Molecular Cell, ***Volume 46***

Supplemental Information

Embryonic Stem Cells Induce Pluripotency in Somatic Cell Fusion through Biphasic Reprogramming

Kara M. Foshay, Timothy J. Looney, Sheila Chari, Frank Fuxiang Mao, Jae Hyun Lee, Li Zhang, Croydon J. Fernandes, Samuel W. Baker, Kayla L. Clift, Jedidiah Gaetz, Chun-Guang Di, Andy Peng Xiang, and Bruce T. Lahn

Supplemental Discussion

Among the 4 chromatin modifying genes that became activated in R1A by fusion with E14, *Satb1* and *Chd7* are involved in modulating chromatin structure while *Dnmt3b* and *Tet1* are involved in regulating DNA methylation. *Satb1* encodes a regulator of higher-order chromatin architecture (Yasui et al., 2002), which has been shown to bind chromatin scaffolds and recruits chromatin modifying complexes such as NuRD (Cai et al., 2003). Downstream effects of Satb1 binding include changes in chromatin looping and regulation of gene expression. Interestingly, knockdown of this gene in ESCs impairs differentiation and upregulates the pluripotency genes *Nanog*, *Klf4*, and *Tbx3* (Savarese et al., 2009). *Chd7* encodes a chromodomain helicase that is commonly found at enhancers and is associated with chromatin containing high levels of H3K4 mono-methylation. Interestingly, the Chd7 protein co-localizes with Oct4, Nanog and Sox2 proteins in the genome, and modulates ESC gene expression (Schnetz et al., 2010). *Dnmt3b* encodes a de novo methyltransferase that is important for establishing methylation patterns in early embryos and ESCs (Okano et al., 1999), and plays a role in proper differentiation (Chen et al., 2003), but is not essential for generating induced pluripotent stem cells (iPSCs) (Pawlak and Jaenisch, 2011). *Tet1* encodes a methylcytosine dioxygenase, which converts 5-methylcytosine (5mC) in DNA to 5-hydroxymethylcytosine (5hmC) (Tahiliani et al., 2009). 5hmC can then serve as an intermediate in a pathway leading to active demethylation (Guo et al., 2011; He et al., 2011; Ito et al., 2011b). Interestingly, this gene is regulated directly by Oct4 and Sox2, is upregulated during iPSC induction, controls *Nanog* expression in ESCs, and plays a role in proper ESC differentiation (Ito et al., 2011a; Koh et al., 2011). Despite this, *Tet1* knockout does not result in an overt phenotype in ESCs or animals, possibly due to redundant gene function among related family members (Dawlaty et al., 2011). 5hmC is found at high levels in ESCs as compared to other tissues, and was speculated to contribute to a distinct chromatin state that maintains the balance between lineage specification and pluripotency (Wu and Zhang, 2011). Together, our data and previous studies on these chromatin modifiers implicate them as potential players in chromatin remodeling and the resetting of DNA methylation during the *cis*-reprogramming process.

