Supporting Information

Fernandez-Tresguerres et al. 10.1073/pnas.1010708107

SI Materials and Methods

Embryos and in Situ Hybridization. Embryos were staged according to Eyal-Giladi and Kochav (1) for pregastrulation stages (EGK series) and according to Hamburger and Hamilton (2) for post-gastrulation stages (HH series). Whole-mount in situ hybridization of chicken embryos was carried out as described (3, 4). A 488-bp chick *Pou2-related* (*Pou2-r*) clone was obtained by PCR (oligos: Pou2 forward CATGTGCAAGCTGAAGCCACTGCT; reverse TCAGTGGCTGCTGTTGTTCATGGAG). The *Fgf4* and *Nodal* probes were a kind gift of Paola Bovolenta (I. Cajal, Madrid), and the cDNAs for *Sex-determining region Y (SRY)-box containing gene 2* (*Sox2*) (clone ChEST878b12, accession number BU282995), *Forkhead box D3* (*FoxD3*) (clone ChEST68l3, accession number BU128393), and *Polyhomeotic-like 1* (*Phc1*) (clone ChEST49d22, accession number BU219008) were obtained from the UK Medical Research Council Geneservice (5). In all cases, embryos of different stages were processed in parallel, and sense probes were used as negative controls.

Expression Profiling by Microarrays. Twenty HH6 or 30 EGK-X stage embryos were dissected for each of three independent replicates. Blastoderm-derived cells (bdC) were purified as described (6). Briefly, 20 intact stage EGK-X areas pellucida for each of three independent replicates were placed into tubes containing PBS and 2% (vol/vol) chicken serum. After centrifugation, PBS was replaced with trypsin 0.05% and incubated for 5 min at room temperature. Dispersed cells were plated into 35-mm tissue-culture dishes and were cultured for 48 h in DMEM containing 10% (vol/vol) FBS, 2% (vol/vol) chicken serum, 100 U/mL penicillin, 100 μg/mL streptomycin, and leukemia inhibitory factor (LIF). The cultured cells were positive for alkaline phosphatase and showed high levels of expression as determined by quantitative PCR of *Pou2-r* and *Nanog*, as previously described (7, 8), but levels of *Sox2* were undetectable, in agreement with our whole-mount in situ analysis. Integrity and concentration of purified total RNA was determined by a 2100 Bioanalyzer (Agilent) and Nanodrop measurement (Agilent ND-1000 spectrophotometer), followed by amplification, hybridization to the Chicken (V2) Gene Expression Microarray, 4 × 44K, (G2519F; Agilent Technologies), scanning, and image analysis following the manufacturer's instructions. Preprocessing and statistical analysis of the data were performed at the Genomics Unit of the Centro Nacional de Investigaciones Cardiovasculares following standard procedures (9, 10).

Three-way comparisons of the expression data were performed with cutoffs of corrected $P < 0.01$ and $-2 > \text{logFC} > 2$ to identify genes overrepresented in one sample versus the others (Table S1). Sets of mouse genes that respond to down-regulation of octamer-binding protein 4 (Oct4; official name, POU domain, class 5, transcription factor 1), Sox2, or Nanog by RNAi or overexpression of Nanog in ES cells were obtained from Sharov et al. (11). Genes down-regulated in response to RNAi or up-regulated in the case of Nanog overexpression were considered to be activated by the core pluripotency factors and as such to be a signature of the embryonic pluripotent stage. Genes up-regulated in response to RNAi or down-regulated in the case of Nanog overexpression were considered to be repressed by the core pluripotency factors and as to be such a set of genes that should not be overexpressed in the embryonic pluripotent stage. The percentage of the orthologs of overrepresented chick genes in each sample (HH6, EGK-X, bdC) present in each mouse set was calculated independently for each embryonic pluripotency factor (Dataset S2) and jointly for all factors (Fig. 1*B*), with similar results. The sets of genes involved in different aspects of pluripotency were obtained from Tang et al. (12) and analyzed as described above.

Sequence Analysis. Chromosomal positions and genomic organization of chick and other vertebrate genes and genomic sequences from the *Fgf4* and *Sox2* loci were obtained from the latest release of the Ensembl genome browser (www.ensembl.org). Synteny analysis was aided by the use of the Genomicus browser, developed by the Dyogen laboratory in Paris (www.dyogen.ens.fr).

