Evolution of the mammalian embryonic pluripotency gene regulatory network

Beatriz Fernandez-Tresguerres[a,1], Susana Cañon[a,1], Teresa Rayon[a], Barbara Pernaute[a,2], Miguel Crespo[a,3], Carlos Torroja[b], and Miguel Manzanares[a,4]

[a]Department of Cardiovascular Developmental Biology and [b]Bioinformatics Unit, Centro Nacional de Investigaciones Cardiovasculares, 28029 Madrid, Spain

Edited* by Janet Rossant, Hospital for Sick Children, University of Toronto, Toronto, ON, Canada, and approved October 6, 2010 (received for review July 22, 2010)

Embryonic pluripotency in the mouse is established and maintained by a gene-regulatory network under the control of a core set of transcription factors that include octamer-binding protein 4 (Oct4; official name POU domain, class 5, transcription factor 1, Pou5f1), sex-determining region Y (SRY)-box containing gene 2 (Sox2), and homeobox protein Nanog. Although this network is largely conserved in eutherian mammals, very little information is available regarding its evolutionary conservation in other vertebrates. We have compared the embryonic pluripotency networks in mouse and chick by means of expression analysis in the pregastrulation chicken embryo, genomic comparisons, and functional assays of pluripotency-related regulatory elements in ES cells and blastocysts. We find that multiple components of the network are either novel to mammals or have acquired novel expression domains in early developmental stages of the mouse. We also find that the downstream action of the mouse core pluripotency factors is mediated largely by genomic sequence elements nonconserved with chick. In the case of *Sox2* and *Fgf4*, we find that elements driving expression in embryonic pluripotent cells have evolved by a small number of nucleotide changes that create novel binding sites for core factors. Our results show that the network in charge of embryonic pluripotency is an evolutionary novelty of mammals that is related to the comparatively extended period during which mammalian embryonic cells need to be maintained in an undetermined state before engaging in early differentiation events.

Embryonic pluripotency is an essential property of a small group of cells of the mammalian blastocyst which transiently keeps them in an indeterminate, uncommitted state. This condition is a consequence of the earliest differentiation events taking place in the preimplantation embryo. The first lineage decision produces the inner cell mass (ICM) and the trophectoderm (TE). The TE produces most of the extraembryonic structures, mainly the placenta. In the second lineage choice, the ICM gives rise to the epiblast (EPI), which will generate the embryo proper, and the primitive endoderm (PE), another extraembryonic tissue (1).

Cells in the ICM and the EPI retain the full potential to develop into all embryonic tissues and germ layers and are the source of ES cells, which have the capacity of indefinite self-renewal and maintenance of pluripotency in tissue culture. Pluripotency results from the expression of a small network of transcription factors that actively maintain the undetermined state and at the same time repress the differentiation program (2). The core members of this network are the products of the *Oct4*; official name, *Pou5f1*, *Nanog*, and *Sox2* genes. These factors act together through auto- and cross-regulatory interactions and also through direct and overlapping binding to multiple locations throughout the genome, where they regulate downstream target genes (3–5).

Thus far, little is known about the evolutionary conservation of the embryonic pluripotency gene regulatory network (EP-GRN); reports are limited to the description of nonmammalian vertebrate homologs of the core mammalian EP transcription factors Oct4 and Nanog. Although the biochemical properties of these factors appear to be conserved to some extent between mouse and other vertebrates (6–8), other aspects, such as the territories of early expression, are not (9).

We examined the pregastrulation chicken embryo for the expression of homologs of important mouse EP-GRN genes. Genomic analysis shows that, although some EP-GRN genes are specific to mammals, core pluripotency factors are present across amniotes. Nonetheless, the expression patterns of these homologs in the early chicken embryo are incompatible with a role in establishing embryonic pluripotency; moreover, the genomic regions that bind core EP transcription factors are poorly conserved between mammals and chick, and key EP regulatory elements have appeared de novo in mammals.

Results

Orthologs of Mammalian EP-GRN Genes Are Not Expressed in the Pregastrulation Chicken Embryo. To carry out a comparison between the early pregastrulation chicken embryo and the mouse pluripotent state of the blastocyst ICM and ES cells, we examined the expression of the chick orthologs of the core components of the EP-GRN at pregastrulation stages. The expression of an *Oct4*-related gene [called *Pou2-related* (*Pou2-r*)] and of *Nanog* had been already described (7, 9), but no studies report the expression of the other core factor of the network, *Sox2*, in early chicken embryos. In contrast with *Pou2-r* (10), which we find expressed throughout the embryo at Eyal-Giladi–Kochav stage X (EGK-X) and later strongly up-regulated in the area pellucida (Fig. 1*A*) as described (7), no expression of *Sox2* was detected until Hamburger–Hamilton stage 4 (HH4), when it is strongly expressed in the neural plate (11).

We next studied the expression of orthologs of mouse genes that act downstream of the core factors, either in maintaining pluripotency, such as *FoxD3* (12), or acting as the first triggers of differentiation, such as the signaling molecules Fgf4 or Nodal that are expressed in the ICM of the blastocyst (13, 14). We also analyzed the polycomb group gene *Phc1*, because it is a direct target of the core EP factors (3, 4), is expressed in ES cells, and maps close to the *Gdf3-Dppa3* pluripotency gene cluster (9).

None of these genes is expressed in the pregastrulation chicken embryo (stage EGK-X) but only at later stages, in line with previously described patterns and domains (Fig. 1*A*) (15–18). These genes, unlike *Nanog* and *Oct4*/*Pou2-r*, are not expressed in primordial germ cells (PGC) of either chick or mouse. These results show that orthologs of many genes involved in embryonic pluripotency and early lineages in the mouse, including *Nanog*, are not expressed in equivalent territories of the early chicken embryo (9).

