest scale DI track to the smallest. The best-scoring set of domain calls was determined at each level by summing path scores for each scale of DI track as determined by the Viterbi algorithm. However, any domain boundary found in a larger scale DI track was passed to smaller scale DI tracks as a required state for that position. Each boundary found was optimized across smaller scale tracks within three positions on either side of the original call prior to being passed down. For example, if a domain boundary was found at positions i to i+1, all smaller scale tracks would be required to include those hidden states at positions i and i+1 as well.

Author Contributions

JEPC and VGC conceived the project. JEPC and TM designed stem cell differentiation experiments. JEPC, AS, BL, and JD designed 5C experiments. JEPC, SD, and YS designed lentiviral experiments. JEPC, TIG, CG, VGC, JD, and RS designed FISH experiments. JEPC, AS, WDW, JKB, MJB, CO, TAH, TIG, CG, and YS performed experiments. MEGS, JEPC, and JPT developed concepts and code for 5C peak calling pipeline and HMM model. JEPC performed post-peak calling analysis with input from

JD, MEGS, and JPT. JEPC analyzed data and wrote the paper with input from JD, JPT, and VGC.

Supplementary References

Aird, D., Ross, M.G., Chen, W.S., Danielsson, M., Fennell, T., Russ, C., Jaffe, D.B., Nusbaum, C., and Gnirke, A. (2011). Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. Genome Biol *12*, R18.

Bau, D., Sanyal, A., Lajoie, B.R., Capriotti, E., Byron, M., Lawrence, J.B., Dekker, J., and Marti-Renom, M.A. (2011). The three-dimensional folding of the alpha-globin gene domain reveals formation of chromatin globules. Nat Struct Mol Biol *18*, 107-114.

Carpenedo, R.L., Sargent, C.Y., and McDevitt, T.C. (2007). Rotary suspension culture enhances the efficiency, yield, and homogeneity of embryoid body differentiation. Stem Cells *25*, 2224-2234.

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. Proc Natl Acad Sci U S A *107*, 21931-21936.

Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. (2002). Capturing chromosome conformation. Science *295*, 1306-1311.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature *485*, 376-380.

Dostie, J., and Dekker, J. (2007). Mapping networks of physical interactions between genomic elements using 5C technology. Nat Protoc *2*, 988-1002.

Dostie, J., Richmond, T.A., Arnaout, R.A., Selzer, R.R., Lee, W.L., Honan, T.A., Rubio, E.D., Krumm, A., Lamb, J., Nusbaum, C.*, et al.* (2006). Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. Genome Res *16*, 1299-1309.

Gheldof, N., Smith, E.M., Tabuchi, T.M., Koch, C.M., Dunham, I., Stamatoyannopoulos, J.A., and Dekker, J. (2010). Cell-type-specific long-range looping interactions identify distant regulatory elements of the CFTR gene. Nucleic Acids Res *38*, 4325-4336.

Guo, C., Gerasimova, T., Hao, H., Ivanova, I., Chakraborty, T., Selimyan, R., Oltz, E.M., and Sen, R. (2011). Two forms of loops generate the chromatin conformation of the immunoglobulin heavy-chain gene locus. Cell *147*, 332-343.

Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc *4*, 44-57.