Supplemental Experimental Procedures

Cell Culture and Fusion

R1A was derived from the Rat-1a rat embryonic fibroblasts (Stone et al., 1987). L6 was derived from the L6 rat myoblasts (ATCC, cat# CRL-1458). Both were derived by transduction of parental cells with an hEF1a-dTomato lentiviral vector containing hygromycin resistance as described previously (Qin et al., 2010), followed by the isolation of a clonal line. C2C12 was derived from the C2C12 mouse myoblasts (ATCC, cat #: CRL-1772), via transduction of parental cells with an hEF1a-EGFP lentiviral vector containing puromycin resistance as described previously (Qin et al., 2010), followed by the isolation of a clonal line. 129TF mouse tail fibroblasts were derived from a 3-week old female of 129 strain background according to published protocol (Xu, 2005), transduced with EGFP-expressing lentivirus, followed by isolation of a clonal line. E14 mouse embryonic stem cells were derived clonally from ES-E14TG2A parental cells (ATCC; catalog #: CRL-1821) transduced with EGFP-expressing lentivirus. The N2A mouse neuroblastoma cells were derived clonally from Neuro-2a parental cells (ATCC; catalog #: CCL-131) transduced with EGFP-expressing lentivirus. E14 and E14Aid cells were cultured under feeder-free conditions in MEM with 10% FBS, non-essential amino acids, sodium pyruvate, LIF, 50 μM β-mercaptoethanol, 3 μM CHIR99021 (Stemgent), and 1 μM PD0325901 (Stemgent). B6NSC mouse neural stem cells were derived and maintained as described previously with minor modification (Ma et al., 2006). C57BL/6 mouse neural tube was dissected at E12.5, mechanically dissociated, and cultured at a density of 5×10^4 cells/ml in DMEM/F-12 medium (Invitrogen) supplemented with B27 (Invitrogen), 20 ng/ml basic fibroblast growth factor (Peprotech), 20 ng/ml epidermal growth factor (Peprotech) and 5 mg/ml heparin (Sigma). For differentiation of B6NSC, cells were plated onto poly-lysine-coated cover-slips in DMED/F-12 medium (Invitrogen) supplemented with B27 (Invitrogen) and 1 mM retinoic acid (Sigma). Cells were fixed with 4% paraformaldehyde and stained before or after differentiation with antibodies against nestin (Millpore; 1:150 dilution) for neural stem cells, Tuj-1 (R&D; 1:200 dilution) for neurons, GFAP (Dako; 1:500 dilution) antibody for astrocytes, and O4 (Abcam;1:200 dilution) for oligodendrocyte. For somatic-ESC fusions, cells were fused in suspension at a 1:1 ratio, with 1 ml PEG 1500MW (50% w/v in serum-free DMEM) added for 1 minute at 37°C, followed by the addition of 3 ml serum-free DMEM over the next two minutes. Cells were centrifuged at 500 RPM for 5 minutes then washed and resuspended at $15x10^6$ cells/ml for FACS sorting. EGFP and dTomato double fluorescent cells were purified using a BD FACS Aria II. For somatic-

somatic fusions, cells were plated together at a 1:1 ratio for at least 2 hours. 5 ml 50% PEG 1500MW was added for 1 minute. Cells were washed and allowed to recover in complete culture media, followed by dissociation and replating at low density in selection media. Daily passaging of fused cells yielded pure fused cultures within 4-6 days.

Generation of a Mouse-Rat Ortholog Library

Ensembl.org was used to obtain fasta libraries of all annotated mouse and rat ORFs, which were converted to protein sequence using blast2protein. BLAST was used to align each mouse protein sequence to the library of rat protein sequences. The rat protein sequence with the highest homology, and therefore the highest BLAST score, was aligned to the mouse protein sequence library following the same procedure. If the highest ranked mouse protein sequence from this reciprocal query was identical to the original mouse protein sequence, then the pair of mouse and rat protein sequences were considered to be orthologous. This method identified 47663 pairs of orthologous mouse and rat ORFs representing 18036 genes. Multiple pairs of orthologous mouse-rat ORFs represented single genes that possess multiple annotated transcripts. In such cases, we used the ORF pair whose mouse member most closely matched a human-verified RefSEQ transcript. If no RefSEQ match was available, we used the ORF pair whose mouse member had been annotated earliest in the ensembl database. Our resulting library contained 18036 genes, each represented by a single pair of orthologous mouse-rat ORFs. Within our library, ORFs of differing size, or mis-annotation of ORF boundaries would produce non-equivalent ORF pairs and would confound the quantification of ortholog expression within the fused cells. To address this problem, we used a 12 base sliding window to assess local homology in each ORF pair. Windows containing 6 or more mismatched bases were redacted with Ns in the ORF library. Thus, the ORFs in each pair are nearly identical in size. We further removed any genes with <80% overall homology between orthologs. We also removed genes where, among the RNA-Seq reads generated form unfused cells of one species that map to either mouse or rat ortholog, >2% map to the ortholog of the wrong species. After applying these filters, our library contained 16344 pairs of mouse-rat orthologs. One potential problem with the identification of occluded genes is wrong orthology assignment in our mouse-rat ortholog library. This could occur if a mouse gene is paired with a rat pseudogene (or closely related gene family member) instead of the true rat ortholog, which would cause the gene to appear occluded in the rat. Sequencing and annotation of the rat genome are less detailed than that of the mouse, which could lead to some mis-paired orthologs. However, given the high homology cutoff we used in orthology assignment, mis-paired orthologs should be rather rare. A second potential problem in the identification of occlude genes is the possibility of regulatory incompatibility of evolutionarily divergent promoters and enhancers in the mouse and rat genomes. However, our previous analysis showed that interspecies incompatibility is not a significant confounding factor in the identification of occluded genes (Lee et al., 2009). Overall, despite some technical limitations, we believe that the great majority of occluded genes are properly identified.