EVOLUTION

Author contributions: B.F.-T., S.C., T.R., B.P., and M.M. designed research; B.F.-T., S.C., T.R., B.P., and M.M. performed research; B.F.-T., S.C., T.R., B.P., M.C., C.T., and M.M. analyzed data; and B.F.-T., S.C., and M.M. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

[1]B.F.-T. and S.C. contributed equally to this work.

[2]Present address: Molecular Embryology Group, Medical Research Council Clinical Sciences Centre, Hammersmith Hospital Campus, Imperial College London, London W12 ONN, United Kingdom.

[3]Present address: Department of Neurology and Neuroscience, Weill Cornell Medical College, New York, NY 10065.

[4]To whom correspondence should be addressed. E-mail: mmanzanares@cnic.es.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1010708107/-/DCSupplemental.

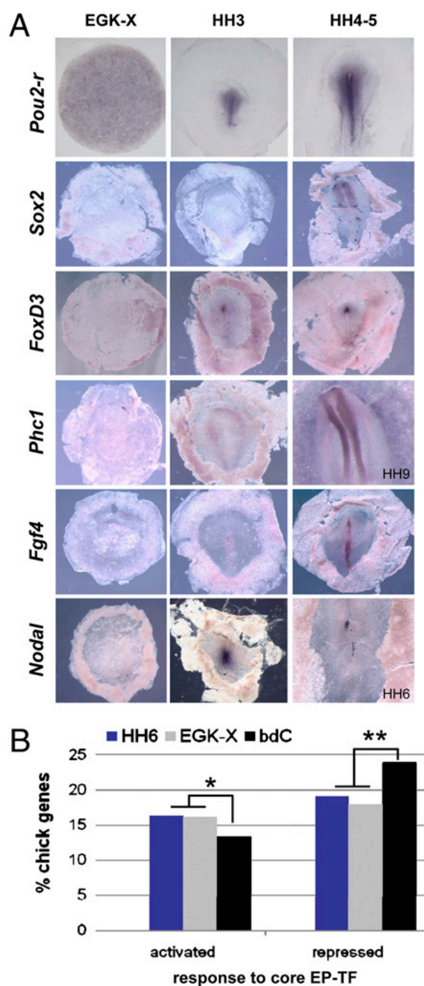

Fig. 1. Chick orthologs of mouse EP-GRN genes are not enriched in early pregastrulation embryos. (*A*) *Pou2-r* is expressed in the earliest EGK-X stage embryos and shortly thereafter is strongly up-regulated in the area pellucida. *Sox2* is first expressed in the neural plate at HH5, *FoxD3* in the node at HH3, *Phc1* in the neural tube at later stages (HH9), *Fgf4* in the primitive streak at HH4, and *Nodal* in the early primitive streak and later is restricted to its anterior portion (HH6). (*B*) Global analysis of gene expression shows that chick orthologs of mouse genes that are activated by core pluripotency factors (EP-TF: Oct4, Sox2, and Nanog) are equally represented in genes up-regulated in HH6 or EGK-X embryos and down-regulated in blastoderm-derived cells (bdC). Genes repressed by core factors show the same trend but are overrepresented in up-regulated genes in bdC. *$P < 0.05$; **$P < 0.01$ (two-tailed Fisher's exact test).

We extended this in situ screen by carrying out global expression analysis on stage EGK-X and HH6 embryos (Dataset S1) and on blastoderm-derived cells that have been claimed to represent the chick equivalent of mouse ES cells (19). We then compared the genes that were up-regulated in each of the chick samples with those in mouse ES cells that respond to changing the levels of the core pluripotency factors Oct4, Sox2, and Nanog (20). We found that genes positively regulated by the core factors, and therefore candidates to be part of the EP-GRN, are not overrepresented in EGK-X embryos compared with HH6 embryos and even are underrepresented in blastoderm-derived cells. On the other hand, genes that are negatively regulated and possibly are involved in differentiation events after the stage of embryonic pluripotency are not enriched in HH6 embryos as compared with EGK-X embryos and are overrepresented in blastoderm-derived cells (Fig. 1*B* and Dataset S2). We found similar trends when we compared the chick dataset with the results of a recent single-cell analysis of mouse embryonic pluripotent cells (21). Genes defined as pluripotent and ES or ICM specific are not overrepresented in chick EGK-X samples as compared with the postgastrulation HH6 embryos. Surprisingly, blastoderm-derived cells show enrichment in genes classified as repressors of pluripotency and a lower proportion of ES-specific or self-renewal genes as compared with the HH6 embryo (Fig. S1 and Dataset S2). These results show that the global expression profile of chick pregastrulation EGK-X embryo or blastoderm-derived cells is not more similar to that of mouse embryonic pluripotent stages than that of postgastrulation chicken embryos.

Several Pluripotency-Related Genes Are Specific to Mammals. We next searched for chick orthologs of mouse genes that are downstream of the core factors as part of the EP-GRN or that are expressed in patterns similar to these core factors and have been extensively used as markers of the pluripotent state. We found that in many cases no chick ortholog is present in the available genomic and transcriptional databases. For example, the chromosomal regions surrounding the mouse genes *Utf1*, *Tex19.1*, *Dppa2*, *Dppa4*, and *Dppa5* conserve synteny with the chick genome, but none of these genes is present in chick, although their immediate neighbors are (Fig. S2). Extensive searches found no evidence for orthologs of these genes in other genomic positions or in other databases. Other mouse EP-GRN genes that are not found in the chicken genome, such as *Dppa1*, *Rex1*, or *Nac1*, are located in regions that show no syntenic conservation at all with chick.