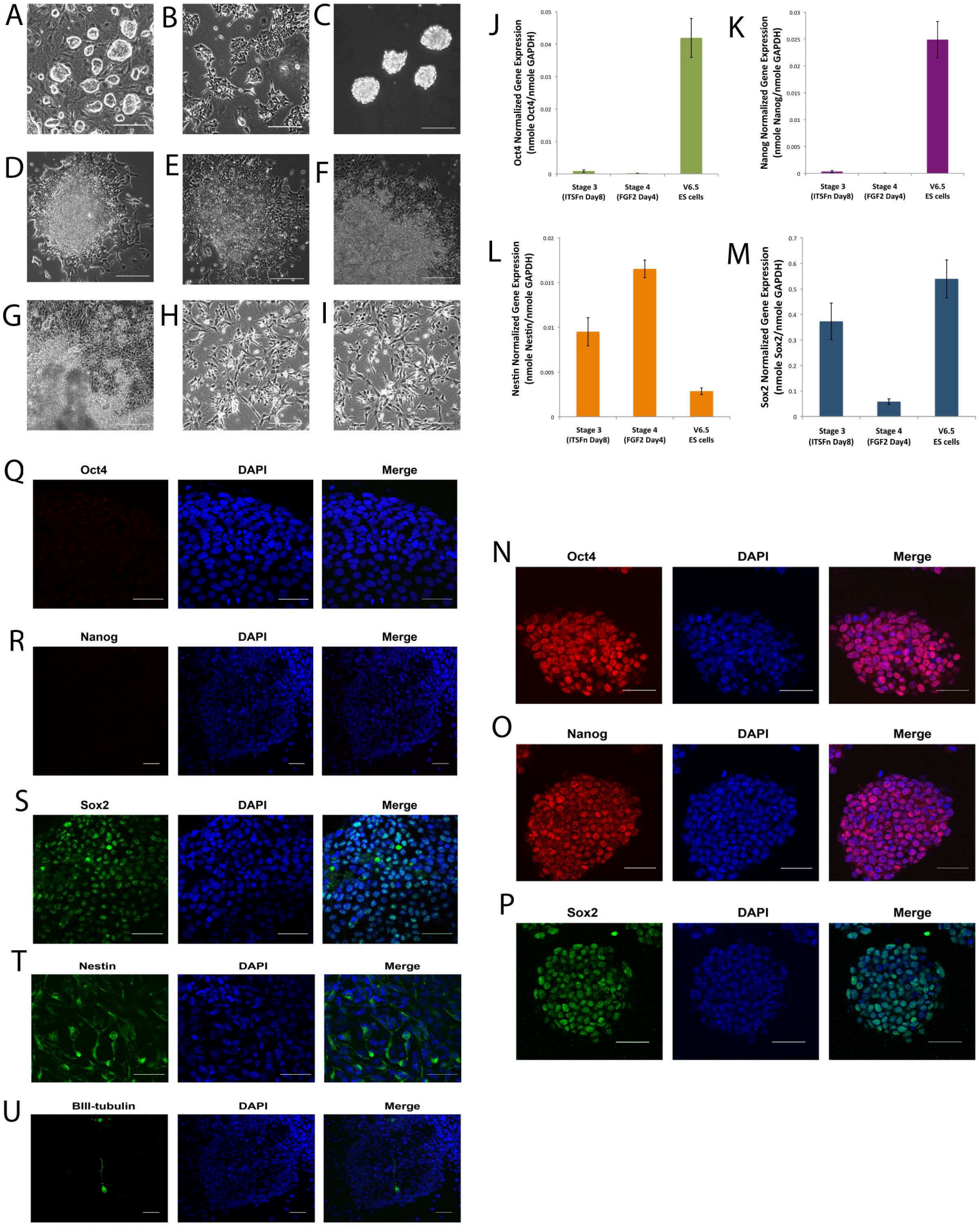
Imakaev, M., Fudenberg, G., McCord, R.P., Naumova, N., Goloborodko, A., Lajoie, B.R., Dekker, J., and Mirny, L.A. (2012). Iterative correction of Hi-C data reveals hallmarks of chromosome organization. Nat Methods *9*, 999-1003.

Kagey, M.H., Newman, J.J., Bilodeau, S., Zhan, Y., Orlando, D.A., van Berkum, N.L., Ebmeier, C.C., Goossens, J., Rahl, P.B., Levine, S.S.*, et al.* (2010). Mediator and cohesin connect gene expression and chromatin architecture. Nature *467*, 430-435.

Lajoie, B.R., van Berkum, N.L., Sanyal, A., and Dekker, J. (2009). My5C: web tools for chromosome conformation capture studies. Nat Methods *6*, 690-691.

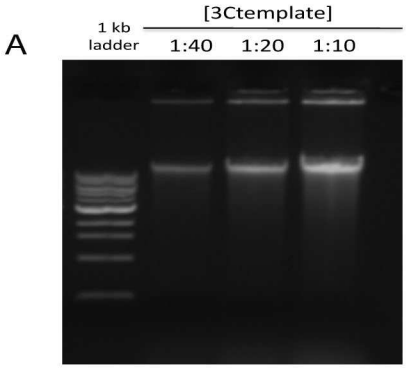
Langmead, B. (2010). Aligning short sequencing reads with Bowtie. Curr Protoc Bioinformatics *Chapter 11*, Unit 11 17.
Meissner, A., Mikkelsen, T.S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B.E., Nusbaum, C., Jaffe, D.B.*, et al.* (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature *454*, 766-770.
Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.K., Koche, R.P.*, et al.* (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature *448*, 553-560.
Okabe, S., Forsberg-Nilsson, K., Spiro, A.C., Segal, M., and McKay, R.D. (1996). Development of neuronal precursor cells and functional postmitotic neurons from embryonic stem cells in vitro. Mech Dev *59*, 89-102.
Phillips, J.E., Burns, K.L., Le Doux, J.M., Guldberg, R.E., and Garcia, A.J. (2008). Engineering graded tissue interfaces. Proc Natl Acad Sci U S A *105*, 12170-12175.
Phillips, J.E., Gersbach, C.A., Wojtowicz, A.M., and Garcia, A.J. (2006). Glucocorticoid-induced osteogenesis is negatively regulated by Runx2/Cbfa1 serine phosphorylation. J Cell Sci *119*, 581-591.
Rousseau, M., Fraser, J., Ferraiuolo, M.A., Dostie, J., and Blanchette, M. (2011). Three-dimensional modeling of chromatin structure from interaction frequency data using Markov chain Monte Carlo sampling. BMC Bioinformatics *12*, 414.
Sanyal, A., Lajoie, B.R., Jain, G., and Dekker, J. (2012). The long-range interaction landscape of gene promoters. Nature *489*, 109-113.
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.
van Berkum, N.L., and Dekker, J. (2009). Determining spatial chromatin organization of large genomic regions using 5C technology. Methods Mol Biol *567*, 189-213.
Wood, A.M., Van Bortle, K., Ramos, E., Takenaka, N., Rohrbaugh, M., Jones, B.C., Jones, K.C., and Corces, V.G. (2011). Regulation of chromatin organization and inducible gene expression by a Drosophila insulator. Mol Cell *44*, 29-38.
Yaffe, E., and Tanay, A. (2011). Probabilistic modeling of Hi-C contact maps eliminates systematic biases to characterize global chromosomal architecture. Nat Genet *43*, 1059-1065.
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 Biol *9*, R137.

