



The endogenous retrovirus ENS-1 provides active binding sites for transcription factors in embryonic stem cells that specify extra embryonic tissue

A. Mey, Hervé Acloque, Emmanuelle Lerat, S. Gounel, V. Tribollet, S. Blanc,
D. Curton, A-M. Birot, A. Nieto, J. Samarut

► To cite this version:

A. Mey, Hervé Acloque, Emmanuelle Lerat, S. Gounel, V. Tribollet, et al.. The endogenous retrovirus ENS-1 provides active binding sites for transcription factors in embryonic stem cells that specify extra embryonic tissue. *Retrovirology*, 2012, 9 (1), pp.21. 10.1186/1742-4690-9-21 . hal-02289815

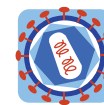
HAL Id: hal-02289815

<https://univ-lyon1.hal.science/hal-02289815>

Submitted on 29 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



RESEARCH

Open Access

The endogenous retrovirus ENS-1 provides active binding sites for transcription factors in embryonic stem cells that specify extra embryonic tissue

Anne Mey^{1*}, Hervé Acloque^{2,3}, Emmanuelle Lerat⁴, Sébastien Gounel¹, Violaine Tribollet¹, Sophie Blanc¹, Damien Curton¹, Anne-Marie Birot¹, M Angela Nieto² and Jacques Samarut^{1*}

Abstract

Background: Long terminal repeats (LTR) from endogenous retroviruses (ERV) are source of binding sites for transcription factors which affect the host regulatory networks in different cell types, including pluripotent cells. The embryonic epiblast is made of pluripotent cells that are subjected to opposite transcriptional regulatory networks to give rise to distinct embryonic and extraembryonic lineages. To assess the transcriptional contribution of ERV to early developmental processes, we have characterized *in vitro* and *in vivo* the regulation of ENS-1, a host adopted and developmentally regulated ERV that is expressed in chick embryonic stem cells.

Results: We show that *Ens-1* LTR activity is controlled by two transcriptional pathways that drive pluripotent cells to alternative developmental fates. Indeed, both Nanog that maintains pluripotency and Gata4 that induces differentiation toward extraembryonic endoderm independently activate the LTR. Ets coactivators are required to support Gata factors' activity thus preventing inappropriate activation before epigenetic silencing occurs during differentiation. Consistent with their expression patterns during chick embryonic development, Gata4, Nanog and Ets1 are recruited on the LTR in embryonic stem cells; in the epiblast the complementary expression of Nanog and Gata/Ets correlates with the *Ens-1* gene expression pattern; and Ens-1 transcripts are also detected in the hypoblast, an extraembryonic tissue expressing Gata4 and Ets2, but not Nanog. Accordingly, over expression of Gata4 in embryos induces an ectopic expression of *Ens-1*.

Conclusion: Our results show that *Ens-1* LTR have co-opted conditions required for the emergence of extraembryonic tissues from pluripotent epiblasts cells. By providing pluripotent cells with intact binding sites for Gata, Nanog, or both, *Ens-1* LTR may promote distinct transcriptional networks in embryonic stem cells subpopulations and prime the separation between embryonic and extraembryonic fates.

Background

Long terminal repeats (LTR) from endogenous retroviruses (ERV) are remnants of transposable elements disseminated in the genome that contain promoter activity [1] and can control nearby genes in different organisms [2-5]. They represent a source of binding sites for transcription factors [6], and some that are

active in embryonic stem (ES) cells have been shown to rewire the Nanog and Oct3/4 transcriptional networks in a species-specific manner [7]. Whether these changes are neutral or reflect species-specific adaptation to conserved developmental processes is not known, but ERV that escape silencing in pluripotent cells have been described in several species [4,8].

ES cells are isolated from the inner cell mass of very early embryos and can generate all the cells of an organism [9], a unique property called pluripotency that is supported by Oct3/4 [10], Sox2 [11] and Nanog [12] transcription factors. Oct3/4 and Nanog inhibit

* Correspondence: anne.mey@lyon.inra.fr; jacques.samarut@ens-lyon.fr

¹Institut de Génétique Fonctionnelle de Lyon, Université de Lyon, Université Lyon 1, CNRS, INRA, Ecole Normale Supérieure de Lyon, 46 allée d'Italie, 69364 Lyon Cedex 07, France

Full list of author information is available at the end of the article

differentiation toward embryonic and extraembryonic lineages, the latter providing nutrient exchange and inductive signals for the embryo [13]. These functions are well conserved in ES cells from different species, including chicken [14]. *In vivo*, the emergence of extraembryonic tissues from pluripotent cells represents the first cell fate decision and precedes the differentiation of the embryonic lineages. Notably in different species, Nanog deficiency makes the cells tolerant to differentiation into extraembryonic endoderm lineages [15-17] allowing the action of Gata-6 [18] and Gata-4 [19,20] transcription factors to drive extraembryonic endoderm formation. However, it is not clear what mechanisms guide pluripotent cells toward embryonic or extraembryonic lineages upon the suppression of the controls exerted by Oct3/4 [21] and Nanog [15].

To better understand the contribution of LTR to the transcriptional networks available in ES cells, we focused our interest on a developmentally regulated ERV and characterized its transcriptional regulation. The *Ens-1* LTR controls the expression of a multigenic family of genes of retroviral origin, *ENS* (Embryonic Normal Stem cell), present only in Galliform species. The *Ens-1* copy presents the most complete coding region and has been maintained in Galliform genomes through negative selection pressure [22] as observed for host-adopted retrotransposons [23]. *Ens-1* also called *Erni*, is expressed in pluripotent cells of the epiblast and later in the prospective neural plate [24,25], where it has been demonstrated to delay the expression of Sox2 [26] affecting the timing of emergence of the definitive neural plate and thus embryonic patterning. *In vitro*, *Ens-1* is expressed in chicken ES (cES) cells [25] and is repressed when ES cells differentiation is induced, mimicking the repression of the *Ens-1* LTR as further development occurs [27]. In addition to the coding regions, more than 800 copies of solo-LTR are disseminated and placed in close contact to host genes in sense or in anti-sense orientations [22] where they might act as alternative promoters [28]. We show here that the *Ens-1* LTR is under the control of both Nanog and Gata factors in such a way that may direct the formation of the extraembryonic endoderm when ES cells exit pluripotency.

Results

A cooperation between distinct DNA motifs controls the activity of the *Ens-1* promoter

The promoter activity of *Ens-1* in the chicken ES (cES) cells is supported by a 455 bp sequence (p455) upstream of the transcription initiation site and isolated from the U3 domain of the whole LTR [25]. The transcription factor CP2 that partially controls p455 activity binds to a domain located within the already explored 277 bp of the 5' end of this region [27]. Deletions within the first

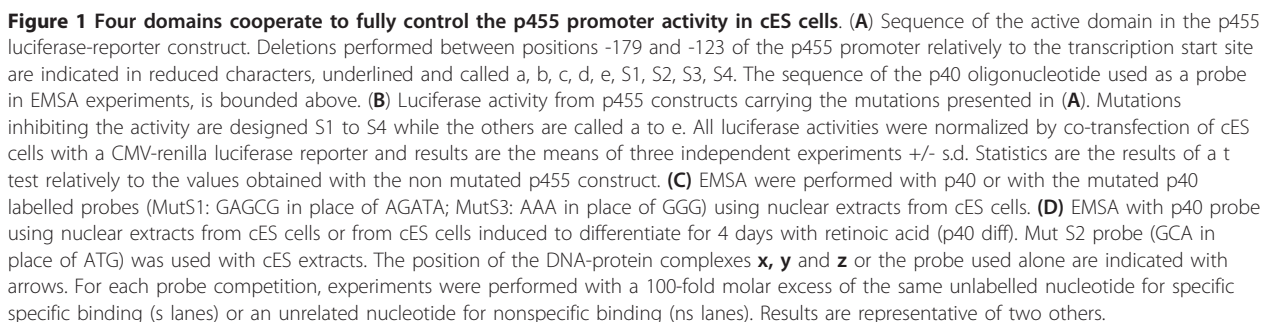
237 bp of p455 placed upstream of a firefly luciferase reporter gene have allowed us to find major determinants for the promoter activity between the positions -179 and -128 (Additional file 1: Figure S1). To reveal active binding sites for transcription factors, a series of site-directed deletions were performed in this region (Figure 1A). Four sites, S1 to S4, contributed to *Ens-1* promoter activity, and single deletions of each of them were not sufficient to abrogate the promoter activity to the same extent as the whole -179/-128 deletion (90% inhibition) (Figure 1B). No inhibition was obtained after site directed deletions downstream of position -128, including those in two putative [27] Gata binding sites (Additional file 1: Figure S1). These results suggest that several elements cooperate for the full promoter activity, as confirmed by the analysis of combined mutations of these four sites which inhibited the activity of the p455 construct with the same efficiency as a large deletion.

Site S2 was separated from site S3 by only three nucleotides suggesting a close contact between factors occupying both sites and a narrow sequence specificity of the inhibiting deletions when compared with other close mutations.

It thus seems that the *Ens-1* promoter is controlled by a combination of DNA binding proteins recognizing distinct and specific motifs and acting in a synergistic manner to promote the transcriptional activity.

The active domains are differently involved in the recruitment of ES specific protein complexes on the promoter

To address the role of the DNA active motifs in the recruitment of protein complexes, electrophoretic mobility shift assays (EMSA) were performed with nuclear extracts from cES cells. A ³²P end-labelled double stranded oligonucleotide, spanning sites S1 to S3 and called p40, was used as a probe and was compared with the same sequence carrying mutations on sites S1, S2 or S3 (Figure 1A). To compare probes of similar sizes, the mutations were made by substitution. The p40 probe formed three different protein-DNA complexes, **x**, **y** and **z** (Figure 1C), the latest being sometimes resumed to a single band as illustrated in Figure 1D. These complexes were competed off by a 100-fold molar excess of unlabelled probe, but only the complex **x** resisted to a 100-fold molar excess of an irrelevant unlabelled probe indicating specific binding only for **x** (Figure 1C). When using nuclear extracts from cES cells induced to differentiate for four days with retinoic acid, no complex was formed on the DNA probe (Figure 1D) indicating that the binding was specific for pluripotent stem cells. As shown in Figure 1C the mutation of the S1 site did not affect complex formation, while mutation of the S3 site inhibited the formation of the complex **x**. The double



In vivo, mutations in S1 and S3 of the full length promoter contained in the U3 domain from the *Ens-1* LTR [25] totally abrogated promoter activity (Figure 2), confirming the importance of these DNA sequences in the regulation of the whole promoter. Activity was also

The Match, Patch and AliBaba2 programs were used to help in the identification of transcription factors that may recognize the promoter active sites (see Figure 3A). The S1 site was recognized as a potential binding site