Core EP-GRN–Binding Cassettes Have Been Acquired in the Mammalian Lineage. In light of the above results, we can predict that the genomic regions bound by pluripotency factors (most likely corresponding to pluripotent specific *cis*-regulatory elements) will be poorly conserved between mouse and chick. We therefore examined the evolutionary conservation of 1,688 noncoding genomic regions bound by Oct4 and Nanog in ES cells (5). Of these regions, 11.55% are not conserved between mouse and any other species, more than half (53.26%) are conserved only in rodents (mouse and rat), approximately a third (32.7%) are conserved in mammals (mouse, rat, human, and dog), and only 2.49% are conserved in chick (Fig. 2 and Dataset S2). This figure is extremely low compared with prior estimates of the overall conservation of noncoding regulatory elements in human, rodent, and chick (22). We confirmed these results using the data from an independent study for the genomic binding in ES cells of 12 transcription factors involved in pluripotency (23). In all cases, conservation of these regions in the chick genome was around 2.5% (Table S1).

We next compared these results with a similar dataset from a GRN known to control a biological process well conserved between mouse and chick. For this comparison we chose the *cis*-regulatory circuitry underlying limb patterning. Applying the approach described above to the 200 top-scoring genomic regions bound by the limb development GRN transcription factor GLI-Kruppel family member GLI3 (Gli3) (24), we found that 26% of these regions are conserved between mouse and chick (Fig. 2 and Dataset S2). It also is noteworthy that a higher proportion of the Gli3-bound regions are conserved in the other mammals analyzed (60.5%, versus 32.7% for Oct4 and Nanog). This comparison shows that regions bound by Oct4 and Nanog in ES cells are significantly less conserved than those bound by Gli3 in the developing limb (in all cases, $P < 0.0001$; two-tailed Fisher's exact test).

We also examined the regions bound in the limb by the transcriptional coactivator p300 that is found at active enhancer elements in a tissue-specific fashion (25). As with Gli3, a large proportion of the regions bound were conserved in chick (22.5%; Fig. 2 and Dataset S2). When we analyzed regions bound by p300 in ES cells (23), we found a degree of conservation similar to that of pluripotency factors (3%; Fig. 2 and Dataset S2). However, the binding of a different chromatin regulator, Suppressor of zeste

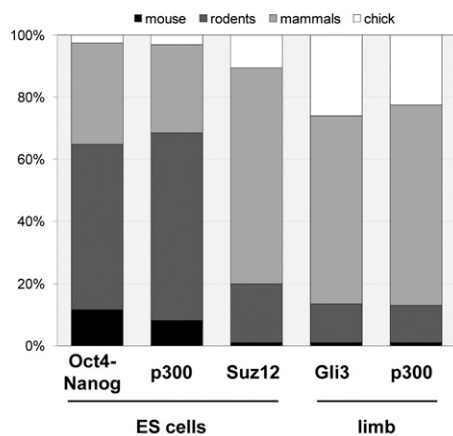

Fig. 2. Genomic regions bound by Oct4 and Nanog are poorly conserved in chick. Compared with regions bound by Gli3 in the limb, a high proportion of Oct4-Nanog–bound or p300-bound regions in ES cells are specific to mouse (black) or are conserved only in rodents (dark gray). Very few of the Oct4-Nanog–bound regions are conserved in chick, compared with those bound by Gli3 (2.49 vs. 26%), and a similar trend is found for p300 binding in ES cells as compared with limbs (3.05 vs. 22.50%). On the other hand, the conservation profile of regions bound in ES cells by Suz12, a chromatin regulator not involved in pluripotency, is more similar to that found for Gli3 and p300 in the limb than for Oct4/Nanog and p300 in ES cells.

12 homolog (Suz12) (23), not involved in pluripotency but involved in ES differentiation programs (26) shows a conservation profile much more similar to that of Gli3 or p300 in the limb than to EP-GRN factors (Fig. 2 and Dataset S2). These results suggest that not only the binding of specific factors but the set of active enhancers in ES cells is much less conserved when compared with chick than those involved in other developmental GRN.

However, there are at least two alternative explanations for the data presented above. One possibility is that the majority of regions bound by Oct4 and Nanog in the mouse genome are *cis*-regulatory elements controlling mouse-specific genes that have no chick orthologs. This explanation is not valid, because the proportion of 1-to-1 chick orthologs is higher in the genes located in the vicinity of Oct4-Nanog–bound regions than in the total mouse gene set (P value < 0.0001; two-tailed χ^2 test). The second possibility is that the low conservation of Oct4-Nanog regions in the chick genome could simply be an artifact of their enrichment in mouse- or rodent-only conserved regions (Fig. 2). To test for this possibility, we calculated the proportion of Oct4-Nanog regions conserved between mouse and human that also were conserved with rat or with chick (Dataset S2). Applying this same strategy to whole genome comparison, it has been previously shown that 23.29% of all noncoding regions conserved among mouse, human, and rat are also conserved in chick (27). In our case we find that for Oct4-Nanog regions this figure is 8.02% and for Gli3 regions it is 32.3% (Fig. S3 and Dataset S2). Therefore the lack of conservation of Oct4-Nanog–bound regions between mouse and chick is still significant when mouse- and rodent-specific genomic regions are excluded from the analysis (P value < 0.0001; two-tailed Fisher's exact test).