Genomic regions bound by Oct4 and Nanog in mouse ES cells were obtained from the ChIP-sequencing data of Marson et al. (13). We included in the analysis regions bound together by four (Oct4, Nanog, Sox2, and Tcf3; 814 regions), three (Oct4, Nanog, and Sox2; 638 regions) or two (Nanog and Oct4; 236 regions) core embryonic pluripotency gene-regulatory network (EP-GRN) transcription factors located 8 kb or less from known genes. We obtained similar results if each set was analyzed independently. The 200 top-scoring genomic regions bound by Gli3 in the mouse limb were obtained from the ChIP-sequencing data from Volkes et al. (14). Genomic positions bound by p300 and Suz12 were obtained from Chen et al. (15) and were extended to 500 bp centered on the bound position. Genomic regions bound by p300 in the developing limb were obtained from Visel et al. (16). We also analyzed genomic regions bound by p300 in the forebrain and the midbrain described in the same paper; results were similar to those obtained for the limb (Table S1). In these cases we used 200 random regions. The coordinates for all data sets were converted to the mouse mm7 (August 2005) assembly and were visualized using the Vista Tracks in the University of California Santa Cruz Genome Browser (17). These data sets were compared with rat, human, dog, and chick. Each genomic region was analyzed individually. A region was scored as conserved if it contained at least one segment of more than 100 bp showing more than 70% similarity in intergenic, intronic, or nontranslated regions. Conserved regions overlapping coding exons were not included. Regions were scored as mammalian specific if they were conserved between mouse and at least one of the two nonrodent species analyzed (human and dog).

For the analysis of the regions bound by 12 transcription factors in ES cells (15), we extended the bound position to 500 bp, removed overlapping coding regions, and checked if they were included in the set of conserved mouse-chick genomic regions downloaded from ECRbase (18).

Multispecies genomic alignments were performed using Vista tools (17), and sequences surrounding the Sox2/Oct4 sites were aligned using Clustal X (19) and were edited with BioEdit (20).

Reporter Constructs. Genomic fragments corresponding to the enhancers from mouse *Sox2* and *Fgf4* and the equivalent chick regions were amplified by PCR using the following BAC templates (http://bacpac.chori.org/) and primers:

mouse *Sox2*: BAC RP24-140C7, AAGGCACCAAGAACCAGAAAT and TCAGCAAGTCCTCTCTGGGTA
mouse *Fgf4*: BAC RP23-294B14, GGTGAAAATATGCACGACCAG and TGAATGCTTCTCTTTGGATGG
chick *Sox2*: BACs CH261-178A15 and CH261-110M, AGATTCAGGCATTTGATCTCG and AAACAAGCGGTGAATTCCTCT

Chick *Fgf4* was amplified form genomic DNA with ATGGTGTGAAAAGTGGCAAAG and TTTGGTGCAGTATTTGGAAGG.

The fragments were cloned in pGem-T Easy vector (Promega), and mutated versions were generated by site-directed mutagenesis (Mutagenex Inc.). The following changes (indicated in lowercase) were introduced:

mouse-to-chick *Fgf4*, CTTTGTTTGGATGCTAAT to aTTTGTTctatTGg-AAT (six substitutions and one deletion)
chick-to-mouse *Fgf4*, ATTTGTTCTATTGG-AAT to cTTTGTTtggaTGctAAT (six substitutions and one insertion)
mouse-to-chick *Sox2*, CATTGTGATGCATAT to tggTGTaAaGaAacc (nine substitutions)
chick-to-mouse *Sox2*, TGGTGTAAAGAAACC to catTGTgAtGcAtat (nine substitutions)

Wild-type and mutated versions then were cloned in a vector containing a human minimal β-globin promoter and the *lacZ* reporter gene (21), a kind gift of Robb Krumlauf (Stowers Institute, Kansas City, KS). These constructs were used for both ES cell transfections and, once linearized and the plasmid backbone removed, for the generation of transgenic embryos. As a positive control we used the previously described *Oct4* distal enhancer element (22) using BAC RP23-152G18 as template and primers AGCGGCCGCCTCTGCTACATGTAAATTTGTCT and AGCGGCCGCCTAAACAAGTACTCAACCCTTGAA.