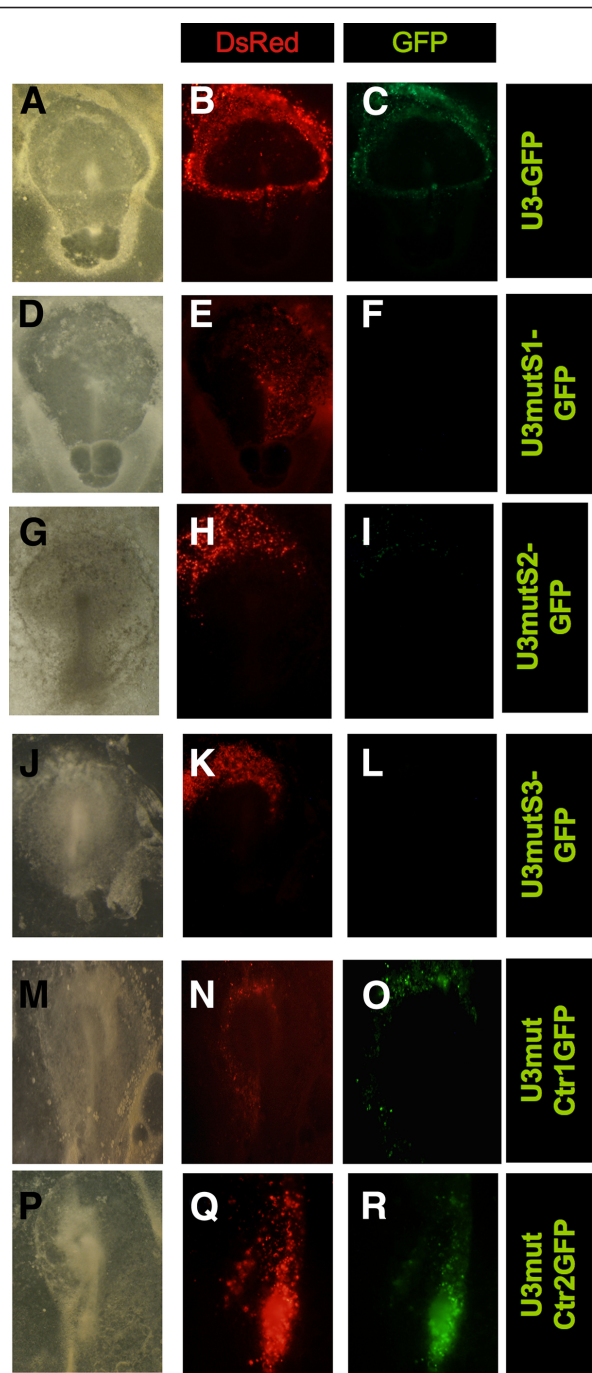


Figure 2 The activity of *Ens-1* promoter *in vivo* involves similar domains as in cES cells. Epiblast cells of pre-primitive streak chicken embryos were cotransfected by electroporation with CMV-DsRed (used to visualize the transfected cells) and U3-GFP vector, in which U3 contains the whole promoter of *Ens-1* (A-C). GFP expression was compared with U3 constructs presenting the MutS1 (D-F), the MutS2 (G-I), or the MutS3 (J-L) mutations (see Figure 1 legend). As negative controls, mutations on non-functional sites were performed upstream of site S1 (Mut Ctr1: deletion of site b, Figure 1A) (M-O) or at position -129 (Mut Ctr2: deletion of the GTGTG motif) (P-R). The same results were obtained in three independent experiments.

for the family of Gata transcription factors (A/T)GATA (A/G) [29] while the site S3 overlapped the core motif GGAA recognized by the Ets 1/2 transcription factors [30] and by the Churchill transcription factor (CGGGAA) expressed in the developing nervous system of chick [31] (Figure 3A). The complete motif for Nanog binding (TAATGG) [32] was also present in the *Ens-1* promoter and encompasses the S2 motif (Figure 3A). No homology with known transcription factor binding sites was found for S4. Alternatively, this motif may indirectly increase transcription by favouring the accessibility of flanking DNA to transcription factors as described for polydT tracts [33]. To identify which of these factors may promote the specific activity of the promoter in ES cells and in the early embryo, their expression patterns were checked by QPCR analysis (Figure 3B) and *in situ* hybridization (Figure 4). As expected *PouV*, the chicken homolog of *Oct4* [14], and *Nanog* were fully repressed in cES cells induced to differentiate by retinoic acid, while *Sox2* that is not expressed before the epiblast is fated to become neural plate in the chicken [34] was strongly induced, even more than *Churchill* (Figure 3B). With respect to the transcription of the six Gata family members, no variation was observed for *Gata1* and *Gata6* mRNA; *Gata3* mRNA was induced while *Gata2*, *Gata4* and *Gata5* were repressed during differentiation. Similarly, *Ets2* was reproducibly induced during the differentiation while *Ets1* was repressed (Figure 3B). Overall these results show that all the candidates for binding to the S1, S2 or S3 sites could be detected in cES cells, but the only ones that were reproducibly repressed during differentiation were *Nanog*, *Ets-1* and members of the Gata family, *Gata2*, *Gata4* and *Gata5*. *In situ* hybridization experiments in whole embryos (Figure 4) showed that all these factors, except *Gata5*, were detected in the epiblast at stage XII/XIII from which ES cells were derived. *Ets2* was also detected at this stage. At gastrulation, *Gata* and *Nanog* showed complementary expression patterns within the ectoderm (HH3, Figure 4), and their combined expression coincided with that of *Ens-1*. Similarly, *Ets1* and *Ets2* also showed complementary transcription patterns until stage HH3 (Figure 4).

To confirm the recruitment of these transcription factors to the promoter, we performed chromatin immunoprecipitation (ChIP) assays using cES cells expressing exogenous *Ets1*, *Gata4* or *Nanog* in fusion with a flag tag. Untagged factors and a flag-tagged version of the ubiquitous YY1 transcription factor were used as negative controls. Results shown in Figure 3C confirmed the specific recruitment of all the tagged proteins on the regulatory domain of the *Ens-1* promoter when compared with flanking regions. These interactions were significantly higher than those obtained with negative

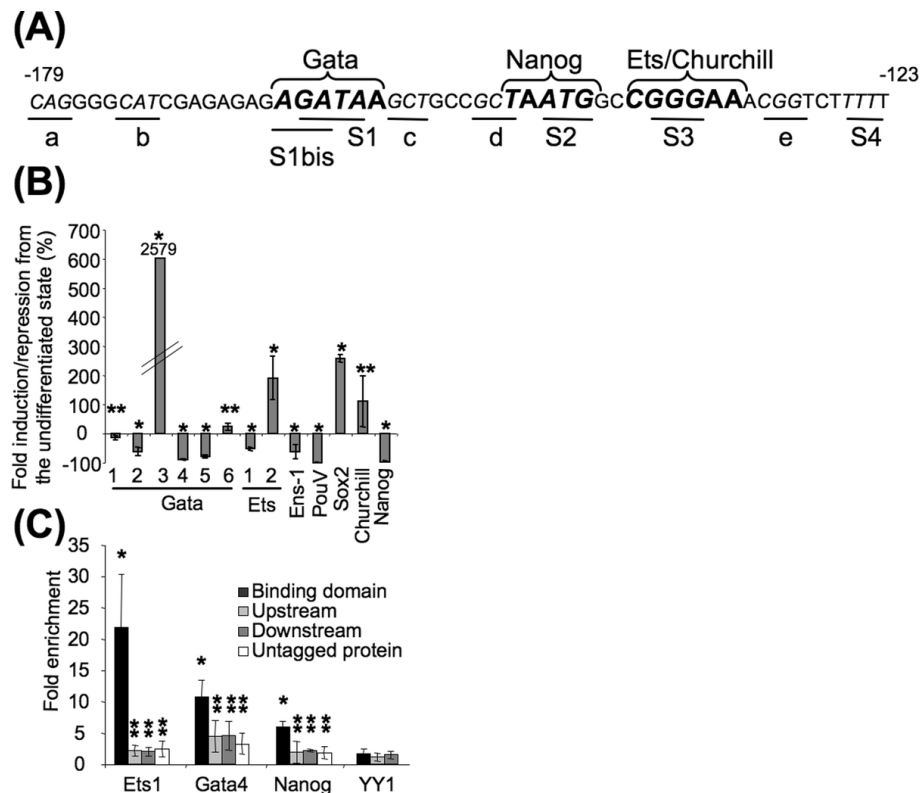


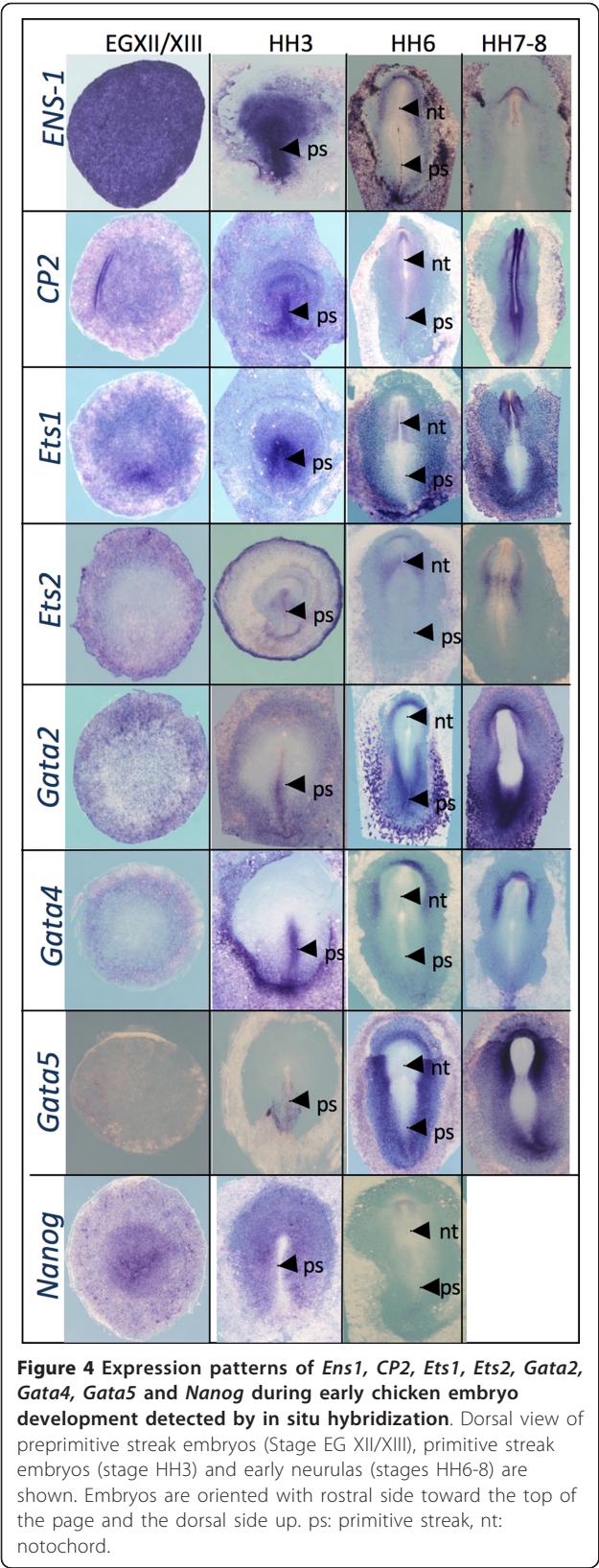
Figure 3 Nanog, Ets1 and Gata4 interact with the functional domain of the *Ens*LTR promoter in cES cells. (A) Binding sites for Gata, Nanog, Ets and Churchill transcription factors in the functional domain of the *Ens-1* LTR are represented in bold and upper characters. The mutations used in Figure 1 are indicated by underline and italic characters. **(B)** Real time PCR analysis performed on cDNA from cES induced to differentiate 48 h with retinoic acid. Expression levels of the indicated transcripts are represented as percentages of the levels obtained in undifferentiated cES. The numbers that are underlined identify the member of the Gata or of the Ets transcription factors family indicated below. Means are from three experiments \pm s.d. For Gata3 s.d. is \pm 1097. T test is relative to the values obtained in undifferentiated cells: * p < 0.05, ** p > 0.05. **(C)** Chromatin-immunoprecipitation of the *Ens-1* LTR promoter with anti-flag antibodies in cES transfected with expression vectors encoding for Nanog, Gata4 or Ets1 proteins in fusion with a flag tag. cES transfected with untagged transcription factors were used as negative controls. Results obtained with the irrelevant transcription factor YY1 in fusion with a flag tag are also represented. Results represent the fold enrichment compared to cells transfected with empty vector. The non-specific binding of the tagged transcription factors on upstream and on downstream regions is represented as well as the non specific interaction of the untagged transcription factors with the binding domain. Means of at least two experiments \pm s.d. are shown. T test is relative to the values obtained with the tagged YY1 transcription factor, * p < 0.05, ** p > 0.05.