Conservation and Evolution of the *Fgf4* and *Sox2* EP Enhancers. We next tried to identify specific cases where we could trace the appearance of a core EP-GRN–regulated element. We focused on elements that have been shown to possess *cis*-regulatory activity in vivo in response to core EP factors. The best characterized set of such regulatory elements corresponds to Sox2/Oct4-regulated genes in which adjacent high-mobility group (HMG) and POU-binding motifs mediate high transcriptional activity (28). To date, functional HMG/POU cassettes have been characterized in detail from eight Sox2/Oct4 target genes, namely *Sox2* and *Oct4* themselves plus *Nanog*, *Lefty1*, *Fgf4*, *Fbxo15*, *Utf1*, and *Dppa4* (ref. 29 and references therein). Only for *Fgf4* and *Sox2* we found that the genomic region containing the HMG/POU cassette shows a partial but significant degree of conservation between mouse, other mammals, and chick (Fig. 3 *A* and *B*). In other cases, either there is no ortholog in chick [as for *Dppa4* or *Utf1* (Fig. S2) or for *Oct4*, which is not a true ortholog but a paralogue of chick *Pou2-r* (10)], or no similarity at all between mouse and chick genomic regions can be detected (as occurs for *Nanog*).

The *Fgf4* HMG/POU cassette is located in the 3′ UTR of the gene, and genomic fragments carrying these sites drive reporter expression in ES cells and the ICM; this expression is strictly dependant on the synergistic action of Sox2 and Oct4 (29–31). To examine the conservation of these sites and surrounding regions, we aligned the 3′ UTR of mouse *Fgf4* (2.3 kb) to 3 kb of genomic sequence immediately downstream from the stop codon of the coding region of *Fgf4* orthologs from other vertebrates (Fig. 3*A*). We found various peaks of conservation distributed along the length of the mouse 3′ UTR and noticed that the Sox2/Oct4-binding site (the thin blue line in Fig. 3*A*) was adjacent to a region conserved among mammals and chick. In this last case, sequence conservation was sufficient to anchor an alignment that contained the HMG/POU cassette (Fig. 3*C*). The HMG half of the cassette is identical in all species except for a single change in the 5′ end (A instead of C) in the chick sequence. Intriguingly, this residue is invariable in all Sox2/Oct4 composite sites identified to date in mouse (32). In the POU site, both chick and platypus show several nucleotide changes and a one base deletion.

The *Sox2* HMG/POU cassette localizes ≈2.5 kb downstream of the stop codon in mouse and is active in pluripotent cells (33). We aligned 4 kb downstream from the stop codon of the mouse gene to the equivalent regions from other vertebrates. As with *Fgf4*, conservation was distributed along the length of the sequence, and the Sox2/Oct4-binding site was located in a region highly conserved among all mammals and poorly but still detectably conserved in chick and lizard (Fig. 3*B*). This region does not correspond to any of the multiple conserved elements previously described as acting as enhancers in the early chicken embryo (11). Closer examination showed that, unlike the *Fgf4* enhancer, the *Sox2* enhancer of non-eutherian mammals contains a perfect POU site, and the HMG site contains only two changes; however, once again, these changes affect the C residue in the first position, which is invariant in all other Sox2/Oct4 sites (Fig. 3*D*). The chick sequence is altered with respect to mouse at multiple positions in both the HMG and the POU sites (Fig. 3*D*).

Activity of the *Fgf4* and *Sox2* EP Enhancers in Pluripotent Cells. Based on studies of other Sox2/Oct4 elements, the nucleotide changes we observe in chick compared with the mouse *Fgf4* and *Sox2* enhancers would be predicted to abolish Oct4 and Sox2 binding and therefore enhancer activity (29, 32, 34, 35). To test this hypothesis, we cloned the corresponding mouse and chick genomic regions from *Fgf4* and *Sox2* in an enhancer-detection vector and tested their activity in mouse ES cells and in preimplantation embryos (Fig. 4). As a positive control, we used the previously described *Oct4* distal enhancer (*Oct4*-DE; Fig. 4 *A* and *B*) that contains the Sox2/Oct4-binding sites and is active both in ES cells and in the ICM of the blastocyst (36). As expected, both mouse fragments were able to drive expression of the reporter in ES cells, but only the *Fgf4* enhancer was active in the blastocyst transgene assay we used (Fig. 4 *A* and *C*). The chick fragments, on the contrary, showed much lower activity in ES cells, and the chick *Fgf4* fragment was not active in the blastocyst (Fig. 4 *A* and *D*). We then proceeded to change the sequence of the Sox2/Oct4 sites from the mouse *Fgf4* and *Sox2* enhancers to that of the equivalent position in the chick sequences (Fig. 3 *C* and *D*). This change dramatically reduced expression of the reporter in ES cells (Fig. 4*A*) and abolished activity of the *Fgf4* enhancer in the blastocyst. To test if the mouse sequence of these sites was sufficient for activity in ES cells, we mutated the corresponding base pairs in the chick fragments to those found in mouse (Fig. 3 *C* and *D*). However, this mutation was not sufficient to make the