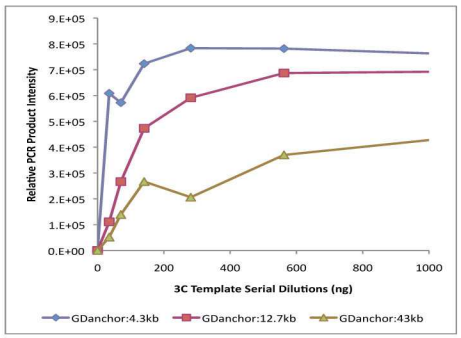
Supplemental Figure 1

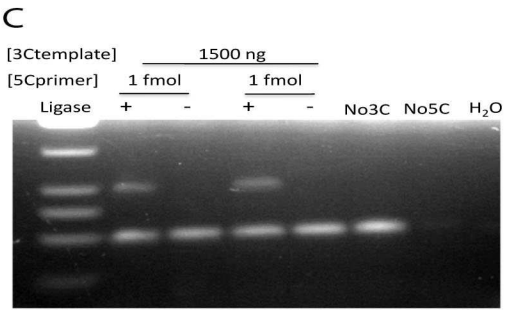


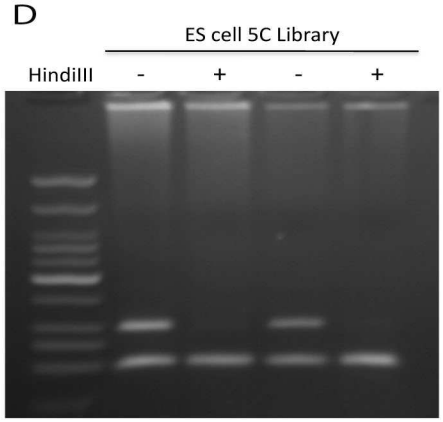
Supplementary _Figure_1__Phillips-Cremins_2013.