ES Cell Culture and Embryo Transgenesis. Mouse ES cell line E14 was grown in gelatin-coated dishes without feeders on DMEM supplemented with LIF, 15% FCS, 1 mM glutamine, 1 mM sodium pyruvate, nonessential amino acids, and 100 μM β-mercaptoethanol. Transient transfections were performed as previously described (23), in this case using 12 μL of Lipofectamine 2000 (Invitrogen) and 2.5 μg of each construct. The pPyCAG-GFP vector was cotransfected in all cases as an internal control for efficiency. Cells were examined 48 h later, and cells expressing *lacZ* and GFP were counted in four random and independent fields (average 300 cells per field) for each well. The number of $lacZ^+$ cells was normalized with respect to GFP^+ cells, and the value for the *Oct4* distal enhancer was given an arbitrary value of 1. Negative controls included a mock transfection (only pPyCAG-GFP) and the empty vector. Three independent transfections were carried out in each case.

ZHBTc4 ES cells, in which Oct4 expression can be repressed by the addition of tetracycline (24), were transfected with the *Oct4* distal enhancer (Oct-DE) construct using monomeric red fluorescent protein as reporter, plus empty pCAGGS, mouse *Oct4*, or chicken *Pou2-r* cloned in the pCAGGS vector [kind gifts of Hisato Kondoh (Graduate School of Frontier Biosciences, Osaka University) and Joaquin Rodriguez-Leon (UEX, Badajoz), respectively]. Reporter activity was measured 48 h after transfection followed by growth with or without tetracycline (10 ng/mL). E14 ES cells were transfected in the same conditions with different combinations of the overexpressing constructs for mouse *Oct4* or chicken *Pou2-r* plus the different versions of the *Fgf4* and *Sox2* enhancers, as well as the equivalent chick genomic fragments. Relative activity of enhancer elements was determined as the proportion of positive cells for each combination of reporter and expression construct.

Transient transgenic blastocysts were generated by pronuclear injection as described in ref. 25. Each construct was microinjected at a concentration of 4 ng/μL, and embryos were cultured in microdrops of M16 medium (Sigma) covered with mineral oil (Sigma) at 37 °C, 5% CO_2 until blastocyst stage, fixed in 1% formaldehyde, 0.2% glutaraldehyde, 2 mM $MgCl_2$, 5 mM EGTA, and 0.02% Igepal for 5 min at room temperature, washed in PBS, and stained for β-galactosidase activity for 24 h at room temperature in the dark. A minimum of 50 blastocysts were used per construct to calculate the percentage of positive embryos. The empty vector containing only the minimal promoter and the *lacZ* reporter was used as a negative control.