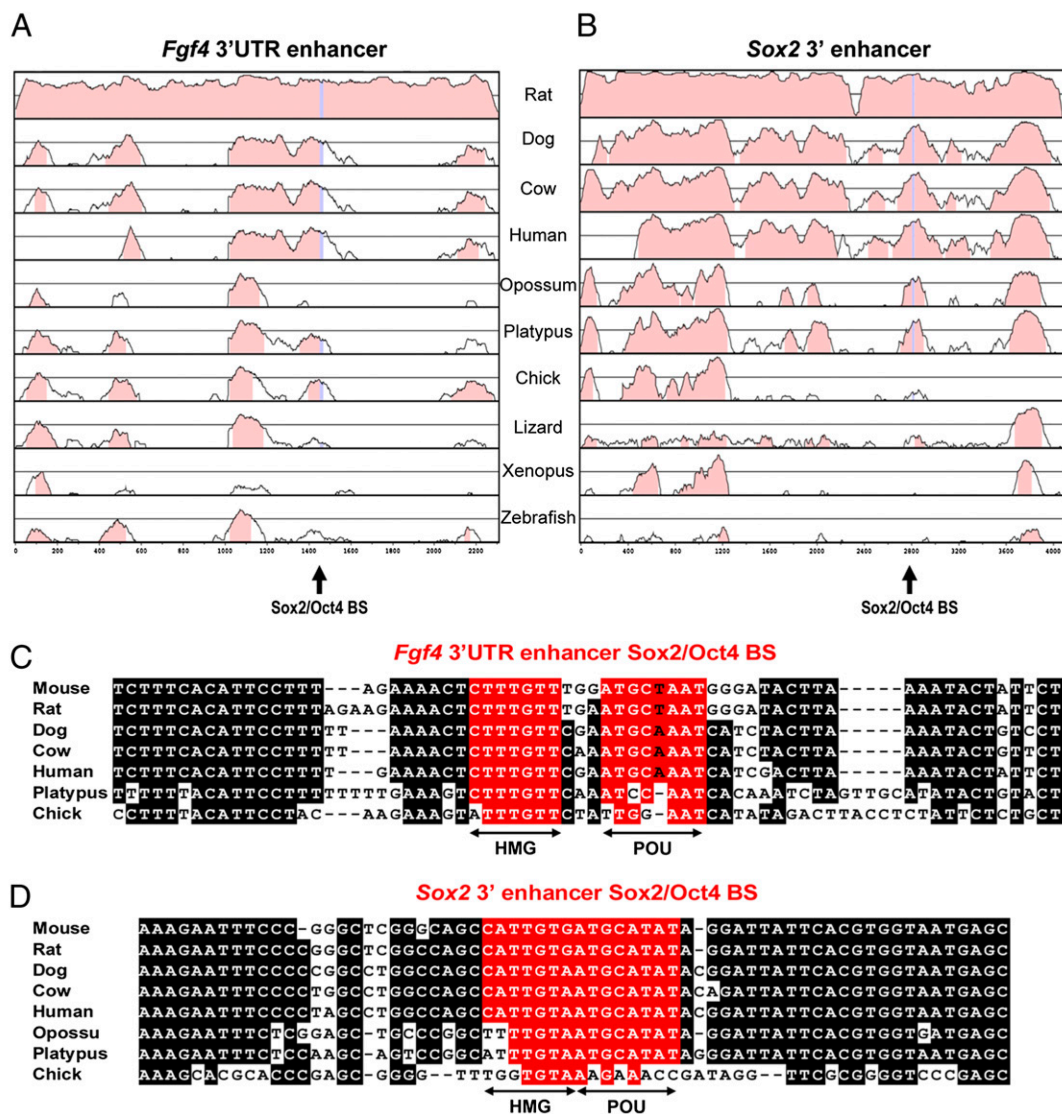

Fig. 3. Conservation of the *Fgf4* and *Sox2* EP enhancers. (*A* and *B*) Vista plots comparing the 3′ UTR from mouse *Fgf4* (*A*) and the 4 kb downstream of the mouse *Sox2* stop codon (*B*) with orthologous regions from other vertebrates. Regions of ≥100 bp that show ≥60% (≥50% in the case of nonmammalian species) sequence identity to the mouse sequence are colored pink. The Sox2/Oct4-binding site (BS) is shown in blue. (*C* and *D*) Sequence comparison of the regions surrounding the Sox2/Oct4 BS from *Fgf4* (*C*) and *Sox2* (*D*). The HMG and POU sites are indicated below the sequence and are highlighted in red. Residues identical across at least five species are shown with a black background. Dashes indicate gaps introduced to maximize the alignment. The opossum and zebrafish *Fgf4* sequences do not align in this region.

chick fragments active in ES cells, because the mutated fragments showed no significant changes as compared with the wild types (Fig. 4*A*), and the mutated chick *Fgf4* fragment was not active in the blastocyst. This result suggests that sequences other than the Sox2/Oct4 cassette contained in the mouse enhancers and not conserved with chick also are needed for correct expression in pluripotent cells.

Although mouse and chick Sox2 peptides are 92% identical, ruling out drastic changes in binding specificities, Oct4-related factors show much greater divergence (7). We therefore tested the function of mouse Oct4 and chick Pou2-r on the pluripotency enhancers described above. Both proteins were equally active on the *Oct4*-DE and rescued enhancer activity when endogenous Oct4 expression was shut down (Fig. S4*A*). On the other hand, overexpression of chick Pou2-r did not have any effect on chick genomic fragments from *Fgf4* or *Sox2* or on the mutated mouse-to-chick version of the enhancers (Fig. S4*B*).

This analysis thus has identified two examples in which overall similarity allows us to identify unambiguously the chick genomic region corresponding to the mouse Sox2/Oct4-responsive elements active during pluripotency but lacking critical sequences required for binding by Sox2 and Oct4. The number of residues that change from chick to mouse is small enough to suggest that changes to these sequences by point mutation and insertion led to the appearance of novel regulatory elements under the control of Sox2 and Oct4 as part of the emerging EP-GRN during mammalian evolution.