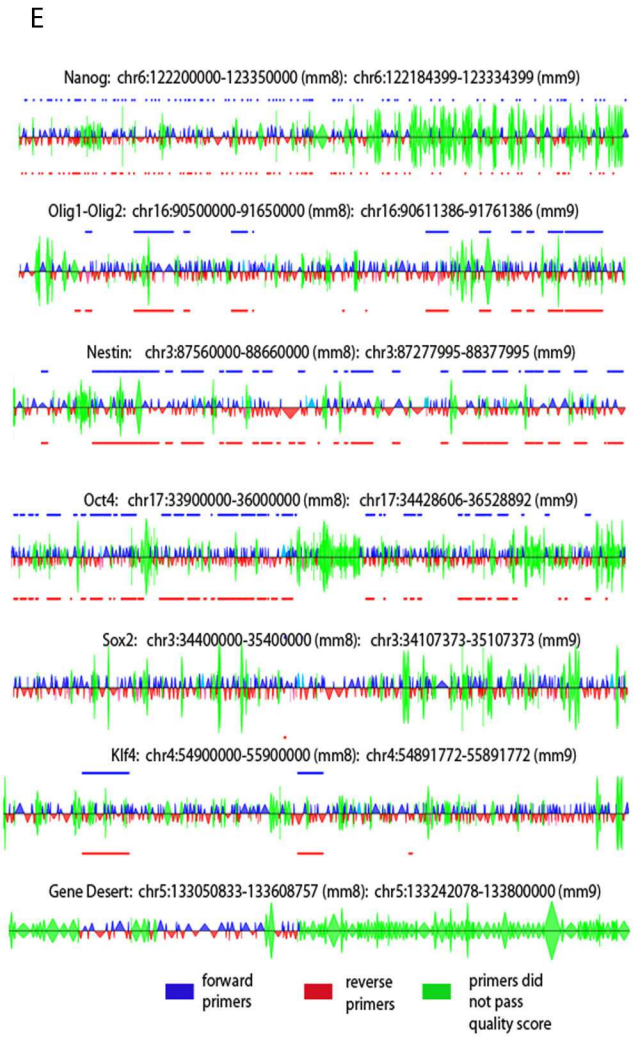
Supplemental Figure 2

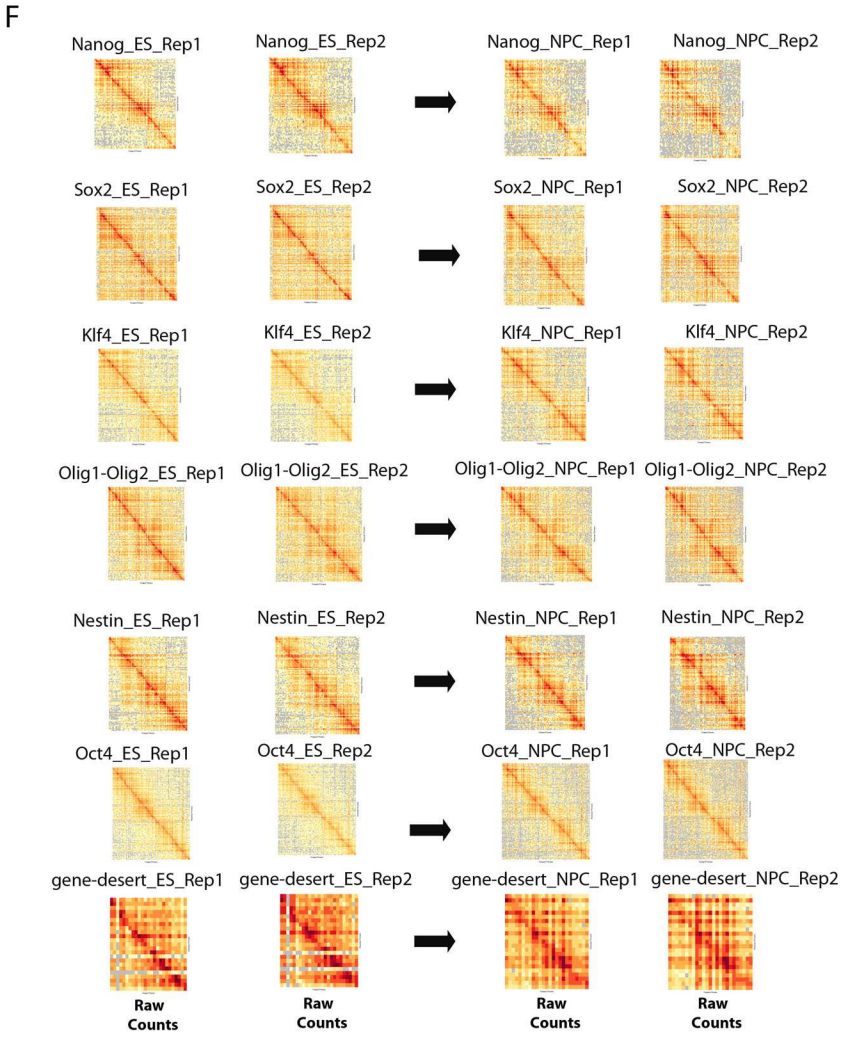