controls. Altogether these data indicate that Nanog, Gata4, and Ets1 expressed in ES cells and repressed upon differentiation are recruited to the *Ens-1* promoter in pluripotent cells and are good candidates to support its activity.

Combinations with Ets are required for Gata or Nanog to support the full *Ens-1* promoter activity

The respective roles of Ets and Gata factors in the control of the *Ens-1* promoter were evaluated in differentiated cells which are devoid of Nanog (see Figure 3B) and in which the promoter activity is repressed [25,27]. Indeed, ES cells induced to differentiate with retinoic acid lost about 90% of *Ens-1* promoter activity, which was not restored by the ectopic expression of any of the

factors alone (Figure 5A). However combinations of Ets-1 or -2 with Gata4 induced promoter activity. The transcription factor Churchill, used instead of Ets due to its potential interaction with the same DNA motif, failed to synergize with the Gata factors (Figure 5A). Thus the combination of Ets and Gata4 is necessary and sufficient to support the *Ens-1* promoter activity in differentiated cells. Some redundancy was observed between the members of the same family independently of their regulation during differentiation (Additional file 2: Figure S2). The endogenous expression of the Ets and Gata family members in differentiated cells may explain the activation of the *Ens-1* promoter by Gata4, Ets1 and Ets2 transfected separately. However, the best results were obtained with Gata/Ets combinations, particularly involving either Ets1



or *Gata4* (Figure 3B) which are limiting in differentiated cells (see Figure 3B). Both *Ets1* and *Gata4* were also tested for their complementarity with *Nanog* (Figure 5B). *Nanog* alone strongly induced promoter activity in differentiated cells when compared with *Gata4* or *Ets1*. *Nanog* combined with *Ets1* reached levels of promoter activity that exceeded even those obtained in undifferentiated cES cells. In contrast no synergy was observed with *Gata4* suggesting that *Ets1* was independently required for both *Gata4* and *Nanog* activities (Figure 5B). When *Ets1* was mixed with both *Gata4* and *Nanog* (Figure 5C), the activity of the promoter was induced at higher levels than those observed with *Nanog*/*Ets1* but with a strong variability between experiments as also observed with *Gata4*/*Ets1* (Figure 5A). Interestingly, in the absence of *Ets1*, the addition of *Gata4* to *Nanog* was not a source of variability in the transcriptional activation of p455. In contrast, the addition of the CP2 transcription factor enhanced the response to *Nanog*/*Ets1* without increasing the variability, but was neutral when the promoter was placed under the control of *Gata4*/*Ets1* (Figure 5C). These data confirmed that CP2 is an enhancer of *Ens-1* promoter activity in ES cells [27]. To better characterize the additive effect of transcription factors, we compared the influence of equal quantities or a 4-fold excess of one of the two co-transfected factors. In both *Gata4*/*Ets1* and *Nanog*/*Ets1* combinations, the results were mainly influenced by a decrease in the *Ets1* level (Figure 5D) but with different outcomes. Limiting *Ets1* impaired the activity of the *Gata4*/*Ets1* dimer. This was not the case with *Nanog* as this combination still restored promoter activity but with a higher variability between experiments. Altogether these results indicate that both *Gata4* and *Nanog* are involved in promoter activity in ES cells and that both cooperate with *Ets1* in an independent manner.

To test whether the active motifs in the *Ens-1* promoter that have been characterized in ES cells were also used in differentiated cells, the combinations that fully restored the promoter activity were tested on p455 and on its mutated versions. As expected, the activation mediated by *Gata4*/*Ets1* or by *Nanog*/*Ets1* was fully impaired by mutations affecting either S1 + S3 or S2 + S3, respectively, leading to significantly lower luciferase activities than those obtained in undifferentiated cells with the wild type promoter. None of the single mutations abrogated the activity obtained with combinations of the transcription factors to the same extent. Indeed, with any of the transcription factors combinations tested, luciferase activities of p455 with single mutations were either not significantly different or higher than those observed in

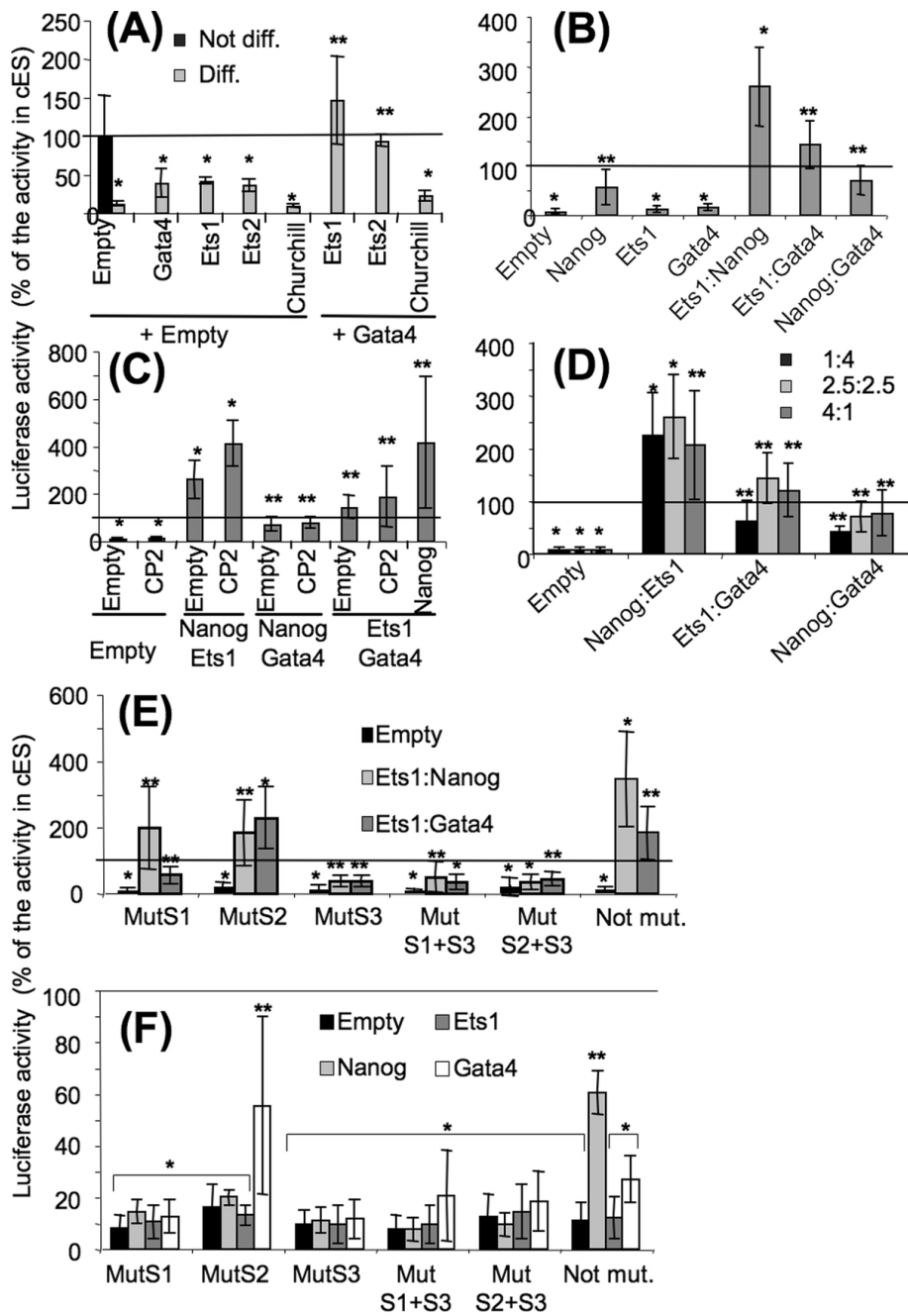


Figure 5 Ets factors synergize with Nanog and are indispensable for Gata4 to restore the activity of the p455 promoter in differentiated cells. (A, B, C, E, F) cES cells induced to differentiate 48 h with retinoic acid (Diff) were transfected with the p455-Luc vector and with pCi-neo vectors expressing the indicated transcription factors tested in combinations or alone. Equal quantities of each vector were transfected. (D) Co-transfections were performed as in (A) but with unequal quantities of each vector. The proportion of each is indicated as parts of five in the figure in the order indicated on the abscissa. (E) Combinations of transcription factors restoring the promoter activity in differentiated cells were tested on p455-Luc mutated on one or two of the activation sites. (F) Effect of mutations on the activity mediated by each transcription factors transfected alone in differentiated cells. All luciferase activities were measured 24 h after transfection and are represented as percentages of the value obtained with p455-Luc in cES cells co-transfected with empty vectors (Not Diff.). Results are the means of at least three independent experiments +/- s.d. In all the figures, t test values are relative to the results obtained in undifferentiated cES cells with p455-Luc: *p < 0.05, **p > 0.05. For (E) only, t test values relative to the p455-Luc activity in differentiated cells transfected with a given combination of transcription factors have also been calculated and are indicated in the text.

undifferentiated cells with p455 (Figure 5E). However, for a given combination of transcription factors, strong differences in the contribution of each binding site were revealed when compared to the activity on p455 in differentiated cells (Figure 5E).

As expected, the mutation in S3 (in an Ets binding site) significantly inhibited the activity mediated by Gata4/Ets1 or by Nanog/Ets1 on the wild type promoter ($p < 0.05$) while the S1 mutation (in a Gata binding site) only inhibited ($p < 0.05$) the activity of the Gata4/Ets1 combination (Figure 5E). In contrast, the S2 mutated version (in a Nanog binding site) did not show significant inhibition of the activity ($p > 0.05$) after ectopic expression of Nanog/Ets1 (Figure 5E) suggesting an indirect effect of Nanog in this combination of exogenous factors. To clarify this result, we evaluated the consequences of the mutations on promoter activity when expressing only Nanog.