Discussion

Pluripotency is a necessary and transient stage in the development of any multicellular organism that passes through an obligatory one-cell stage as part of its life cycle. Research in the mouse and other mammals has identified the main factors and signals involved in embryonic pluripotency, but we still have little insight into how deeply conserved the EP-GRN is and how it appeared during evolution. Recent studies in nonmammalian vertebrates argue for conservation of pluripotency and its genetic control (6, 7), whereas comparative genome-wide studies be-

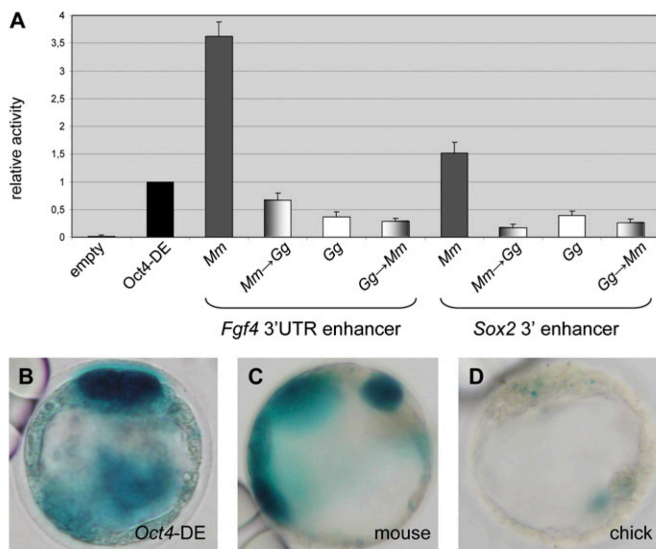

Fig. 4. Activity of the *Fgf4* and *Sox2* EP enhancers. (*A*) Relative enhancer activity in ES cells of the genomic fragments from the *Fgf4* and *Sox2* locus of mouse (*Mm*) and chick (*Gg*), as well as versions where the mouse Sox2/Oct4 site was changed to the equivalent chick sequence (*Mm*→*Gg*) or the chick sequence was changed to include the mouse Sox2/Oct4 site (*Gg*→*Mm*). Transfection efficiency was normalized and expressed as relative value with respect to the activity of the *Oct4* distal enhancer (*Oct4*-DE). (*B*–*D*) Enhancer activity in transgenic mouse blastocysts of the mouse *Fgf4* 3′ UTR enhancer (*C*) and the equivalent chick genomic region (*D*). The embryo shown in *D* is representative for the unspecific punctuated pattern equal to that obtained even when the empty vector is used, as well as for the mutated versions of both the mouse and chick fragments. The activity of the mouse *Oct4*-DE in the inner cell mass is shown as a positive control (*B*).

tween different mammals point to a high degree of plasticity of the EP-GRN (37, 38). Our results comparing mouse and chick help clarify this apparent contradiction, because they show that, although core pluripotency factors are present in other vertebrates, the way they are connected in the EP-GRN is unique to mammals. Co-option, duplication, and the appearance of novel genes, as well as new regulatory interactions between them, have occurred during the evolution of the EP-GRN.

The freshly laid chicken egg contains an embryo that has not started gastrulation and in which two distinct concentric regions can be observed: the central area pellucida, from which the embryo proper will develop, and the exterior area opaca, which will form the majority of the extraembryonic membranes. This arrangement is topologically similar to the mouse blastocyst, with the separation of embryonic and extraembryonic lineages, and has been proposed to align more precisely with the early developmental stages of other vertebrate embryos (39). The lack of pregastrulation expression of orthologs of EP-GRN genes by this stage (EGK-X) rules out a role for them in establishing the epiblast fate of the chicken embryo before gastrulation as happens in mouse. Furthermore the pregastrulation chicken embryo does not show a general enrichment in the expression of orthologs of mouse pluripotency-related genes, again suggesting that the EP-GRN is not deployed in the same manner in the early mouse and chicken embryo.

Some of the genes of the mouse EP-GRN are novel and specific to mammals and in many instances, such as *Gdf3* (9) and *Rex1* (40), arose by gene duplication. It is tempting to speculate that such gene duplications were major events behind the emergence of the pluripotency regulatory network, as has been suggested in the case of *Oct4* (10). However, targets of core pluripotency factors (5) are not enriched in mammalian-specific genes. Therefore, if the EP-GRN was newly assembled in mammals, as the expression data suggest, it must have occurred not only through the appearance of novel genes but also through the recruitment of preexisting ones.

The most obvious way in which genes can gain new expression domains is through the appearance of novel *cis*-regulatory elements in their vicinity. We find that the genomic regions bound by the core EP-GRN factors show very little conservation between mouse and chick. Furthermore, they show lower conservation among mammals than evolutionarily older GRN such as that controlling limb development, in line with recent findings on the rewiring of peripheral components of the EP-GRN that could be mediated in part by transposons (37, 38). We found two examples that confirm our prediction that novel elements regulated by core pluripotency factors have appeared in mammals. Although the sequences surrounding the Sox2/Oct4 cassette are conserved between mammals and chick in both the *Fgf4* and *Sox2* ES cell enhancers, the binding sites per se show critical nucleotide changes that render the chick versions inactive in functional assays in mouse ES cells and preimplantation embryos. Importantly, this lack of activity is not the result of binding differences between mouse and chick factors. Furthermore, we show that the Sox2/Oct4 cassette is necessary but not sufficient for activity in pluripotent cells. This result means that the mouse EP enhancers contain other nonconserved sequences necessary for activity that remain to be identified.

Our results raise the question of why a new GRN in charge of maintaining embryonic pluripotency would have arisen in mammals. We believe the answer may lie in the peculiarities of the early mammalian embryo. Most vertebrate embryos grow by a series of quick cell divisions (39), sustained by the yolk supplied in the egg, and patterning is initiated shortly after fertilization by maternal and/or external factors.