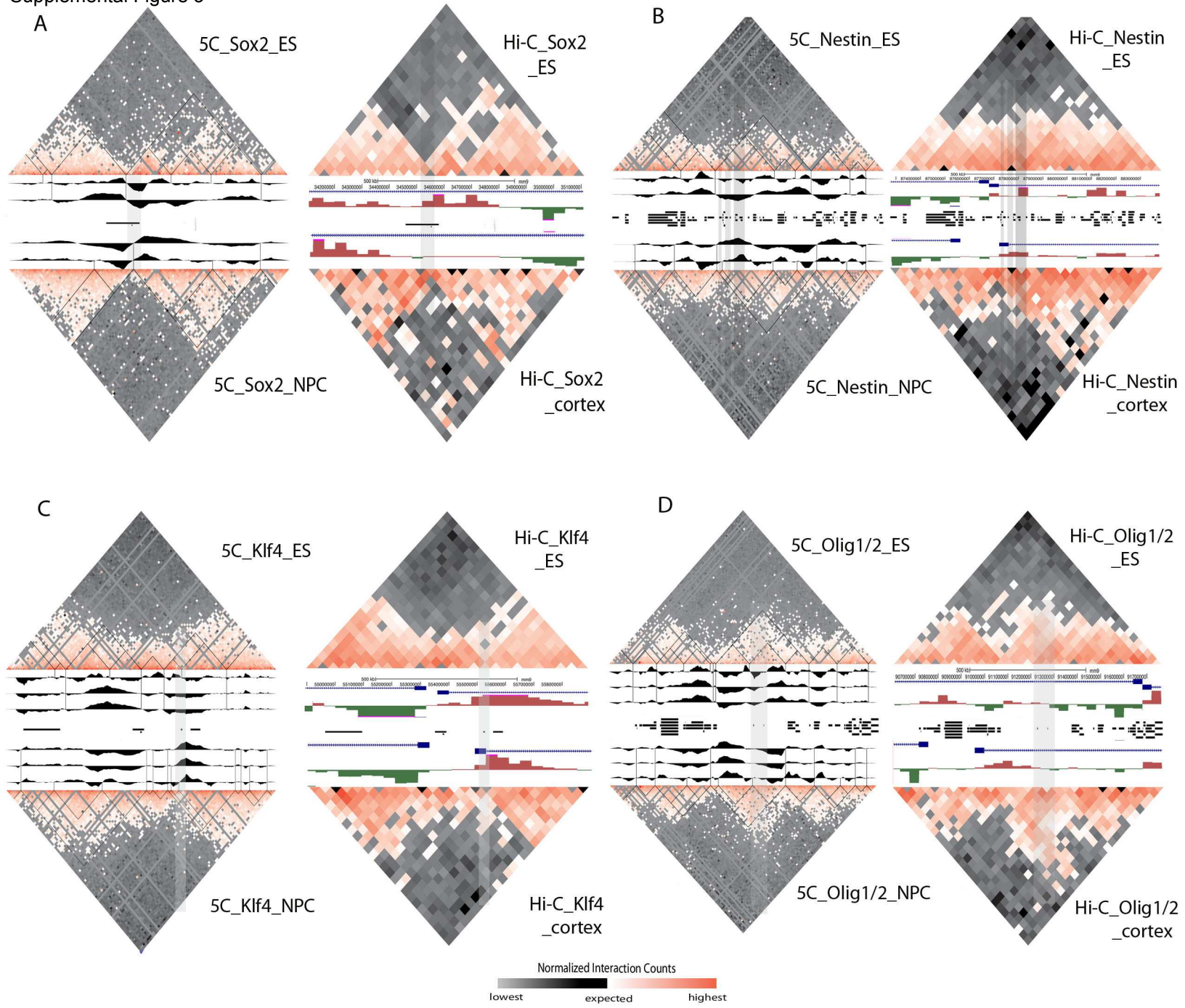






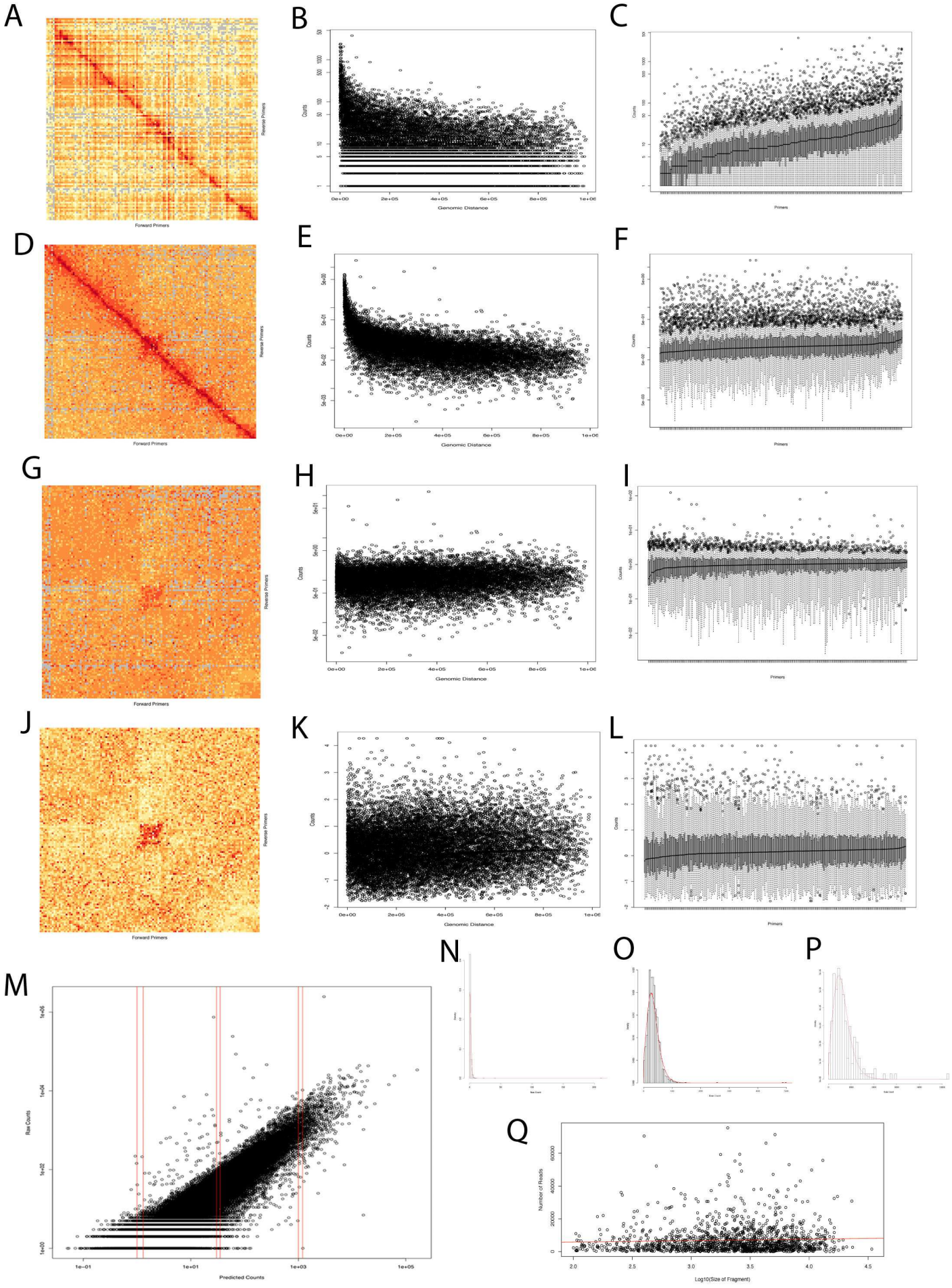
Supplementary_Figure_2_Phillips-Cremins_2013.