Mapping of RNA-Seq Reads

To eliminate ambiguity we selected reads that matched perfectly to only one single location within the combined ORF library of both mouse and rat. On average, selecting for perfectly, uniquely matching reads reduced the total number of reads per ORF by 46% for both mouse and rat. Calculation of the correlation between the proportion of mouse and rat reads lost for each gene confirmed that read reduction was highly similar for each ORF member, and indicated that asymmetric read reduction would not confound our data analysis. In 129TF, 0.38% of reads aligned perfectly and uniquely to the rat genome, while in B35 rat neuroblastoma cells, 0.46% of reads mapped perfectly and uniquely to the mouse genome. These low percentages indicated that our mapping criteria were highly effective.

RT-PCR and RT-PCR-Seq

RNA was isolated using standard protocols and treated with DNase I. cDNA conversion was conducted using the Superscript III First-Strand Synthesis System with random hexamers according to manufacturer's protocol (Invitrogen). PCR was done using 4 ng cDNA per 20 µl reaction and 35 cycles. Standard sequencing reactions were set up using 5 µl of PCR product, treated with shrimp alkaline phosphatase and exonuclease I (USB), and sequenced using the ABI Big Dye sequencing protocol (Applied Biosystems). Sequencing reactions were purified using ethanol precipitation, run on an ABI 3730xl DNA Analyzer (Applied Biosystems), and analyzed using Sequencher (GeneCodes Corporation).

Aid Expression Studies

E14Aid cells were derived clonally from E14, by transduction with a lentiviral vector carrying hEF1a-m*Aid* and blasticidin resistance. Primer sequences used for RT-PCR are as follows: ORF forward – TCAGCCTGAGGATTTTCACC; ORF reverse – TACAAGGGCAAAAGGATGCG; ORFtoUTR forward – CGCATCCTTTTGCCCTTGTA; ORFtoUTR reverse – CAGTAGATGGCGATGTTG; ORFtoTransgene forward - CGCATCCTTTTGCCCTTGTA; ORFtoTransgene reverse – ACGGGCCACAACTCCTCATA.

DNA Methylation Analysis

Mouse-specific primers used for amplification of bisulfite-treated DNA were as published (Imamura et al., 2006), and rat-specific primers were as follows: rat CR1: GATGGGGATTTAAGTAATTGGT and CCTCTAACCTTAACCTCTAACCC; rat CR4: GTTGAGTTTGGATAGGAAGGTT and AACTTCCCCACAATAAACAAAT.

Supplemental References

Cai, S., Han, H.J., and Kohwi-Shigematsu, T. (2003). Tissue-specific nuclear architecture and gene expression regulated by SATB1. Nature genetics *34*, 42-51.

Chen, T., Ueda, Y., Dodge, J.E., Wang, Z., and Li, E. (2003). Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b. Mol Cell Biol *23*, 5594-5605.

Dawlaty, M.M., Ganz, K., Powell, B.E., Hu, Y.C., Markoulaki, S., Cheng, A.W., Gao, Q., Kim, J., Choi, S.W., Page, D.C., and Jaenisch, R. (2011). Tet1 is dispensable for maintaining pluripotency and its loss is compatible with embryonic and postnatal development. Cell stem cell *9*, 166-175.

Guo, J.U., Su, Y., Zhong, C., Ming, G.L., and Song, H. (2011). Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. Cell *145*, 423-434.

He, Y.F., Li, B.Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L.*, et al.* (2011). Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. Science (New York, N.Y *333*, 1303-1307.

Imamura, M., Miura, K., Iwabuchi, K., Ichisaka, T., Nakagawa, M., Lee, J., Kanatsu-Shinohara, M., Shinohara, T., and Yamanaka, S. (2006). Transcriptional repression and DNA hypermethylation of a small set of ES cell marker genes in male germline stem cells. BMC Dev Biol *6*, 34.

Ito, S., D'Alessio, A.C., Taranova, O.V., Hong, K., Sowers, L.C., and Zhang, Y. (2011a). Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. Nature *466*, 1129-1133.

Ito, S., Shen, L., Dai, Q., Wu, S.C., Collins, L.B., Swenberg, J.A., He, C., and Zhang, Y. (2011b). Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. Science (New York, N.Y *333*, 1300-1303.

Koh, K.P., Yabuuchi, A., Rao, S., Huang, Y., Cunniff, K., Nardone, J., Laiho, A., Tahiliani, M., Sommer, C.A., Mostoslavsky, G.*, et al.* (2011). Tet1 and Tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells. Cell stem cell *8*, 200-213.