Results shown in Figure 5F indicate that any of the single or double mutations significantly impaired the reactivation of the promoter mediated by Nanog, confirming the indirect effect and the requirement for intact S1, S2 and S3 binding sites. Results were different when expressing the other transcription factors. Although Gata4 alone failed to reactivate the p455 construct (Figure 5A and 5F), it reactivated the S2 mutant version to levels comparable to those obtained with the intact promoter in undifferentiated cells (Figure 5F). This was not the case with Nanog alone that restored the promoter activity on wild type p455 only. This indicates that the indirect effect of Nanog on the promoter activity required its interaction with the S2 site, probably releasing another DNA binding protein acting as a repressor. EMSA experiments in human HEK293 cells confirmed the direct interaction of the transfected chicken protein Gata4 with S1 site in non-ES cells (Additional file 3: Figure S3).

In contrast to Gata4, Ets1 was not able to increase the activity of a S2 mutant version of the promoter, indicating that the activity mediated by the combination of Nanog and Ets1 could not be only due to an indirect effect. The 2.5 (Figure 5B) to 4 (Figure 5E) fold increase in promoter activity in differentiated versus undifferentiated cells further illustrates the synergy between Nanog and Ets1.

Altogether these results reveal a functional interplay between Gata4, Nanog and Ets1 that requires intact S1, S2 and S3 binding sites, respectively. They also show the major role played by Ets1 to promote the activity mediated by Gata4 and by Nanog and also that Nanog partly acts in an indirect manner, probably by competition with another DNA binding protein acting as activity repressor on the S2 site.

Gata4 induces the ectopic expression of *Ens-1* in the developing embryo and is associated with the expression of *Ens-1* in extraembryonic tissues

Nanog and Gata6/Gata4 exert opposite influences on mammalian pluripotent cells toward the epiblast and the extraembryonic endoderm respectively [35]. In the chick embryo the hypoblast is equivalent to the mouse anterior visceral endoderm that derives from the extraembryonic endoderm [36]. The hypoblast expressed Gata4 and Ets2, but not Nanog (Figure. 6A). In agreement with our results in differentiated cells, *in situ* hybridizations in whole embryos revealed the expression of *Ens-1* in the hypoblast in addition to the embryonic epiblast (Figure 6B). Altogether these results are compatible with a role for Gata4 in the maintenance of the *Ens-1* LTR activity in extraembryonic tissues.

To confirm that Gata and Ets factors can support the activity of the *Ens-1* promoter *in vivo*, Gata4 and Ets1 were overexpressed by electroporation in stage HH3 chick embryos. Interestingly, Gata4 electroporation, but not Ets1, was sufficient to induce *Ens-1* ectopic expression (Figure 6C). Since Gata4 could not induce *Ens-1* promoter activity when used alone in differentiated cells, these results may be due to the endogenous expression of Ets1 and Ets2 at the electroporation site (see Figure 4) that can cooperate with Gata4. In contrast, Gata factors were not expressed at the electroporation site but rather overlapped with *Ens-1* expression (see Figure 4), discarding the possibility of testing the activity of Ets1 ectopically. These results confirmed that Gata4 is an activator of the *Ens-1* LTR *in vivo* acting independently from Nanog. The electroporation of CP2 poorly induced the ectopic expression of *Ens-1* confirming that its enhancer activity mainly depends of the control exerted by Nanog that was repressed at that stage (see Figure 4).

Surprisingly, overexpression of Gata/Ets failed to reactivate the expression of *Ens-1* in ES cells induced to differentiate *in vitro* with retinoic acid. To assess whether epigenetic regulations were responsible for this lack of response, differentiated cells were treated with drugs that change the chromatin status such as histone deacetylase inhibitors (valproic acid and trichostatin A) or a DNA methyl transferase inhibitor (5-aza-cytidine). These treatments were sufficient to fully reactivate and even reach higher levels of *Ens-1* transcripts when compared to the expression in ES cells (Figure 7A). These drugs did not restore the expression of *PouV* and *Nanog* excluding that our results were an indirect consequence of cell reprogramming. Histone deacetylase inhibitors did not restore the expression of *Gata4*, *Ets1*, or any of the Gata factors repressed during differentiation (Figure 7B) indicating that the transcription factors present in

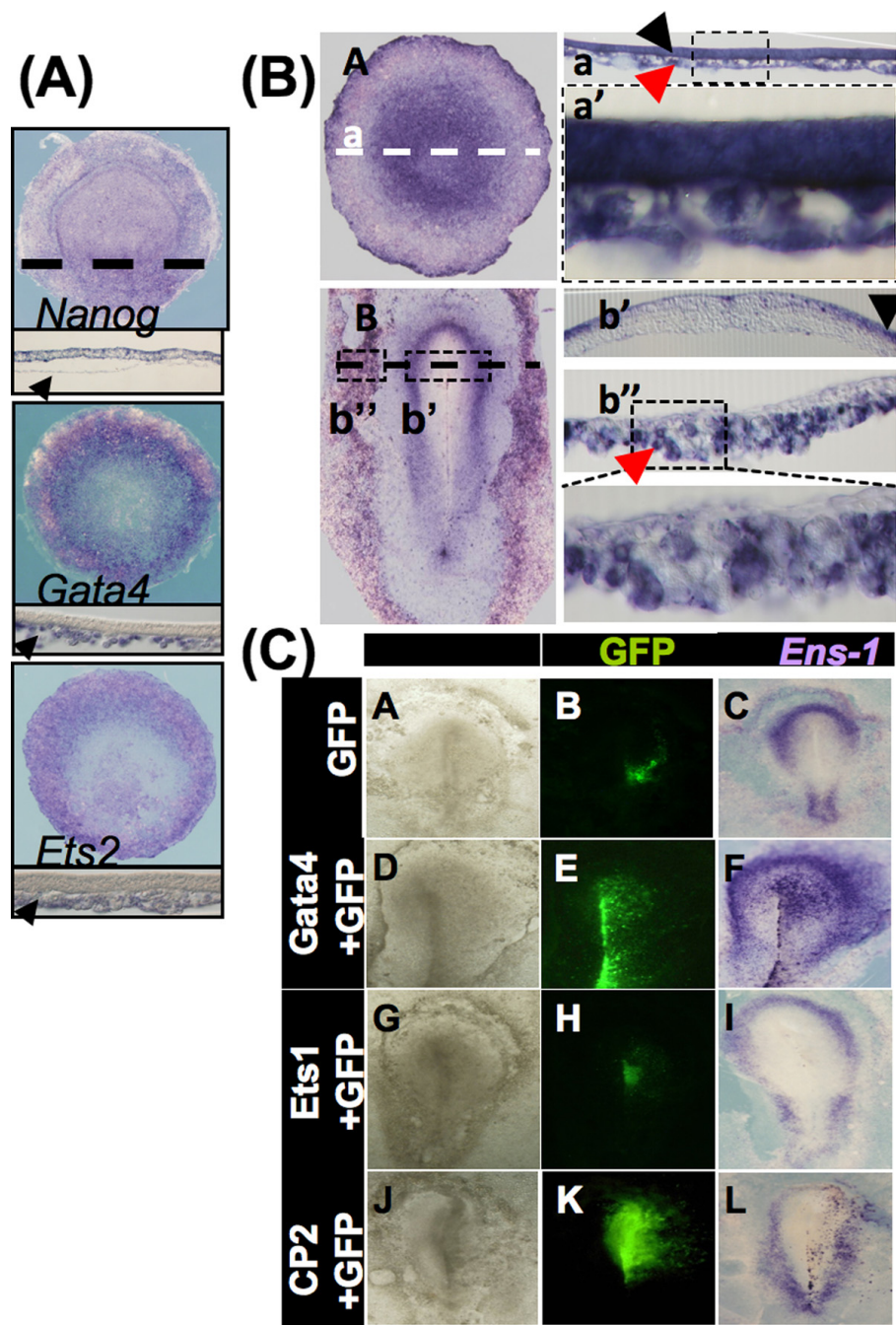
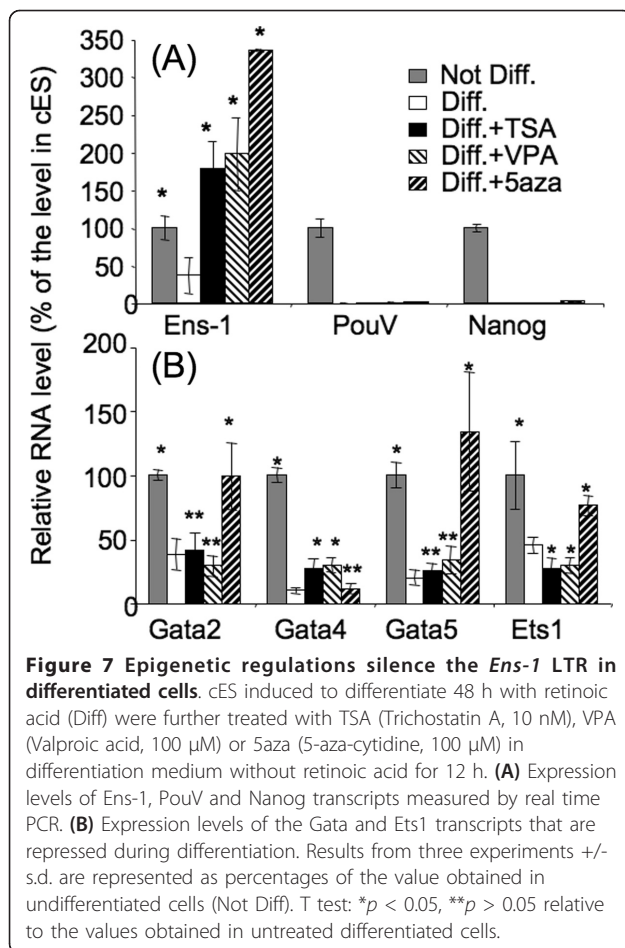


Figure 6 Gata4 induces the expression of *Ens-1* in vivo. (A) *Nanog*, *Gata4* and *Ets2* transcripts detected by in situ hybridization in stage XIII (EG) chick blastula. Sections highlight specific expression either in the epiblast or in the hypoblast (black arrow). (B) *In situ* hybridization in chick stage XIII (EG) blastula A and in stage V (H&H) gastrula B. Sections show *Ens-1* expression both in the epiblast (black arrow on a, enlargement on a') and in the hypoblast (red arrow in a, enlargement on a') at blastula stages, and later, at gastrulation, at the border of neural and non neural ectoderm (black arrow on b') and in the extraembryonic mesoderm (red arrow on b'', enlargement below). Dotted lines indicate sections' level. (C) Epiblasts of preprimitive streak chicken embryos were electroporated for the ectopic expression of GFP alone (A-C) or conjointly with the transcription factors Gata4 (D-F), Ets1 (G-I) and CP2 (J-L). *Ens-1* transcripts were detected ten hours later by *in situ* hybridization. Results obtained in 10 out of 19 electroporated embryos with Gata4 (F), in 4 out of 17 with CP2 (L), in 13 out of 13 with Ets1 (I) and in 12 out of 12 (C) with GFP alone.