The situation is radically different in mammals, where eggs are devoid of yolk and intrauterine nourishment is provided by the TE-derived placenta. Initial cell divisions of the mammalian embryo are extremely slow (39), and until implantation there is hardly any growth. Recent evidence indicates that embryonic patterning in mammals is not initiated until periimplantation stages, long after fertilization (1). In fact, the first differentiation event to occur is the specification of the TE lineage, which is unique to mammals. Another key characteristic of the mouse embryo is that patterning is initiated by intrinsic and zygotic mechanisms: No external trigger or maternal signal is needed to establish the mayor body axes. Therefore in mammalian embryos the delay in pattern formation and the uncoupling from external cues necessitate prolonged maintenance of embryonic cells in an undetermined and quiescent state. The evolutionary appearance of the EP-GRN would have permitted this state, with the core factors (Oct4-Sox2-Nanog) silencing genes involved in early differentiation and specification (3, 4) but leaving them ready to be expressed in a controlled spatial and temporal fashion.

Some light on the origin of the EP-GRN can be shed by the fact that *Nanog* and *Pou5f1* paralogues are expressed in the PGC of mouse and chicken embryos (7, 9), but other EP-GRN genes are not expressed in the PGC in either species. A possible scenario is that *Nanog* and *Pou5*-like genes may act in PGC to prevent their differentiation and maintain their germline potential until differentiation of the mature germ cells. This function therefore may be an older, evolutionarily conserved role for these genes in vertebrates that later was co-opted and expanded to form the network responsible for the same function in early embryonic cells of mammals.

Materials and Methods

Full details of materials and methods are given in *SI Materials and Methods*.

Embryos and in Situ Hybridization. Embryos were staged according to Eyal-Giladi and Kochav (41) for pregastrulation stages (EGK series) and according to Hamburger and Hamilton (42) for postgastrulation stages (HH series).

Expression Profiling by Microarrays. Total RNA from 20 HH6 or 30 EGK-X stage embryos or blastoderm-derived cells from 20 EGK-X dissected areas pellucida were hybridized to the Chicken (V2) Gene Expression Microarray, 4 × 44K, (G2519F; Agilent Technologies). Three independent replicates for each stage

were used. Stage-specific overrepresented genes were selected by three-way comparison among samples with cutoffs of corrected *P* values < 0.01 and −2 > logFC > 2 (Dataset S1). Sets of mouse genes that respond to down-regulation of Oct4, Sox2, or Nanog by RNAi or overexpression of Nanog in ES cells were obtained from Sharov et al. (20), and sets of genes involved in different aspects of pluripotency were obtained from Tang et al. (21).

Sequence Analysis. Chromosomal positions, genomic organization and sequences were obtained from the latest release of the Ensembl genome browser (www.ensembl.org). Genomic regions bound by different factors were obtained from Oct4 and Nanog in ES cells (5); 12 factors, p300, and Suz12 in ES cells (23); Gli3 in limbs (24); and p300 in limbs (25).

Reporter Constructs and Enhancer Assays. Genomic fragments corresponding to the enhancers from mouse *Sox2* and *Fgf4* and the equivalent chick regions were amplified by PCR using BAC templates. The mutated versions were generated by site-directed mutagenesis. Wild-type and mutated versions then were cloned in a vector containing a human minimal β-globin promoter and the *lacZ* reporter gene, a kind gift from Robb Krumlauf (Stowers Institute, Kansas City, KS). These constructs were used for both ES cell transfections and for the generation of transgenic embryos.

Mouse E14 ES cells were grown in gelatin-coated dishes without feeders and were cotransfected with the pPyCAG-GFP vector as an internal control. Cells were examined 48 h later, and cells expressing *lacZ* and GFP were quantified. Three independent transfections were carried out in each case. Transient transgenic blastocysts were generated by pronuclear injection. A minimum of 50 blastocysts per construct was used to calculate the percentage of positive embryos.

ACKNOWLEDGMENTS. We thank Robb Krumlauf (Stowers Institute, Kansas City, KS), Ian Chambers (MRC-CRM, Edinburgh), Hisato Kondoh (Graduate School of Frontier Biosciences, Osaka University), Joaquín Rodriguez-Leon (UEX, Badajoz, Spain), and Paola Bovolenta (I. Cajal, Madrid) for reagents; the Genomics Unit of the Centro Nacional de Investigaciones Cardiovasculares for help with microarrays; Tristan Rodriguez, Miguel Torres, and Ian Chambers for comments and suggestions; and Simon Bartlett for critical reading of the manuscript. This work was supported by Grants BFU2008-00838 and CONSOLIDER-25120 from the Spanish Government and Grant CAM S-SAL-0190-2006 from the Regional Government of Madrid. The Centro Nacional de Investigaciones Cardiovasculares is supported by the Spanish Ministry of Science and Innovation and the ProCNIC Foundation.