Lee, J.H., Bugarija, B., Millan, E.J., Walton, N.M., Gaetz, J., Fernandes, C.J., Yu, W.H., Mekel-Bobrov, N., Vallender, T.W., Snyder, G.E.*, et al.* (2009). Systematic identification of cis-silenced genes by trans complementation. Human molecular genetics *18*, 835-846.

Ma, B.F., Liu, X.M., Xie, X.M., Zhang, A.X., Zhang, J.Q., Yu, W.H., Zhang, X.M., Li, S.N., Lahn, B.T., and Xiang, A.P. (2006). Slower cycling of nestin-positive cells in neurosphere culture. Neuroreport *17*, 377-381.

Okano, M., Bell, D.W., Haber, D.A., and Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell *99*, 247-257.

Pawlak, M., and Jaenisch, R. (2011). De novo DNA methylation by Dnmt3a and Dnmt3b is dispensable for nuclear reprogramming of somatic cells to a pluripotent state. Genes & development *25*, 1035-1040.

Qin, J.Y., Zhang, L., Clift, K.L., Hulur, I., Xiang, A.P., Ren, B.Z., and Lahn, B.T. (2010). Systematic comparison of constitutive promoters and the doxycycline-inducible promoter. PLoS One *5*, e10611.

Savarese, F., Davila, A., Nechanitzky, R., De La Rosa-Velazquez, I., Pereira, C.F., Engelke, R., Takahashi, K., Jenuwein, T., Kohwi-Shigematsu, T., Fisher, A.G., and Grosschedl, R. (2009). Satb1 and Satb2 regulate embryonic stem cell differentiation and Nanog expression. Genes & development *23*, 2625-2638.

Schnetz, M.P., Handoko, L., Akhtar-Zaidi, B., Bartels, C.F., Pereira, C.F., Fisher, A.G., Adams, D.J., Flicek, P., Crawford, G.E., Laframboise, T.*, et al.* (2010). CHD7 targets active gene enhancer elements to modulate ES cell-specific gene expression. PLoS genetics *6*, e1001023.

Stone, J., de Lange, T., Ramsay, G., Jakobovits, E., Bishop, J.M., Varmus, H., and Lee, W. (1987). Definition of regions in human c-myc that are involved in transformation and nuclear localization. Mol Cell Biol *7*, 1697-1709.

Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L., and Rao, A. (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science (New York, N.Y *324*, 930-935.

Wu, H., and Zhang, Y. (2011). Tet1 and 5-hydroxymethylation: a genome-wide view in mouse embryonic stem cells. Cell Cycle *10*, 2428-2436.

Xu, J. (2005). Preparation, culture, and immortalization of mouse embryonic fibroblasts. Curr Protoc Mol Biol *Chapter 28*, Unit 28 21.

Yasui, D., Miyano, M., Cai, S., Varga-Weisz, P., and Kohwi-Shigematsu, T. (2002). SATB1 targets chromatin remodelling to regulate genes over long distances. Nature *419*, 641-645.