Supplemental Figure 3



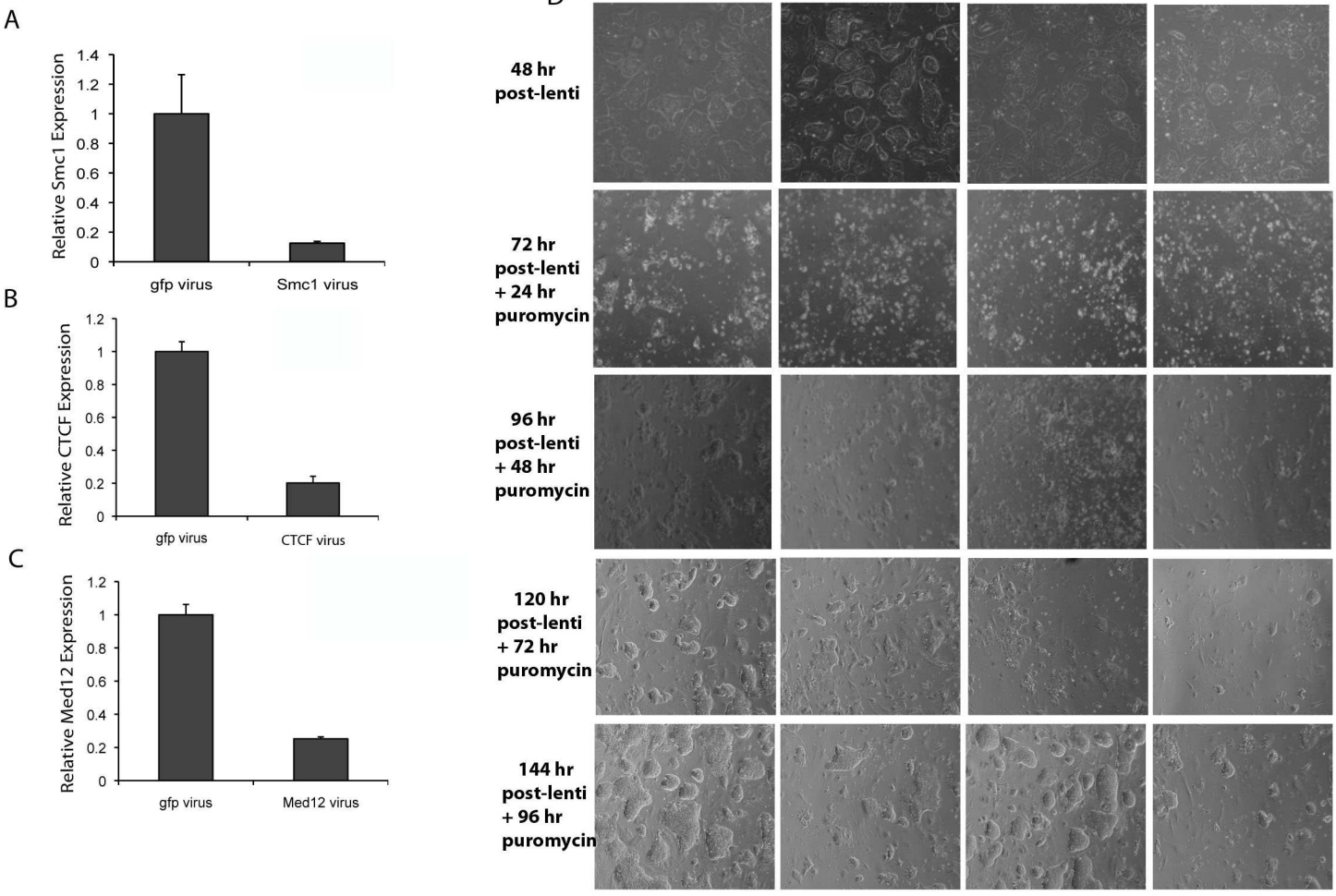
Supplementary_Figure_3_Phillips-Cremins_2013.