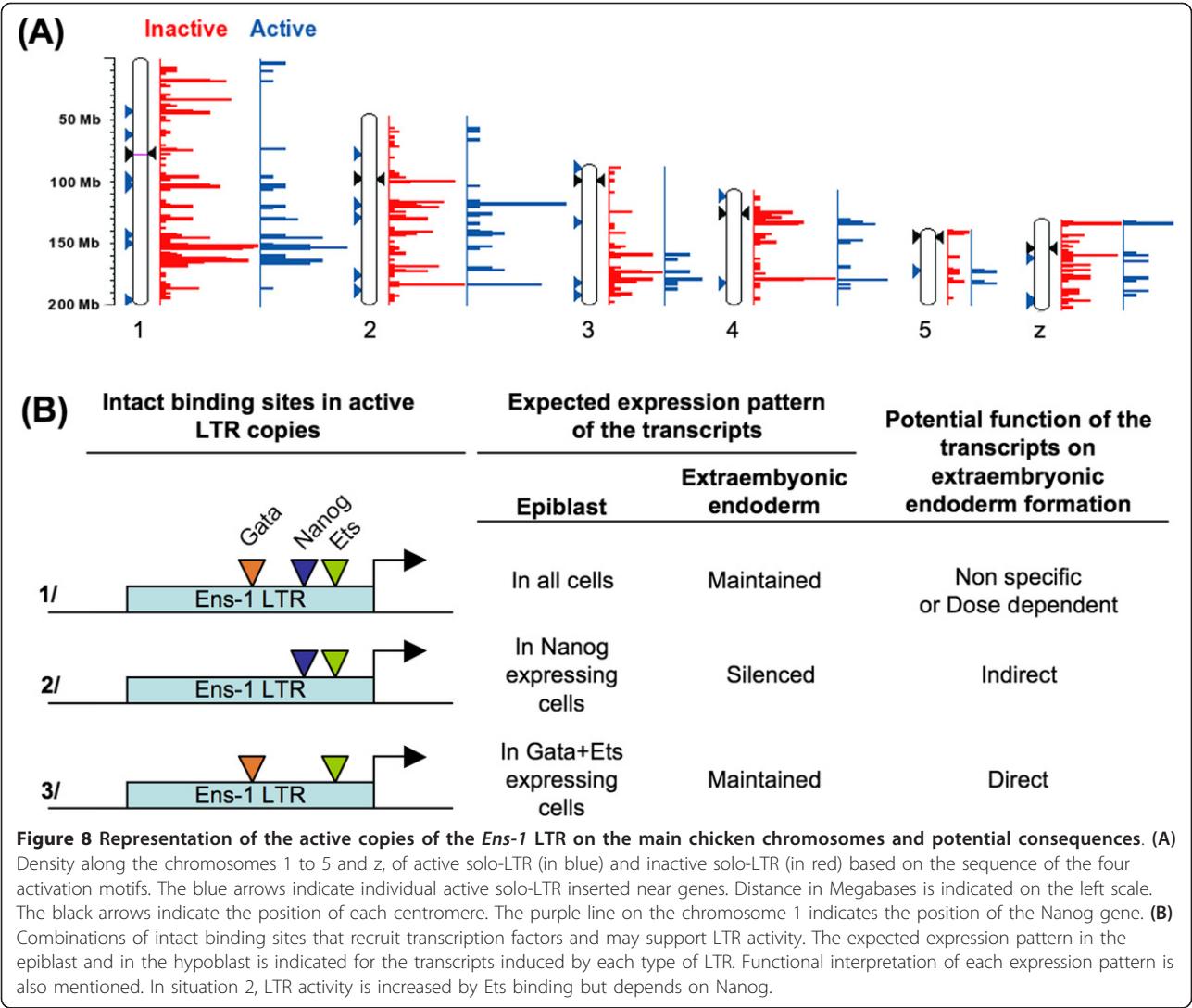
differentiated cells were sufficient to activate the promoter.

These results are in agreement with the existence of an additional level of regulation that involves epigenetic silencing, thereby restricting the accessibility of *Ens-1* LTR for the binding of transcription factors. This is likely to occur at developmental stages where *Ens-1* is no longer necessary and is in fact not expressed.

Distribution of the active copies of the *Ens-1* promoter in the chicken genome

The characterization of *Ens-1* LTR allowed us to perform a new analysis of their distribution on the latest version of the chicken genome considering only copies that contain the intact S1, S2, S3 and S4 sites required for full promoter activity. A total of 227 potentially active solo-LTR were revealed. Among them, 44 were located at less than 20 kb from genes or inside genes (list of 71 genes available upon request), and the rest were classified as intergenic. It is worth noting here that an intact TATA box-like sequence (GATAAAA) [27] was found in 200 out of the 227 (88%) potentially active solo-LTR and in 40 out of the 44 mentioned (91%) that

are located close to genes. Therefore most of the LTR with four intact activation domains may support direct transcription. Most of them (225 out of 227, 99%) also have an intact CP2 binding domain (CNRG-N6-CNRG) shown in a previous paper to support the enhancer activity of the *Ens-1* promoter [27]. This observation also supports a role for the active LTR in the transcriptional regulation of more distant sequences. The different insertions located on the main chromosomes (1,2,3,4,5 and z) were represented for active and inactive copies (solo-LTR with at least one mutation in the motifs S1 to S4) (Figure 8A). Both were located in the same clusters except for those in the regions surrounding the centromeres where we found the inactive but not the active solo-LTR. This was particularly true for chromosomes 1, 2, 3 and 5. These results suggest a non-random distribution of the active and inactive solo-LTR regarding heterochromatin enriched regions. Interestingly, on chromosome 1, the pluripotency gene *Nanog* was located near the centromere and at distance from active solo LTR. No insertion was found on chromosome 17, which contains the *PouV* gene that sustains cES pluripotency [14]. This suggests that pluripotency supporting genes are preserved from the influence of *Ens-1* LTR, in agreement with a putative role for the *Ens-1* LTR in the control of host genes exerted either in *cis* or in *trans*. To address whether the nearby genes belong to a particular functional category, gene ontology annotations of the genes were retrieved from the Ensembl database using the Biomart tool [37]. The most represented categories were genes whose products are associated with the membrane (about 25%) and the intermediate filaments components (approx. 12%) as previously observed with other adopted repetitive elements [38]. LTR with four intact motifs may support transcription in all the cells of the epiblast and later maintain expression in the hypoblast (Figure 8B situation 1/) as observed for *Ens-1* although at different levels (see Figure 6B panel A), perhaps reflecting differences in transcription factor concentrations. This heterogeneity in the epiblast may account for the emergence of cells forming the extraembryonic endoderm or may reflect the association with a maintained or a delayed function. Interestingly, some nearby genes of solo-LTR have already been involved in embryonic or extraembryonic development such as *Klf-6* [39], *teneurin-4* fragment [40] or the conserved microRNA *miR-7b* involved in the inhibition of *Fos* [41] that is required for extraembryonic endoderm epithelial organization [42]. Alternatively, LTR copies presenting only *Nanog* (Figure 8B situation 2/) or *Gata/Ets* (Figure 8B situation 3/) binding motifs would show a more restricted transcriptional pattern in the epiblast. These situations are compatible with the priming of



neighbouring genes for future silencing or with their subsequent role as cells exit pluripotency as summarised in Figure 8B presenting three potential functional situations. Importantly, they are all consistent with a role of *Ens-1* LTR in the first developmental decisions through the production of differentially regulated transcripts in pluripotent cells.

Discussion

This study addresses the question of whether the activity of an ERV LTR at very early developmental stages might be a source of targeted variability in the regulatory pathways operating in pluripotent cells. We show that the balance between the two opposite regulatory pathways controlling the decision between embryonic and extra-embryonic tissues is co-opted in the regulation of the *Ens-1* LTR promoter and correlates with its pattern of transcriptional activity.

We show that in ES cells, the *Ens-1* LTR is controlled by the pluripotency specific transcription factor Nanog as described for other ERV that escape silencing [7]. However, this LTR is also activated by a combination of Gata and Ets transcription factors, two families that are not restricted to pluripotent cells, but whose members were found here to be expressed in ES cells and in the epiblast in agreement with a previous report [43]. At early developmental stages, the expression pattern of Gata and Ets transcription factors is likely to promote that of the ERV gene *Ens-1*, even in the absence of Nanog, such as in the primitive streak. In agreement with this observation, electroporation of Gata4 induced ectopic expression of *Ens-1* in embryonic tissues that expressed Ets1 but not Nanog. In mammals, Gata4 over-expression is sufficient to transform ES cells into the extraembryonic endoderm lineage [20], and in agreement with this, we found expression of Gata4 in the

chick hypoblast along with that of *Ets-2*. We show here that *Ens-1* was also expressed in this extraembryonic tissue. In post-streak embryonic tissues, a good correlation was also observed between *Gata4* and *Ens-1* expression patterns until the formation of the neural plate (HH6), but *Ets* factors were not detected anymore suggesting the involvement of additional regulators. Indeed, in differentiated cells the activation domain attributed to *Nanog* was shown to play an indirect role in the activity mediated by *Gata4* in the absence of *Ets1*. At later developmental stages *Ens-1* is repressed in the whole embryo [24,25] despite the wider expression of *Gata* [44,45] and *Ets* [46] factors. In accordance with the silencing mechanisms reported for other ERV [4,8], epigenetic regulations are likely to restrict the promoter accessibility in irrelevant tissues as observed here in differentiated cells where the expression of *Ens-1* was strongly induced by treatment with epigenetic modifying drugs but not by the overexpression of *Gata/Ets*. Our results thus reveal that the *Gata/Ets* combination plays a role at developmental stages that correlates with the emergence of the hypoblast while distinct mechanisms are likely to occur in the neural plate [24,47] and later.

Several lines of evidence indicate that both *Nanog* and *Ets/Gata* regulations are active in ES cells. First, all were expressed in ES cells; secondly, we found that *Nanog*, *Gata4* and *Ets1* were recruited on the *Ens-1* promoter active domain in ES cells; thirdly, the suppression of the *Nanog* binding site only partially suppressed the promoter activity in ES cells and reciprocally with mutations in the *Gata* or in the *Ets* binding sites; fourthly, the activity mediated by *Nanog* is increased by *Ets-1*. This dual regulation is in agreement with the recent demonstration in the mouse that *Nanog* is required for the extraembryonic endoderm formation [48] that relies on the heterogeneity of ES cells populations. Accordingly distinct levels of the pluripotency marker *Nanog* are found among individual ES cells, *Nanog*-low cells expressing higher levels of extraembryonic endoderm markers [49] and being more prone to differentiate [50]. Similarly the epiblast that concentrates pluripotent cells, also contains cellular sub-populations expressing extraembryonic endoderm markers in different species [51,52]. These cells are likely precursors of the extraembryonic endoderm. Heterogeneity in the expression levels of *Nanog*, *Ets* and *Gata* factors was also observed here in stage XII/XIII chick epiblast where the periphery of the embryo that concentrates *Gata* factors expressed lower levels of *Nanog* transcripts. Despite this heterogeneity in the distribution of its regulating transcription factors, the expression level of *Ens-1* was maintained in the whole epiblast probably reflecting the ability of both regulation pathways to support the *Ens-1* LTR activity in cells with distinct fates. In embryonic tissues, *Ens-1*

regulates the timing of *Sox2* activation and thus the emergence of the definitive neural plate [26]. The present data support an earlier role of *Ens-1* or co-regulated sequences in the process that drives the formation of the hypoblast from pluripotent cells. Accordingly, *Ens-1* expression was found to be maintained in the hypoblast.