1. Eyal-Giladi H, Kochav S (1976) From cleavage to primitive streak formation: A complementary normal table and a new look at the first stages of the development of the chick. I. General morphology. *Dev Biol* 49:321–337.
2. Hamburger V, Hamilton HL (1992) A series of normal stages in the development of the chick embryo. 1951. *Dev Dyn* 195:231–272.
3. Acloque H, Wilkinson DG, Nieto MA (2008) In situ hybridization analysis of chick embryos in whole-mount and tissue sections. *Methods Cell Biol* 87:169–185.
4. Ariza-McNaughton L, Krumlauf R (2002) Non-radioactive in situ hybridization: Simplified procedures for use in whole-mounts of mouse and chick embryos. *Int Rev Neurobiol* 47:239–250.
5. Boardman PE, et al. (2002) A comprehensive collection of chicken cDNAs. *Curr Biol* 12: 1965–1969.
6. Petitte JN, Liu G, Yang Z (2004) Avian pluripotent stem cells. *Mech Dev* 121: 1159–1168.
7. Lavial F, et al. (2007) The Oct4 homologue PouV and Nanog regulate pluripotency in chicken embryonic stem cells. *Development* 134:3549–3563.
8. Pain B, et al. (1996) Long-term in vitro culture and characterisation of avian embryonic stem cells with multiple morphogenetic potentialities. *Development* 122: 2339–2348.
9. Bolstad BM, Irizarry RA, Astrand M, Speed TP (2003) A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* 19:185–193.
10. Smyth GK (2004) Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol 3*: Article3.
11. Sharov AA, et al. (2008) Identification of Pou5f1, Sox2, and Nanog downstream target genes with statistical confidence by applying a novel algorithm to time course microarray and genome-wide chromatin immunoprecipitation data. *BMC Genomics* 9:269.
12. Tang F, et al. (2010) Tracing the derivation of embryonic stem cells from the inner cell mass by single-cell RNA-Seq analysis. *Cell Stem Cell* 6:468–478.
13. Marson A, et al. (2008) Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. *Cell* 134:521–533.
14. Vokes SA, Ji H, Wong WH, McMahon AP (2008) A genome-scale analysis of the cis-regulatory circuitry underlying sonic hedgehog-mediated patterning of the mammalian limb. *Genes Dev* 22:2651–2663.
15. Chen X, et al. (2008) Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* 133:1106–1117.
16. Visel A, et al. (2009) ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* 457:854–858.
17. Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I (2004) VISTA: Computational tools for comparative genomics. *Nucleic Acids Res* 32(Web Server issue):W273–279.
18. Loots G, Ovcharenko I (2007) ECRbase: Database of evolutionary conserved regions, promoters, and transcription factor binding sites in vertebrate genomes. *Bioinformatics* 23:122–124.
19. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 25:4876–4882.
20. Hall TA (1999) BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids. Symp. Ser.* 41:95–98.
21. Yee SP, Rigby PW (1993) The regulation of myogenin gene expression during the embryonic development of the mouse. *Genes Dev* 7(7A):1277–1289.
22. Yeom YI, et al. (1996) Germline regulatory element of Oct-4 specific for the totipotent cycle of embryonal cells. *Development* 122:881–894.
23. Robertson M, et al. (2006) Nanog retrotransposed genes with functionally conserved open reading frames. *Mamm Genome* 17:732–743.
24. Niwa H, Miyazaki J, Smith AG (2000) Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nat Genet* 24: 372–376.
25. Nagy A, Gertsensten M, Vintersten K, Behringer R (2003) *Manipulating the Mouse Embryo: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, Plainview, NY), 3rd Ed.
26. Okuda A, et al. (1998) UTF1, a novel transcriptional coactivator expressed in pluripotent embryonic stem cells and extra-embryonic cells. *EMBO J* 17:2019–2032.
27. van den Boom V, et al. (2007) UTF1 is a chromatin-associated protein involved in ES cell differentiation. *J Cell Biol* 178:913–924.
28. Nishimoto M, et al. (2005) Oct-3/4 maintains the proliferative embryonic stem cell state via specific binding to a variant octamer sequence in the regulatory region of the UTF1 locus. *Mol Cell Biol* 25:5084–5094.
29. Zhao Y, et al. (2008) Two supporting factors greatly improve the efficiency of human iPSC generation. *Cell Stem Cell* 3:475–479.
30. Kuntz S, et al. (2008) Tex19, a mammalian-specific protein with a restricted expression in pluripotent stem cells and germ line. *Stem Cells* 26:734–744.
31. Popova NV, Morris RJ (2004) Genetic regulation of mouse stem cells: Identification of two keratinocyte stem cell regulatory loci. *Curr Top Microbiol Immunol* 280:111–137.
32. Tanaka TS, et al. (2006) Esg1, expressed exclusively in preimplantation embryos, germline, and embryonic stem cells, is a putative RNA-binding protein with broad RNA targets. *Dev Growth Differ* 48:381–390.
33. Pierre A, et al. (2007) Atypical structure and phylogenomic evolution of the new eutherian oocyte- and embryo-expressed KHDC1/DPPA5/ECAT1/OOEP gene family. *Genomics* 90:583–594.
34. Maldonado-Saldivia J, et al. (2007) Dppa2 and Dppa4 are closely linked SAP motif genes restricted to pluripotent cells and the germ line. *Stem Cells* 25:19–28.

35. Masaki H, Nishida T, Kitajima S, Asahina K, Teraoka H (2007) Developmental pluripotency-associated 4 (DPPA4) localized in active chromatin inhibits mouse embryonic stem cell differentiation into a primitive ectoderm lineage. *J Biol Chem* 282:33034–33042.
36. Chakravarthy H, et al. (2008) Identification of DPPA4 and other genes as putative Sox2:Oct-3/4 target genes using a combination of in silico analysis and transcription-based assays. *J Cell Physiol* 216:651–662.
37. Inoue N, et al. (1999) New gene family defined by MORC, a nuclear protein required for mouse spermatogenesis. *Hum Mol Genet* 8:1201–1207.
38. Sayers EW, et al. (2010) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* 38:D5–D16.

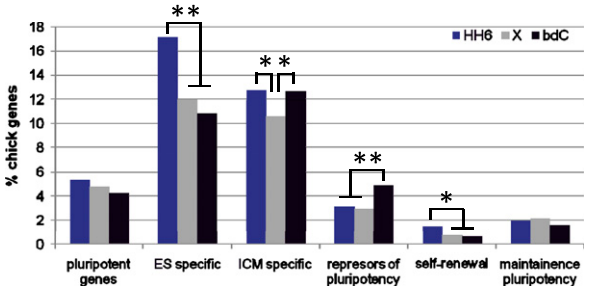

Fig. S1. Expression profiling shows no enrichment of orthologs of mouse pluripotency-related genes in the pregastrulation chick. The distribution of orthologs of chick genes overrepresented in HH6 and EGK-X embryos and in bdC in defined sets of pluripotency-related genes (12) shows that general pluripotent genes are equally distributed in the three sets of chick genes, but ES and self-renewal genes are overrepresented in the HH6 set. Genes classified as repressors of pluripotency are overrepresented in the bdC set. *$P < 0.05$; **$P < 0.01$ (two-tailed Fisher's exact test).