1. Rossant J, Tam PP (2009) Blastocyst lineage formation, early embryonic asymmetries and axis patterning in the mouse. *Development* 136:701–713.
2. Niwa H (2007) How is pluripotency determined and maintained? *Development* 134: 635–646.
3. Boyer LA, et al. (2005) Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* 122:947–956.
4. Loh YH, et al. (2006) The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nat Genet* 38:431–440.
5. Marson A, et al. (2008) Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. *Cell* 134:521–533.
6. Dixon JE, et al. (2010) Axolotl Nanog activity in mouse embryonic stem cells demonstrates that ground state pluripotency is conserved from urodele amphibians to mammals. *Development* 137:2973–2980.
7. Lavial F, et al. (2007) The Oct4 homologue PouV and Nanog regulate pluripotency in chicken embryonic stem cells. *Development* 134:3549–3563.
8. Morrison GM, Brickman JM (2006) Conserved roles for Oct4 homologues in maintaining multipotency during early vertebrate development. *Development* 133: 2011–2022.
9. Cañón S, Herranz C, Manzanares M (2006) Germ cell restricted expression of chick Nanog. *Dev Dyn* 235:2889–2894.
10. Niwa H, Sekita Y, Tsend-Ayush E, Grützner F (2008) Platypus Pou5f1 reveals the first steps in the evolution of trophectoderm differentiation and pluripotency in mammals. *Evol Dev* 10:671–682.
11. Uchikawa M, Ishida Y, Takemoto T, Kamachi Y, Kondoh H (2003) Functional analysis of chicken Sox2 enhancers highlights an array of diverse regulatory elements that are conserved in mammals. *Dev Cell* 4:509–519.
12. Hanna LA, Foreman RK, Tarasenko IA, Kessler DS, Labosky PA (2002) Requirement for Foxd3 in maintaining pluripotent cells of the early mouse embryo. *Genes Dev* 16: 2650–2661.
13. Niswander L, Martin GR (1992) Fgf-4 expression during gastrulation, myogenesis, limb and tooth development in the mouse. *Development* 114:755–768.
14. Takaoka K, et al. (2006) The mouse embryo autonomously acquires anterior-posterior polarity at implantation. *Dev Cell* 10:451–459.
15. Chapman SC, Schubert FR, Schoenwolf GC, Lumsden A (2002) Analysis of spatial and temporal gene expression patterns in blastula and gastrula stage chick embryos. *Dev Biol* 245:187–199.
16. Kos R, Reedy MV, Johnson RL, Erickson CA (2001) The winged-helix transcription factor FoxD3 is important for establishing the neural crest lineage and repressing melanogenesis in avian embryos. *Development* 128:1467–1479.
17. Shamim H, Mason I (1999) Expression of Fgf4 during early development of the chick embryo. *Mech Dev* 85:189–192.
18. Tomotsune D, Shirai M, Takihara Y, Shimada K (2000) Regulation of Hoxb3 expression in the hindbrain and pharyngeal arches by rae28, a member of the mammalian Polycomb group of genes. *Mech Dev* 98:165–169.
19. Petitte JN, Liu G, Yang Z (2004) Avian pluripotent stem cells. *Mech Dev* 121: 1159–1168.
20. 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.
21. 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.
22. Hillier LW, et al.; International Chicken Genome Sequencing Consortium (2004) Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* 432:695–716.
23. Chen X, et al. (2008) Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* 133:1106–1117.
24. 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.
25. Visel A, et al. (2009) ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* 457:854–858.
26. Pasini D, Bracken AP, Hansen JB, Capillo M, Helin K (2007) The polycomb group protein Suz12 is required for embryonic stem cell differentiation. *Mol Cell Biol* 27: 3769–3779.
27. Prabhakar S, et al. (2006) Close sequence comparisons are sufficient to identify human cis-regulatory elements. *Genome Res* 16:855–863.
28. Reményi A, et al. (2003) Crystal structure of a POU/HMG/DNA ternary complex suggests differential assembly of Oct4 and Sox2 on two enhancers. *Genes Dev* 17: 2048–2059.
29. 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.
30. Fraidenraich D, Lang R, Basilico C (1998) Distinct regulatory elements govern Fgf4 gene expression in the mouse blastocyst, myotomes, and developing limb. *Dev Biol* 204:197–209.
31. Yuan H, Corbi N, Basilico C, Dailey L (1995) Developmental-specific activity of the FGF-4 enhancer requires the synergistic action of Sox2 and Oct-3. *Genes Dev* 9: 2635–2645.
32. Chew JL, et al. (2005) Reciprocal transcriptional regulation of Pou5f1 and Sox2 via the Oct4/Sox2 complex in embryonic stem cells. *Mol Cell Biol* 25:6031–6046.
33. Tomioka M, et al. (2002) Identification of Sox-2 regulatory region which is under the control of Oct-3/4-Sox-2 complex. *Nucleic Acids Res* 30:3202–3213.
34. Kuroda T, et al. (2005) Octamer and Sox elements are required for transcriptional cis regulation of Nanog gene expression. *Mol Cell Biol* 25:2475–2485.
35. Rodda DJ, et al. (2005) Transcriptional regulation of nanog by OCT4 and SOX2. *J Biol Chem* 280:24731–24737.
36. Yeom YI, et al. (1996) Germline regulatory element of Oct-4 specific for the totipotent cycle of embryonal cells. *Development* 122:881–894.
37. Kunarso G, et al. (2010) Transposable elements have rewired the core regulatory network of human embryonic stem cells. *Nat Genet* 42:631–634.
38. Xie D, et al. (2010) Rewirable gene regulatory networks in the preimplantation embryonic development of three mammalian species. *Genome Res* 20:804–815.
39. O'Farrell PH, Stumpff J, Su TT (2004) Embryonic cleavage cycles: how is a mouse like a fly? *Curr Biol* 14:R35–R45.
40. Kim JD, Faulk C, Kim J (2007) Retroposition and evolution of the DNA-binding motifs of YY1, YY2 and REX1. *Nucleic Acids Res* 35:3442–3452.
41. 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.
42. Hamburger V, Hamilton HL (1992) A series of normal stages in the development of the chick embryo. 1951. *Dev Dyn* 195:231–272.