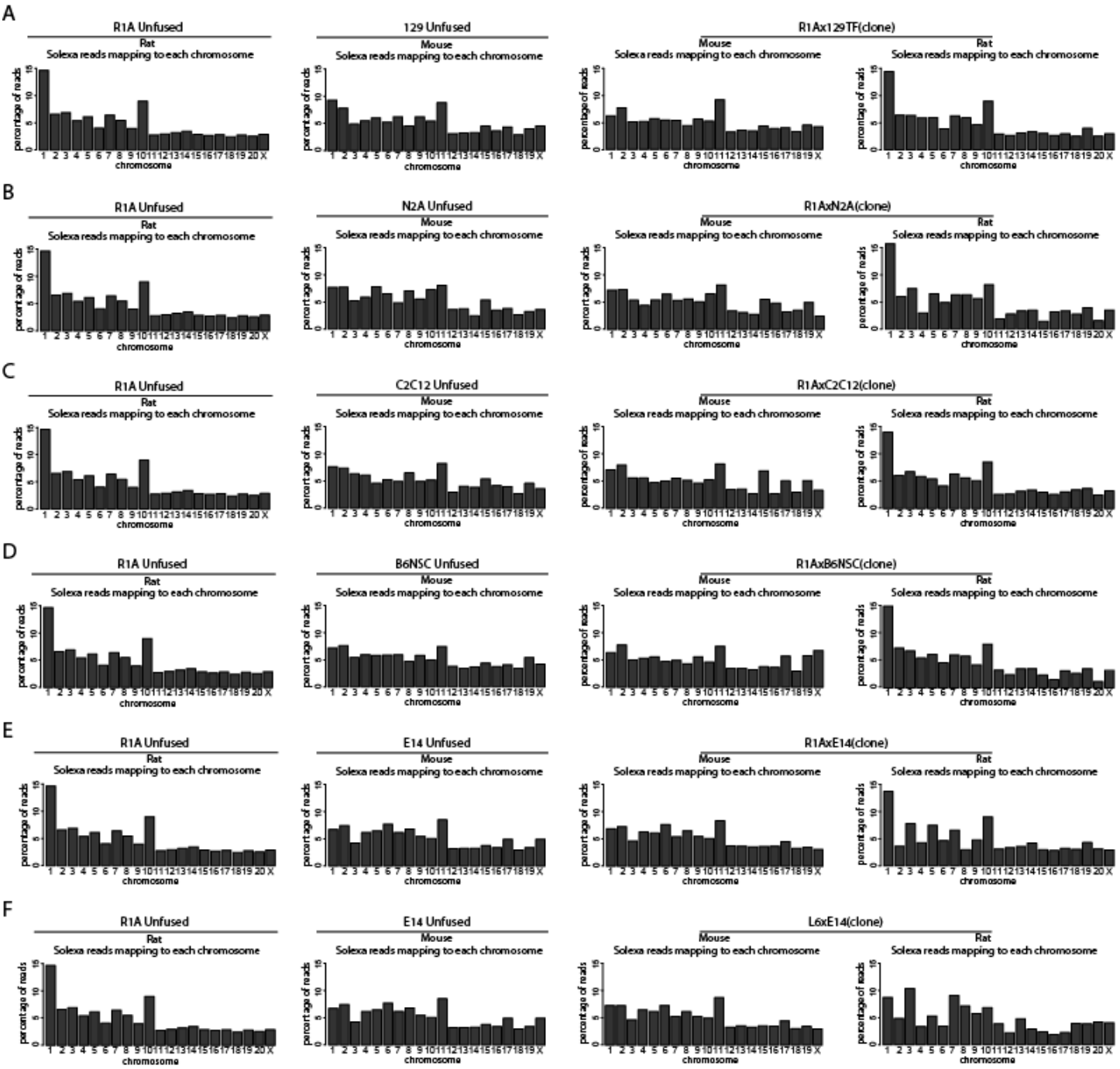

Figure S1. Bioinformatic Chromosome Analysis Showing that Fusion Clones Express Genes from All Chromosomes, Related to Figure 1

(A–F) Graphs were generated by mapping RNA-Seq reads to a reference library consisting of the combined mouse-rat genomes. The proportion of these reads mapping to each chromosome was then tallied and displayed as a bar plot. Data from unfused parental cell lines are presented as a control. All fusion clones contain the full complement of chromosomes from each genome, and the proportion of reads mapped to each chromosome is generally comparable to that of the unfused parental cell line.