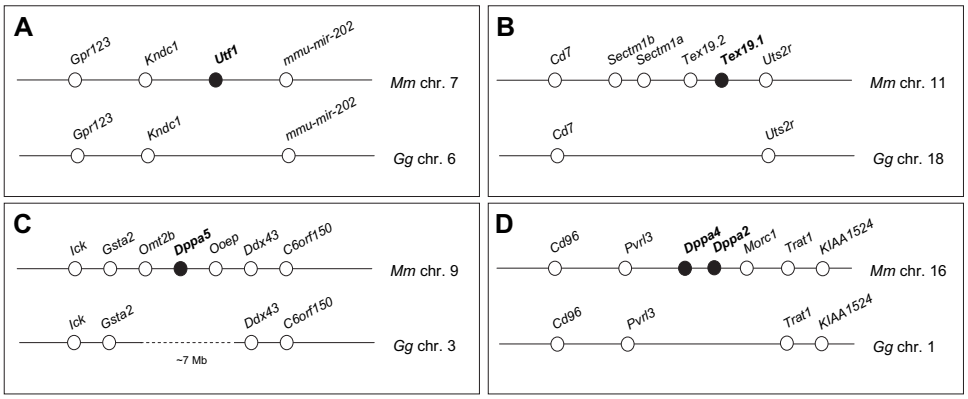

Fig. S2. Lack of EP-GRN orthologs in the chick genome. Mouse *Utf1* (*A*), *Tex19.1* (*B*), *Dppa5* (*C*), and the tandem duplicates *Dppa2* and *Dppa4* (*D*) lie in regions of conserved synteny between mouse and the chick but are absent from the latter. The mouse (*Mm*) and chick (*Gg*) chromosomes (chr.) where the genes are located are indicated. The diagrams are not shown to scale. *Utf1* is a chromatin-associated factor involved in controlling the initiation of ES cell differentiation and expressed in the inner cell mass (ICM) of the blastocyst (26, 27). It is a direct functional target of the Oct4/Sox2 transcriptional complex (28), and it recently has been shown to enhance the efficiency of human induced pluripotent cell generation (29). *Tex19.1* and its tandem duplicate *Tex19.2* have been described previously as mammal-specific and expressed in pluripotent stem cells (30). It is noteworthy that these genes are adjacent to another pair of tandem duplicates present in mouse but not in chick, *Sectm1a* and *Sectm1b* (*B*). These genes belong to the Ig superfamily and are closely related to the neighboring *Cd7* gene, which is conserved in chick. This observation indicates that this region has undergone extensive duplication and gene gain during mammalian evolution. *Dppa5*, identified because it has an expression pattern similar to that of *Oct4*, is flanked by *Oocyte maturation beta* (*Omt2β*) and *Oocyte-expressed protein homolog* (*Ooep*), two other mammal-specific genes (*C*). These genes are expressed in a pattern very similar to that of *Dppa5* in the embryo and ES cells (31, 32); furthermore, *Ooep* and *Dppa5* belong to the same family (33). *Dppa2* and *Dppa4* are tandem duplicated genes expressed in the ICM and later in primordial germ cells (34); interestingly, forced expression of *Dppa4* in ES cells drives them to a primitive ectoderm lineage (35). Furthermore, Dppa4 is a direct target of the Oct4/Sox2 transcriptional complex (36). The *Dppa2/4* duplication must have occurred late in mammalian evolution, because the opossum contains only one *Dppa2/4* gene. The neighboring *Microrchidia 1* (*Morc1*) gene is involved in spermatogenesis (37) and also has no ortholog in chick (*D*). Unigene-based EST expression profiling (38) shows that this gene also is expressed in the preimplantation mouse embryo.