Experiments performed in differentiated cells revealed that *Nanog* as well as the *Gata4/Ets1* combination can restore the promoter activity normally observed in ES cells. However, additional regulations are likely to favor one or the other regulation pathway as illustrated by the transcription factor CP2 that is an enhancer of the *Ens-1* promoter activity in ES cells [27], but was shown here to solely promote the activation mediated by *Nanog*.

This balance between opposite active binding sites provided by the *Ens-1* LTR is thus coherent with the requirements supporting the emergence of the extraembryonic endoderm from pluripotent tissues and may contribute to this progress. Active copies of the *Ens-1* LTR may support in pluripotent cells the specific priming of genes involved later in the extraembryonic endoderm or in the neural plate formation. This is illustrated with *Ens-1* that is expressed in cES cells but involved later during the neural plate formation [26]. Alternatively the epigenetic silencing of the *Ens-1* LTR during differentiation may serve to repress irrelevant genes [53,54] according to the heterochromatin formation during ES cells differentiation [55].

Genome-wide studies combined with transcriptome analysis have concluded that LTR promoters may have an impact on tissue specific transcription [56,57]. We show here that only 26% (227 on 874 [22]) of the *Ens-1* LTR copies in the chicken genome contain an intact activation domain that supports transcriptional activity in the early embryo. Most of them are far from genes, and one third is localized in or at less than 20 kb of host genes, a distance that is compatible with a promoter activity and may direct gene functions in specific tissues. They may also act as enhancers inducing the transcription in both orientations [25] of non-coding RNA from intergenic loci [58]. Such sequences are known to be important players during development [59] and may be involved in the guidance of chromatin-modifying complexes on specific targets [60] as required during ES cells differentiation [61-63]. The presence of active LTR near genes already involved in embryonic or in extraembryonic development as listed here is in favour of relevant species-specific adaptations.

Conclusion

Our results demonstrate that the *Ens-1* LTR support gene expression in pluripotent and in extraembryonic tissues thus providing conditions for cell priming

compatible with the unclarified emergence of extraembryonic endoderm cells from ES cells. In addition to *Ens-1*, transcriptome analysis based on active LTR distribution along the genome will serve as a basis to explore their contribution in the regulation of other genes and their role in defining ES cells subpopulations with distinct cell fates.

Methods

Cell culture and DNA transfection

The culturing of cES cells and their differentiation by retinoic acid 10^{-6} M have been previously described [27]. cES cells or cES cells induced to differentiate 48 hours by retinoic acid were transfected using Lipofectamine 2000 reagent (Invitrogen) with p455-Firefly Luciferase reporter construct containing the *Ens-1* minimal promoter, the Renilla Luciferase reporter construct, pRL-CMV, to normalize for transfection efficiency and expression vectors for transcription factors used alone or in combinations. Except when mentioned, an equal quantity of each expression vector was used and total DNA quantity between conditions was maintained constant using empty vector. Twenty-four hours after transfection the Firefly and Renilla Luciferase luminescences were successively measured using Dual Luciferase Assay (Promega) as described by the manufacturer. Firefly reporter gene values were normalized to the activity of the Renilla luciferase. Chromatin modifying drugs Trichostatin A (TSA), Valproic acid (VPA) and 5-azacytidine (5-aza) were from Sigma.

HEK (Human Embryonic Kidney) 293 cells were from ATCC (CRL-1573) and cultivated in DMEM (Gibco) supplemented with 10% fetal bovine serum (Perbio) as recommended by the supplier.

DNA binding assays

The preparation of DNA binding assays and nuclear extracts was performed as previously described [27] using double-stranded DNA probes labelled with ^{32}P -ATP (Amersham). For competition experiments, a 100-fold molar excess of unlabelled double-stranded nucleotide, synthesized by Sigma, was incubated for 10 min with nuclear extract prior to the addition of the labeled double-stranded probe. For supershift experiments in HEK293 cells anti-Flag M2 antibody was from Sigma. Whole IgG purified from mice were from Zymed.

DNA constructs and site directed mutagenesis

The luciferase reporter construct p455-Luc was done using the pGL2 basic vector (Promega) as previously described and includes the LTR sequence from -455 to +83 of the transcription start site [27]. Targeted mutagenesis by deletions was performed using the Quick Change Mutagenesis Kit (Stratagene) following the

manufacturer's instructions. The U3-GFP construct expresses a GFP reporter gene placed under the control of the sequence from -738 to +83 of the transcription start site described before [25]. This largest construct contains the 738 bp of the promoter defined previously [27] which is present in the U3 region of the LTR of *Ens* genes [25]. Both p455-Luc and U3-GFP contain 83 bp from the R region 5' end starting downstream the transcription start site [25].

Site directed mutations were performed by replacements or by deletions of bases as indicated in the figure legends. Large deletions in the constructs p455 del 128-87, del 128-57 del 179-128, del 179-87, del 179-57 and del 237-31 were obtained by ligation of two PCR amplified fragments surrounding the deletion. The transcription factors were cloned in a pCi-Neo expression vector (Promega) modified to introduce a flag tag at the N-terminal part of the protein. Ets-2 (Genbank:X07202), Gata3 (Genbank:XM_417294), Gata4 (Genbank:XM_420041) and Gata5 (Genbank:NM_205421) were amplified from chicken ES cells cDNA. Chicken Ets-1 (p54, Genbank:X13026) was given by Dr B. Wasylyk [64], Churchill (Genbank:AF238863) was a gift from Dr C. Stern [31]. For over-expression experiments in the embryo, transcription factor coding sequences were placed under the control of the CAG promoter (a chicken b-actin promoter combined with CMV enhancer) to ensure strong expression [65]. The predictions for transcription factor binding sites were carried out using the indicated programs available online (<http://www.gene-regulation.com/pub/programs.html>).

Real time PCR

RNA was extracted using an RNeasy kit with on-column DNase digestion (Qiagen). Reverse transcription was carried out with 1 μg of RNA and SuperScript III (Invitrogen). Real-Time PCR was performed using the MXP-300P PCR-system (Stratagene), Mix-Quantitect SYBR Green (Qiagen) as reagent under the following cycling conditions: 40 cycles at 95°C for 30 seconds, at 55°C for 1 minute and 72°C for 30 seconds. Samples were run in duplicate and gene expression levels were calculated using Delta Ct (<http://www.gene-quantification.info/>) normalized with the chicken 40S ribosomal protein S17 as house-keeping gene. The primers used are listed in Table 1.

Chromatin immunoprecipitation

ES cells were transfected using Lipofectamine 2000 (Invitrogen) with pCi vectors encoding for the indicated transcription factor in fusion with a Flag tag. The following day, cells were fixed with formaldehyde, collected and lysed in 50 mM Tris-HCl (pH 8.1), 10 mM EDTA, 0.5 mM EGTA, 1% (w/v) SDS, and proteases inhibitor for 5 minutes on ice. The cells were then sonicated to

Table 1 Oligonucleotides used for gene expression analysis

Gene	Sense (5'-3')	Anti-sense (5'-3')
RS 17*	ACACCCGTCTGGGCA ACGACT	CCCGCTGGATGCGCTTCATCA
Gata1	CACTGCACTCCGACATCCA	GTACCAAGATCCCACAGTCCTT
Gata5	CTTCCATTACGAAGTACAGACAGCAC	GGACACCGACACAATGCCTTG
Gata6	GAATTCAGACGAGGAAACGAAAACC	ACGTAGATGTTGGAGTCATAGGAAC
Ets-1	CCAGCTTCATCACAGAGTCTACC	AGGGATAGTCGTTCTCGTACTTGAG
Ets-2	CAGAGGAATGCTCAAGCGGC	GCACTTCCTGGAGCGTTTGA
Churchill	ATCATCACCTACGACCACTG	CAGGGTTACAACTGCCTTCA
Sox2*	GCAGAGAAAAGGGAAAAAGGA	TTTCCTAGGGAGGGGTATGAA
Nanog*	CAGCAGACCTCTCCTTGACC	TTCCTTGCCCACTCTCACC
PouV*	GTTGTCCGGGTCTGGTCT	GTGGAAAGGTGGCATGTAGAC
Ens-1	CACCACTCAGGACCCAAAGT	GGGGATGAAACCTTTTGGT

Primers for *gallus* Gata2, Gata3 and Gata4 were purchased from Qiagen

* Primers previously described [64]

shear chromatin to a final average size of 200-600 base pairs. Sonicated chromatin was diluted 1/10 with the following buffer: 20 mM Tris-HCl (pH 8.1), 1% (v/v) Triton X-100, 150 mM NaCl, 1 mM EDTA and proteases inhibitors. Immunoprecipitation was carried out overnight at 4°C with 50 ml of agarose beads coated with anti-Flag M2 antibody (Sigma). Beads were washed extensively, and bound material was eluted by two incubation rounds of 15 minutes in 1% (w/v) SDS and 0.1 M NaHCO₃ at room temperature. Cross-linking was reversed by incubation 4 h at 65°C in 200 mM NaCl and 100 mg/ml proteinase K. The DNA was purified using Mini-Elute columns (Qiagen). Aliquots were used for quantitative real-time PCR as described above. The oligonucleotides used were designed either inside the p455 region (GAGGAACAAGTCCAGGCAAG; GATGGCCATTTTCCTTGAGA), or 1000 bp upstream (CCCACGGTACACAATGAACA; GCTAGG-GAGCCCTTTAACCA), or 1000 bp downstream (TGGTGTGGTGTGTTGCAGTTT; CCCTTTGTTGAG-GAAAGCAC) from the *Ens-1* copy on chromosome 5. The region bounded by the primers designed inside the p455 sequence is located between positions -278 and -93 from the transcription start site.

Embryos

Fertilised chicken eggs were purchased from Granja Santa Isabel, Cordoba, Spain. Eggs were incubated, opened and staged according to Eyal-Giladi and Kochav [66] for the pre-primitive streak stages and Hamburger and Hamilton [67] for subsequent stages. They were dissected and fixed overnight in 4% paraformaldehyde in phosphate buffered saline (PBS) at 4°C.

Chicken embryos electroporation

Stage 2-4 [67] chick embryos were explanted, washed in PBS, and placed upside down over an electroporation

chamber (NEPAGEN) containing a platinum electrode connected to the negative pole. After visualizing the embryo, a solution containing expression plasmids (2 mg/ml in PBS with 0.1% Fastgreen and 6% sucrose) was injected between the vitelline membrane and the epiblast. An anodal electrode was placed over the hypoblast to cover the injected area and contact was made with PBS. A train of electric pulses (5 pulses, 4 Volts, 50 ms, 0.5 Hz) were applied using an Intracell TSS10 pulse stimulator (Intracell). The embryos were then placed in culture as described [68] and allowed to develop until they reach the required stage. Embryos were then photographed with a Leica MZFLIII dissecting microscope to record GFP or DsRed expression and fixed overnight in 4% paraformaldehyde (PFA) in PBS at 4°C for subsequent processing.