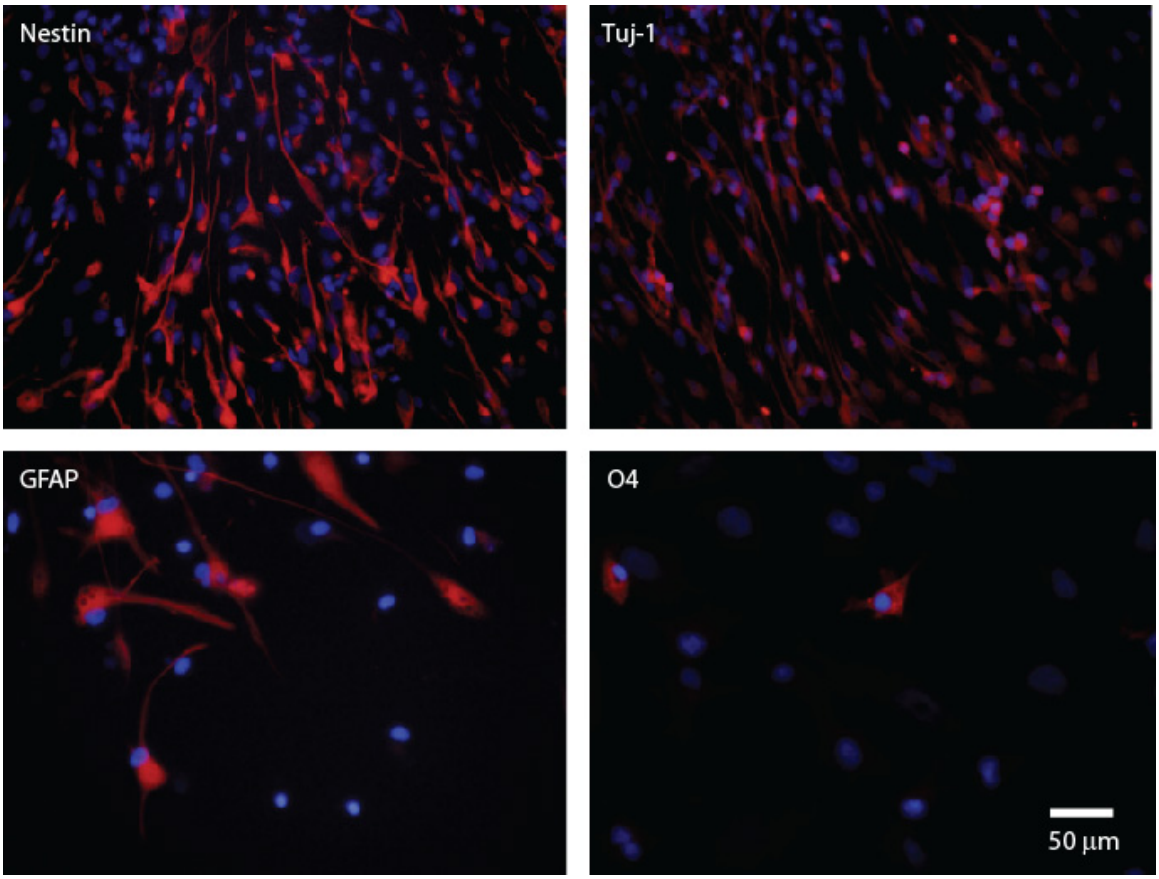

Figure S2. Demonstration of the Multipotential Nature of B6NSC Mouse Neural Stem Cells, Related to Figure 1

Before differentiation, cells were stained for nestin (marker for neural stem cells), and after differentiation, for Tuj-1 (neurons), GFPA (astrocytes), and O4 (oligodendrocytes). Antibody staining is in red while DAPI counterstaining is in blue.

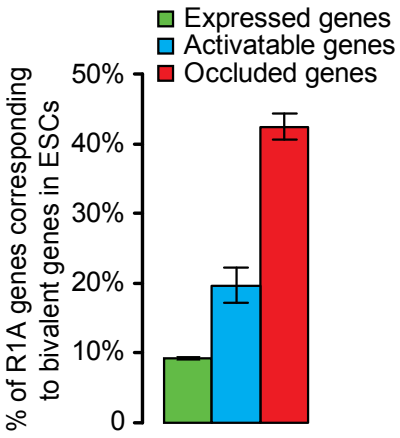

Figure S3. More Occluded than Activatable R1A Genes Correspond to Gene Possessing Bivalent Domains in Mouse ESCs, Related to Results (Behavior of Chromatin-Modifying Genes in Somatic-ESC Fusions)

Table S5. Expression Levels (Transcripts per Genome) of Chromatin Modifiers Activated in the R1A Genome by R1A-E14 Fusions, Related to Results (Behavior of Chromatin-Modifying Genes in Somatic-ESC Fusions)

Gene name	Genome of origin	Unfused cells	Fused day 2	Fused day 4	Fused day 8	Fused clone
Satb1	E14	8.5	2.1	3.9	8	5.4
	R1A	0	0	0.6	8.8	30.8
Chd7	E14	7.2	6.4	10.5	8.3	11.1
	R1A	0	0	0.4	3.5	11.4
Dnmt3b	E14	33.2	21	16.8	38.9	40
	R1A	0	0	0.6	7	42.3
Tet1	E14	63.2	29.7	46.9	36.9	49.9
	R1A	1.4	0.4	0.5	4.3	40.9

Table S6. Expression Levels (Transcripts per Genome) of Chromatin Modifiers Activated in the L6 Genome by L6-E14 Fusion, Related to Results (Behavior of Chromatin-Modifying Genes in Somatic-ESC Fusions)

Gene name	Genome of origin	Unfused cells	Fused clone
Satb1	E14	8.5	2.6
	L6	0.09	17.3
Chd7	E14	7.2	4.5
	L6	1.3	5.7
Dnmt3b	E14	33.2	83.4
	L6	0.7	51.3
Tet1	E14	63.2	57.9
	L6	0	40.8