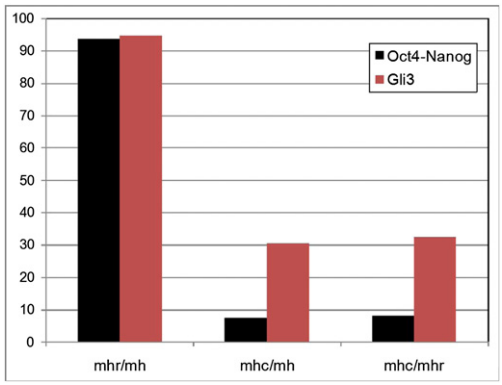

Fig. S3. Genomic regions bound by Oct4 and Nanog are poorly conserved in chick, irrespective of their conservation in mammals. The set of Oct4-Nanog–bound regions has an equal proportion of conserved mammalian noncoding elements as the Gli3 set (mhr/mh) but a much lower number of regions is conserved in chick (mhc/mh). Except for the mhr/mh comparison, all differences between the Oct4-Nanog and Gli3 data sets are highly significant ($P < 0.0001$). mh, genomic regions conserved between mouse and human; mhc, regions conserved between mouse and humans that also are conserved in chick; mhr, regions conserved between mouse and humans that also are conserved in rat.

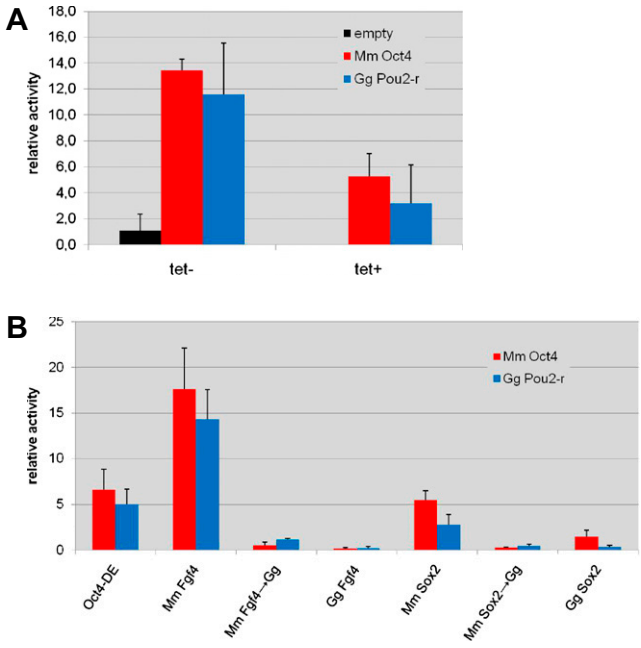

Fig. S4. Overexpression of mouse *Oct4* and chick *Pou2-r* has similar effects on pluripotency enhancers in ES cells. (*A*) The activity of the Oct4-DE in ZHBTc4 ES cells is increased by the overexpression of both mouse Oct4 and chick Pou2-r in similar degrees [tetracycline negative (tet-)]. Shutdown of endogenous *Oct4* expression by the addition of tetracycline (tet^+) abolishes the activity of the Oct4-DE (empty), but enhancer activity is recovered when mouse *Oct4* or chick *Pou2-r* are overexpressed. (*B*) Mouse *Fgf4* and *Sox2* pluripotency enhancers (Mm Fgf4, Mm Sox2) respond in a similar degree to overexpression of mouse *Oct4* or chick *Pou2-r* in E14 ES cells. Neither the chick genomic regions equivalent to the mouse enhancers (Gg Fgf4, Gg Sox2) nor the mouse enhancers where the Sox2/Oct4 site was changed to the equivalent chick sequence (Mm Fgf4→Gg, Mm Sox2→Gg) showed any increased response to overexpression of chick *Pou2-r*. The activity of the Oct-DE is shown as a control. Bars indicate SD.

Table S1. Conservation between mouse and chick of genomic regions surrounding the sites bound by 12 transcription factors involved in pluripotency of mouse ES cells

Factor	Total sites	Conserved sites	% Conserved sites
Nanog	10,343	285	2.76
Oct4	3,761	106	2.82
Sox2	4,526	128	2.83
Esrrβ	21,648	544	2.51
Zfx	10,338	231	2.23
Smad1	1,126	31	2.75
Stat3	2,546	63	2.47
Tcfcp2l1	26,910	688	2.56
E2f	20,699	493	2.38
Klf4	10,875	245	2.25
cMyc	3,422	79	2.31
nMyc	7,182	172	2.39

The proportion of conserved sites was calculated removing those that overlap coding regions.

Other Supporting Information Files

Dataset S1 (XLSX)
Dataset S2 (XLSX)