Supplemental Figure 4



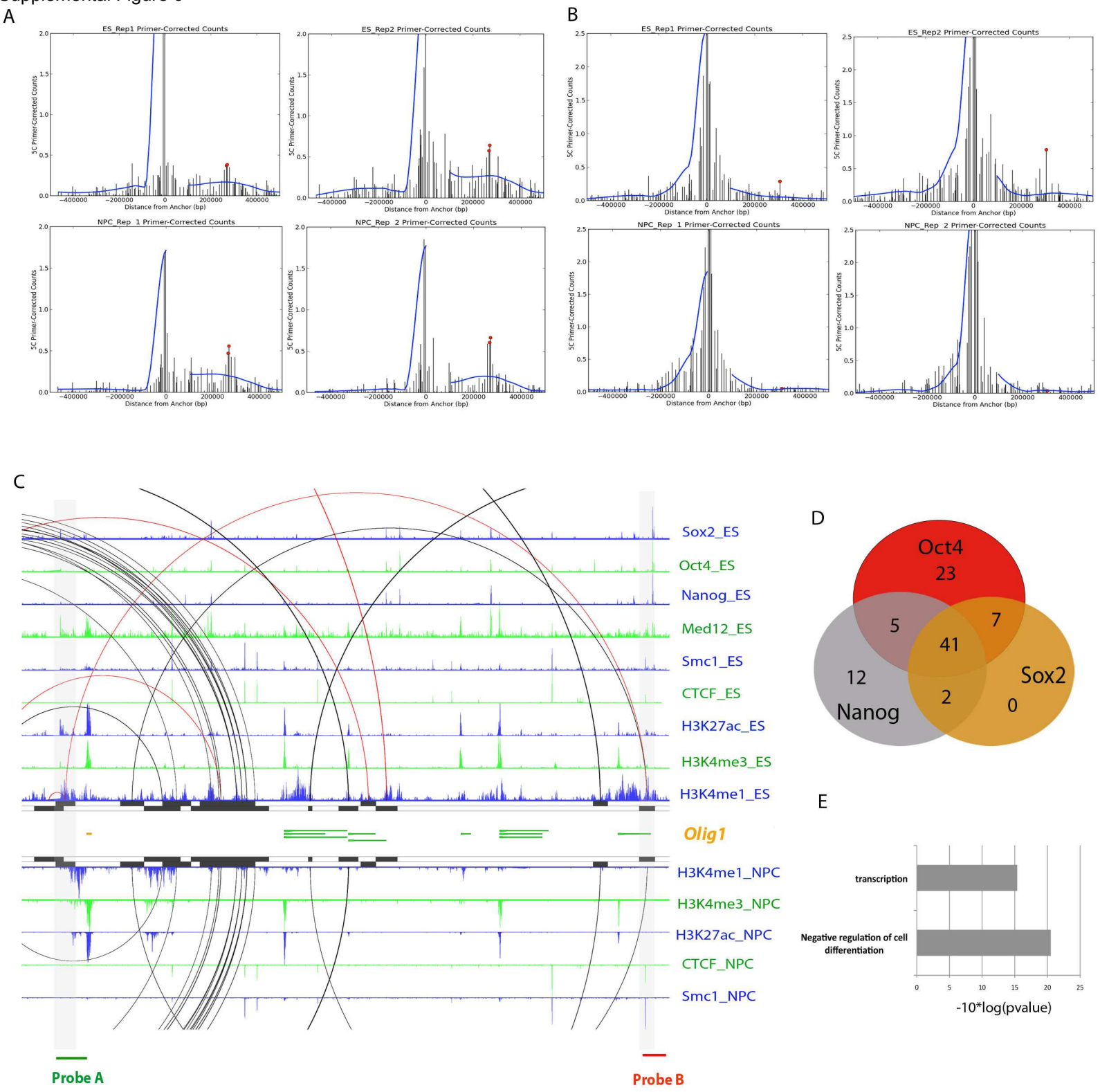
Supplementary_Figure_4_Phillips-Cremins_2013.

Supplemental Figure 5



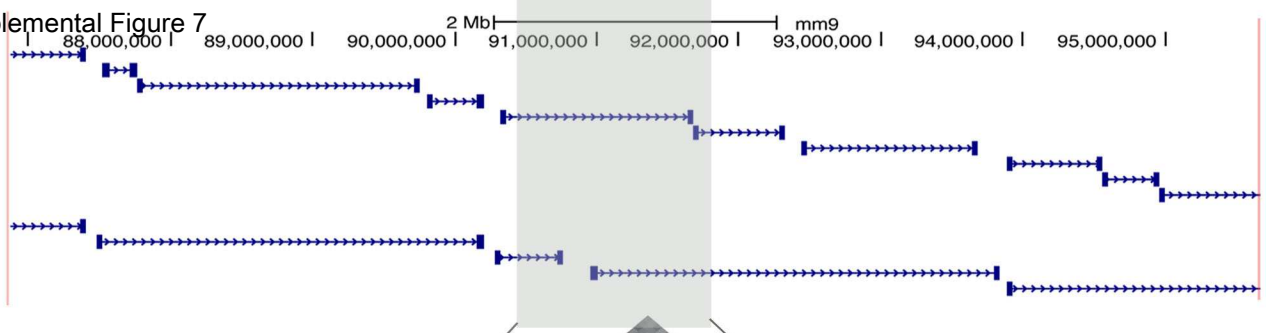
Supplementary_Figure_5_Phillips-Cremins_2013.

Supplemental Figure 6



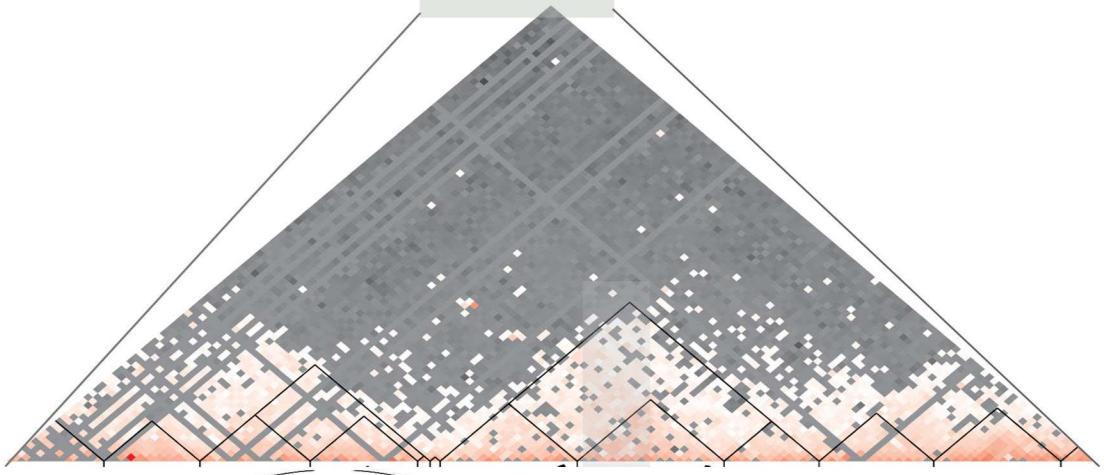
Supplementary _Figure_6_Phillips-Cremins_2013.