Whole mount in situ hybridization

Whole-mount in situ hybridization was carried out in chick embryos at various stages of development as previously described [69]. Digoxigenin-labelled probes for ENS1 [25], CP2 [27], Ets1, Ets2, Gata2, Gata4, Gata5 and Nanog with digoxigenin-UTP (Roche) were synthesized. After hybridization, embryos were fixed in 4% paraformaldehyde in PBS, washed in PBS, and photographed in whole-mount under a Leica M10 dissecting scope. Subsequently, some embryos were embedded in gelatine and sectioned in a vibratome at 40 µm. These slices were photographed using an Olympus DP70 digital camera mounted on a Leica DMR microscope with Nomarski optics.

Detection and localization of solo-LTR in the chicken genome

Using the sequence of the *Ens-1* LTR as a reference, the occurrence of solo-LTR were searched in the latest version of the chicken genome (release galGal3) that was

downloaded from the UCSC website (<http://hgdownload.cse.ucsc.edu/downloads.html>) using blastn [70]. We classified each solo-LTR as active or inactive according to the presence or absence of the motifs identified as being essential to its activity by using the program fuzz-nuc from the EMBOSS package [71]. We thus obtained 227 active solo-LTR and 916 inactive solo-LTR. Using the Ensembl facilities [72] (<http://www.ensembl.org/index.html>), we mapped the distribution of each type of solo-LTR on the chromosomes. By combining the positions of the active solo-LTR and those of the genes, we determined the active solo-LTR inserted close to genes (at less than 20 kb or inside the genes). Genes ontology annotations were retrieved from the Ensembl database using the Biomart tool [37] (<http://www.ensembl.org/index.html>).

Additional material

Additional file 1: Figure S1. The -179 to -128 bp residue upstream of the transcription initiation site is required for the p455 promoter activity.

(A) Schematic representation of the wild-type p455 reporter construct used in transfection experiments and position of the different deletion edges of the constructs used in (B). (B) cES cells were transfected with wild type or with one of the deleted p455 (p455 Del) luciferase reporters illustrated in (A). All luciferase activities were normalized by co-transfection with a CMV-renilla luciferase reporter. (C) The contribution of two putative Gata binding sites to the promoter activity of p455 was examined. Site directed deletions were performed in one of the following positions: Gata n°2: TATC -111/-114 or Gata n°3: TATC +47/+50. The deletion in site S1 described in Figure 1 and showing inhibition is used as reference. Luciferase activities obtained with these constructs in cES cells were compared to that obtained with wild type p455 as indicated in (B). Means are \pm s.d. of at least three independent experiments. Statistics are from t tests relative to the value obtained with p455.

Additional file 2: Figure S2. Redundancy between Gata factors to restore the activity of the p455 promoter in differentiated cells.

Experiments were performed as indicated in Figure 4 with equal quantities of pCi-neo vectors expressing the indicated transcription factors transfected in cES cells induced to differentiate 48 h with retinoic acid (Diff). Results are percentages of the value obtained with p455-Luc in cES cells transfected with empty vector. All the results are the means of three independent experiments \pm s.d. T test: * $p < 0.05$, ** $p > 0.05$, relative to the values obtained in undifferentiated cells (Not diff.).

Additional file 3: Figure S3. Direct interaction of Gata4 with the site S1 of the promoter sequence.

HEK293 cells were transfected with expression vectors encoding for the Gata4 protein in fusion with a flag tag or untagged as control. Nuclear extracts were used for EMSA assays with the labelled p40 probe or with p40 probes mutated in sites S1, S2 or S3. Supershifts were performed using an anti-Flag antibody or whole IgG as control; both used at 1 μ g per lane. On the right is represented the result obtained with untransfected HEK293 cells. This lane is from the same gel but moved from the opposite side. Results are from one experiment representative of two.

Acknowledgements

We thank Dr. B. Wasylyk and Dr. C. Stern for Ets1 and Churchill expression vectors respectively. We thank Dr. F. Laval and Dr. B. Pain for qPCR primers sequences. This work was supported by grants from the ANR (Agence Nationale pour la Recherche) Genanimal (ANR-06-GANI-007-01) and the ANR Blanc (ANR-09-BLAN-0141) to JS, grants from INRA to AM and grants from

the Spanish Ministry of Science and Innovation (BFU2008-01042, CONSOLIDER-INGENIO 2010 CSD2007-00017 and CSD2007-00023) to MAN.

Author details

¹Institut de Génomique Fonctionnelle de Lyon, Université de Lyon, Université Lyon 1, CNRS, INRA, Ecole Normale Supérieure de Lyon, 46 allée d'Italie, 69364 Lyon Cedex 07, France. ²Instituto de Neurociencias de Alicante, CSIC-UMH, Avda. Ramón y Cajal s/n, 03550 Sant Joan d'Alacant, Spain. ³Laboratoire de Genetique Cellulaire-INRA, ENVT, Chemin de Borde Rouge BP52627, 31326 Castanet Tolosan, France. ⁴Université de Lyon, Lyon, F-69003, France; université Lyon 1, Villeurbanne, F-69622 cedex, France; CNRS, UMR5558, Laboratoire de Biométrie et Biologie Evolutive, 43 bd du 11 Novembre 1918, Villeurbanne F-69622 cedex, France.

Authors' contributions

Conceived and designed the experiments: AM. Performed the experiments: SG, VT, SB, DC, AMB, HA, EL. AM. Analyzed the data: AM, HA, EL. Wrote the manuscript: AM. Revised the manuscript: HA, EL, JS, MAN. Gave final approval of the version to be published: AM, JS, MAN.

Competing interests

The authors declare that they have no competing interests.

Received: 21 July 2011 Accepted: 15 March 2012

Published: 15 March 2012

References

1. Brosius J: Genomes were forged by massive bombardments with retroelements and retrosequences. *Genetica* 1999, **107**:209-238.
2. Medstrand P, van de Lagemaat LN, Dunn CA, Landry JR, Svenback D, Mager DL: Impact of transposable elements on the evolution of mammalian gene regulation. *Cytogenet Genome Res* 2005, **110**:342-352.
3. van de Lagemaat LN, Landry JR, Mager DL, Medstrand P: Transposable elements in mammals promote regulatory variation and diversification of genes with specialized functions. *Trends Genet* 2003, **19**:530-536.
4. Peaston AE, Evisikov AV, Graber JH, de Vries WN, Holbrook AE, Solter D, Knowles BB: Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. *Dev Cell* 2004, **7**:597-606.
5. Jordan IK, Rogozin IB, Glazko GV, Koonin EV: Origin of a substantial fraction of human regulatory sequences from transposable elements. *Trends Genet* 2003, **19**:68-72.
6. Bourque G, Leong B, Vega VB, Chen X, Lee YL, Srinivasan KG, Chew J-L, Ruan Y, Wei C-L, Ng HH, Liu ET: Evolution of the mammalian transcription factor binding repertoire via transposable elements. *Genome Res* 2008, **18**:1752-1762.
7. Kunarso G, Chia N-Y, Jeyakani J, Hwang C, Lu X, Chan Y-S, Ng H-H, Bourque G: Transposable elements have rewired the core regulatory network of human embryonic stem cells. *Nat Genet* 2010, **42**:631-634.
8. Rowe HM, Trono D: Dynamic control of endogenous retroviruses during development. *Virology* 2011, **411**:273-287.
9. O'Shea KS: Self-renewal vs. differentiation of mouse embryonic stem cells. *Biol Reprod* 2004, **71**:1755-1765.
10. Nichols J, Zevnik B, Anastassiadis K, Niwa H, Klewe-Nebenius D, Chambers I, Scholer H, Smith A: Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell* 1998, **95**:379-391.
11. Avilion AA, Nicolis SK, Pevny LH, Perez L, Vivian N, Lovell-Badge R: Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes Dev* 2003, **17**:126-140.
12. Chambers I, Colby D, Robertson M, Nichols J, Lee S, Tweedie S, Smith A: Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell* 2003, **113**:643-655.
13. Arnold SJ, Robertson EJ: Making a commitment: cell lineage allocation and axis patterning in the early mouse embryo. *Nat Rev Mol Cell Biol* 2009, **10**:91-103.
14. Laval F, Acloque H, Bertocchini F, Macleod DJ, Boast S, Bachelard E, Montillet G, Thenot S, Sang HM, Stern CD, et al: The Oct4 homologue PouV and Nanog regulate pluripotency in chicken embryonic stem cells. *Development* 2007, **134**:3549-3563.
15. Mitsui K, Tokuzawa Y, Itoh H, Segawa K, Murakami M, Takahashi K, Maruyama M, Maeda M, Yamanaka S: The homeoprotein Nanog is