Supplemental Figure 7

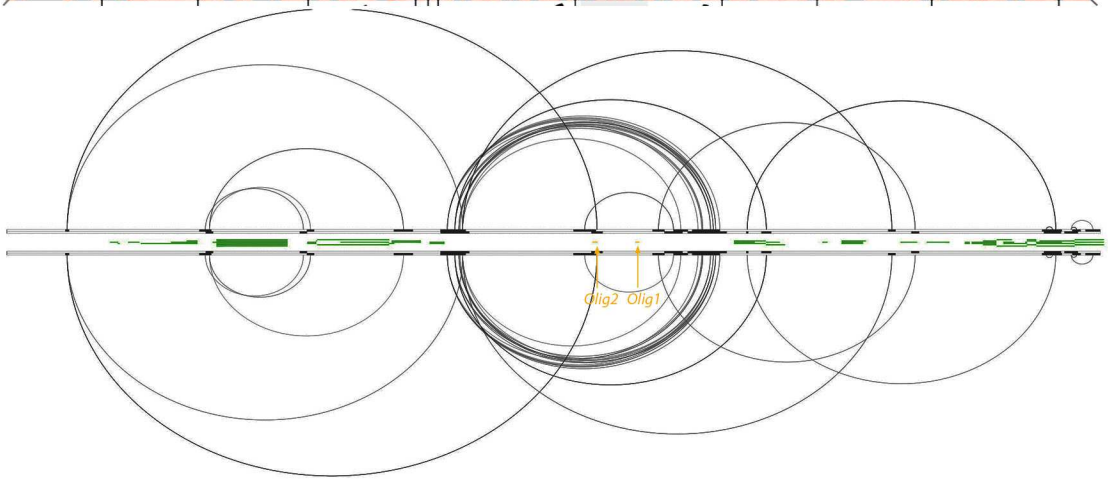


Topological Domains
Hi-C_ES

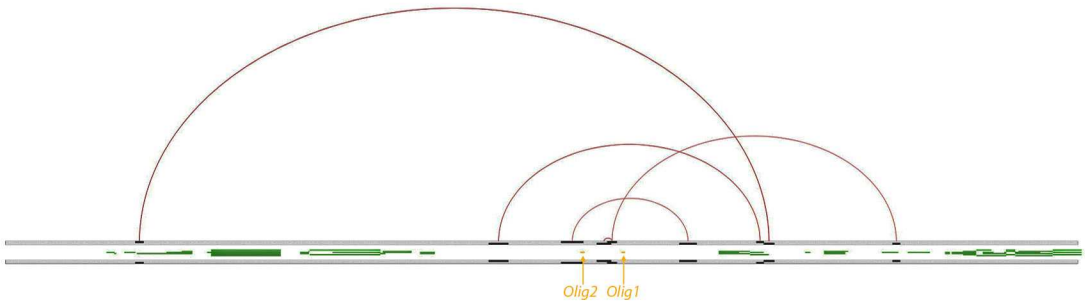
Topological Domains
Hi-C_cortex



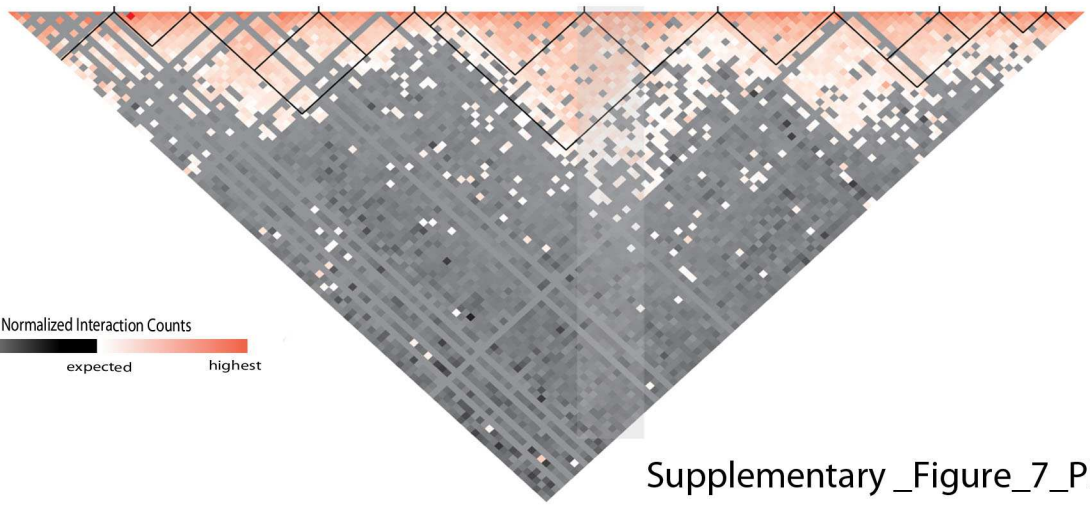
5C_ES_cells
Olig1-Olig2_region



Constitutive
3-D interactions
connected by
constitutive
CTCF+Smc1



ES-specific
3-D interactions
connected by
ES-specific Smc1



5C_NPCs_
Olig1-Olig2_region

Supplementary _Figure_7_Phillips-Cremins_2013.