- required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell* 2003, **113**:631-642.
16. Hyslop L, Stojkovic M, Armstrong L, Walter T, Stojkovic P, Przyborski S, Herbert M, Murdoch A, Strachan T, Lako M: **Downregulation of NANOG Induces Differentiation of Human Embryonic Stem Cells to Extraembryonic Lineages.** *Stem Cells* 2005, **23**:1035-1043.
17. Hough SR, Clements I, Welch PJ, Wiederholt KA: **Differentiation of Mouse Embryonic Stem Cells after RNA Interference-Mediated Silencing of OCT4 and Nanog.** *Stem Cells* 2006, **24**:1467-1475.
18. Koutsourakis M, Langeveld A, Patient R, Beddington R, Grosveld F: **The transcription factor GATA6 is essential for early extraembryonic development.** *Development* 1999, **126**:723-732.
19. Soudais C, Bielinska M, Heikinhoimo M, MacArthur CA, Narita N, Saffitz JE, Simon MC, Leiden JM, Wilson DB: **Targeted mutagenesis of the transcription factor GATA-4 gene in mouse embryonic stem cells disrupts visceral endoderm differentiation in vitro.** *Development* 1995, **121**:3877-3888.
20. Fujikura J, Yamato E, Yonemura S, Hosoda K, Masui S, Nakao K, Miyazaki J-i, Niwa H: **Differentiation of embryonic stem cells is induced by GATA factors.** *Genes Dev* 2002, **16**:784-789.
21. Niwa H, Miyazaki J, Smith AG: **Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells.** *Nat Genet* 2000, **24**:372-376.
22. Lerat E, Birot AM, Samarut J, Mey A: **Maintenance in the chicken genome of the retroviral-like cENS gene family specifically expressed in early embryos.** *J Mol Evol* 2007, **65**:215-227.
23. Sekita Y, Wagatsuma H, Nakamura K, Ono R, Kagami M, Wakisaka N, Hino T, Suzuki-Migishima R, Kohda T, Ogura A, et al: **Role of retrotransposon-derived imprinted gene, Rtl1, in the feto-maternal interface of mouse placenta.** *Nat Genet* 2008, **40**:243-248.
24. Streit A, Berliner AJ, Papanayotou C, Sirulnik A, Stern CD: **Initiation of neural induction by FGF signalling before gastrulation.** *Nature* 2000, **406**:74-78.
25. Acloque H, Risson V, Birot AM, Kunita R, Pain B, Samarut J: **Identification of a new gene family specifically expressed in chicken embryonic stem cells and early embryo.** *Mech Dev* 2001, **103**:79-91.
26. Papanayotou C, Mey A, Birot AM, Saka Y, Boast S, Smith JC, Samarut J, Stern CD: **A mechanism regulating the onset of Sox2 expression in the embryonic neural plate.** *PLoS Biol* 2008, **6**:e2.
27. Acloque H, Mey A, Birot AM, Gruffat H, Pain B, Samarut J: **Transcription factor cCP2 controls gene expression in chicken embryonic stem cells.** *Nucleic Acids Res* 2004, **32**:2259-2271.
28. Cohen CJ, Lock WM, Mager DL: **Endogenous retroviral LTRs as promoters for human genes: A critical assessment.** *Gene* 2009, **448**:105-114.
29. Merika M, Orkin SH: **DNA-binding specificity of GATA family transcription factors.** *Mol Cell Biol* 1993, **13**:3999-4010.
30. Fisher RJ, Mavrothalassitis G, Kondoh A, Papas TS: **High-affinity DNA-protein interactions of the cellular ETS1 protein: the determination of the ETS binding motif.** *Oncogene* 1991, **6**:2249-2254.
31. Sheng G, dos Reis M, Stern CD: **Churchill, a zinc finger transcriptional activator, regulates the transition between gastrulation and neurulation.** *Cell* 2003, **115**:603-613.
32. Jauch R, Ng CKL, Saikatendu KS, Stevens RC, Kolatkar PR: **Crystal Structure and DNA Binding of the Homeodomain of the Stem Cell Transcription Factor Nanog.** *J Mol Biol* 2008, **376**:758-770.
33. Segal E, Widom J: **Poly(dA:dT) tracts: major determinants of nucleosome organization.** *Curr Opin Struct Biol* 2009, **19**:65-71.
34. Rex M, Orme A, Uwanogho D, Tointon K, Wigmore PM, Sharpe PT, Scotting PJ: **Dynamic expression of chicken Sox2 and Sox3 genes in ectoderm induced to form neural tissue.** *Dev Dyn* 1997, **209**:323-332.
35. Chazaud C, Yamanaka Y, Pawson T, Rossant J: **Early Lineage Segregation between Epiblast and Primitive Endoderm in Mouse Blastocysts through the Grb2-MAPK Pathway.** *Dev Cell* 2006, **10**:615-624.
36. Bertocchini F, Stern CD: **The Hypoblast of the Chick Embryo Positions the Primitive Streak by Antagonizing Nodal Signaling.** *Dev Cell* 2002, **3**:735-744.
37. Haider S, Ballester B, Smedley D, Zhang J, Rice P, Kasprzyk A: **BioMart Central Portal, unified access to biological data.** *Nucleic Acids Res* 2009, **37**:W23-W27.
38. Lowe CB, Bejerano G, Haussler D: **Thousands of human mobile element fragments undergo strong purifying selection near developmental genes.** *Proc Natl Acad Sci* 2007, **104**:8005-8010.
39. Blanchon L, Bocco JL, Gallot D, Gachon A-M, L'Emery D, D'Échelotte P, Dastugue B, Sapin V: **Co-localization of KLF6 and KLF4 with pregnancy-specific glycoproteins during human placenta development.** *Mech Dev* 2001, **105**:185-189.
40. Kenzelmann-Broz D, Tucker RP, Leachman NT, Chiquet-Ehrismann R: **The expression of teneurin-4 in the avian embryo: potential roles in patterning of the limb and nervous system.** *Int J Dev Biol* 2010, **54**:1509-1516.
41. Lee H-J, Palkovits Ms, Young WS: **miR-7b, a microRNA up-regulated in the hypothalamus after chronic hyperosmolar stimulation, inhibits Fos translation.** *Proc Natl Acad Sci* 2006, **103**:15669-15674.
42. Yang D-H, Smith ER, Cai KQ, Xu X-X: **c-fos elimination compensates for disabled-2 requirement in mouse extraembryonic endoderm development.** *Dev Dyn* 2009, **238**:514-523.
43. Chapman S, Matsumoto K, Cai Q, Schoenwolf G: **Specification of germ layer identity in the chick gastrula.** *BMC Dev Biol* 2007, **7**:91-107.
44. Laverriere AC, MacNeill C, Mueller C, Poelmann RE, Burch JB, Evans T: **GATA-4/5/6, a subfamily of three transcription factors transcribed in developing heart and gut.** *J Biol Chem* 1994, **269**:23177-23184.
45. Minko K, Bollerot K, Drevon C, Hallais MF, Jaffredo T: **From mesoderm to blood islands: patterns of key molecules during yolk sac erythropoiesis.** *Gene Expr Patterns* 2003, **3**:261-272.
46. Tahtakran SA, Selleck MAJ: **Ets-1 expression is associated with cranial neural crest migration and vasculogenesis in the chick embryo.** *Gene Expr Patterns* 2003, **3**:455-458.
47. Albazerchi A, Stern CD: **A role for the hypoblast (AVE) in the initiation of neural induction, independent of its ability to position the primitive streak.** *Dev Biol* 2007, **301**:489-503.
48. Messerschmidt DM, Kemler R: **Nanog is required for primitive endoderm formation through a non-cell autonomous mechanism.** *Dev Biol* 2010, **344**:129-137.
49. Singh AM, Hamazaki T, Hankowski KE, Terada N: **A Heterogeneous Expression Pattern for Nanog in Embryonic Stem Cells.** *Stem Cells* 2007, **25**:2534-2542.
50. Chambers I, Silva J, Colby D, Nichols J, Nijmeijer B, Robertson M, Vrana J, Jones K, Grotewold L, Smith A: **Nanog safeguards pluripotency and mediates germline development.** *Nature* 2007, **450**:1230-1234.
51. Canham MA, Sharov AA, Ko MSH, Brickman JM: **Functional Heterogeneity of Embryonic Stem Cells Revealed through Translational Amplification of an Early Endodermal Transcript.** *PLoS Biol* 2010, **8**:e1000379.
52. Debeb BG, Galat V, Epplé-Farmer J, Iannaccone S, Woodward WA, Bader M, Iannaccone P, Binas B: **Isolation of Oct4-Expressing Extraembryonic Endoderm Precursor Cell Lines.** *PLoS One* 2009, **4**:e7216.
53. Huda A, Bowen NJ, Conley AB, Jordan IK: **Epigenetic regulation of transposable element derived human gene promoters.** *Gene* 2011, **475**:39-48.
54. Macfarlan TS, Gifford WD, Agarwal S, Driscoll S, Lettieri K, Wang J, Andrews SE, Franco L, Rosenfeld MG, Ren B, Pfaff SL: **Endogenous retroviruses and neighboring genes are coordinately repressed by LSD1/KDM1A.** *Genes Dev* 2011, **25**:594-607.
55. Meshorer E, Misteli T: **Chromatin in pluripotent embryonic stem cells and differentiation.** *Nat Rev Mol Cell Biol* 2006, **7**:540-546.
56. Conley AB, Piriyaopongsa J, Jordan IK: **Retroviral promoters in the human genome.** *Bioinformatics* 2008, **24**:1563-1567.
57. Faulkner GJ, Kimura Y, Daub CO, Wani S, Plessy C, Irvine KM, Schroder K, Cloonan N, Steptoe AL, Lassmann T, et al: **The regulated retrotransposon transcriptome of mammalian cells.** *Nat Genet* 2009, **41**:563-571.
58. Pi W, Zhu X, Wu M, Wang Y, Fulzele S, Eroglu A, Ling J, Tuan D: **Long-range function of an intergenic retrotransposon.** *Proc Natl Acad Sci* 2010, **107**:12992-12997.
59. Pauli A, Rinn JL, Schier AF: **Non-coding RNAs as regulators of embryogenesis.** *Nat Rev Genet* 2011, **12**:136-149.
60. Khalil AM, Guttman M, Huarte M, Garber M, Raj A, Rivea Morales D, Thomas K, Presser A, Bernstein BE, van Oudenaarden A, et al: **Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression.** *Proc Natl Acad Sci* 2009, **106**:11667-11672.

61. Fazzio TG, Huff JT, Panning B: **An RNAi Screen of Chromatin Proteins Identifies Tip60-p400 as a Regulator of Embryonic Stem Cell Identity.** *Cell* 2008, **134**:162-174.
62. Gaspar-Maia A, Alajem A, Polesso F, Sridharan R, Mason MJ, Heidersbach A, Ramalho-Santos Jo, McManus MT, Plath K, Meshorer E, Ramalho-Santos M: **Chd1 regulates open chromatin and pluripotency of embryonic stem cells.** *Nature* 2009, **460**:863-868.
63. Caillier M, Thenot S, Tribollet V, Birot AM, Samarut J, Mey A: **Role of the Epigenetic Regulator HP1gamma in the Control of Embryonic Stem Cell Properties.** *PLoS One* 2010, **5**:e15507.
64. Schneikert J, Lutz Y, Wasyluk B: **Two independent activation domains in c-Ets-1 and c-Ets-2 located in non-conserved sequences of the ets gene family.** *Oncogene* 1992, **7**:249-256.
65. Niwa H, Yamamura K, Miyazaki J: **Efficient Selection for High-Expression Transfectants with a Novel Eukaryotic Vector.** *Gene* 1991, **108**:193-199.
66. Eyal-Giladi H, Kochav S: **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.** *Developmental Biology* 1976, **49**:321-337.
67. Hamburger V, Hamilton H: **A series of normal stages in the development of the chick embryo.** *J Morphol* 1951, **88**:49-92.
68. Chapman SC, Collignon J, Schoenwolf GC, Lumsden A: **Improved method for chick whole-embryo culture using a filter paper carrier.** *Dev Dyn* 2001, **220**:284-289.
69. Aclouque H, Wilkinson DG, Nieto MA, Dr. Marianne B-F: **Chapter 9 In Situ Hybridization Analysis of Chick Embryos in Whole[hypen (true graphic)] Mount and Tissue Sections.** In *Methods in Cell Biology. Volume 87.* Academic Press; 2008:169-185.
70. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucleic Acids Res* 1997, **25**:3389-3402.
71. Rice P, Longden I, Bleasby A: **EMBOSS: The European Molecular Biology Open Software Suite.** *Trends Genet* 2000, **16**:276-277.
72. Flicek P, Amodè MR, Barrell D, Beal K, Brent S, Chen Y, Clapham P, Coates G, Fairley S, Fitzgerald S, et al: **Ensembl 2011.** *Nucleic Acids Res* 2011, **39**: D800-D806.

doi:10.1186/1742-4690-9-21

Cite this article as: Mey et al.: The endogenous retrovirus ENS-1 provides active binding sites for transcription factors in embryonic stem cells that specify extra embryonic tissue. *Retrovirology* 2012 **9**:21.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

