Sp8 regulatory function in the limb bud ectoderm

Rocío Pérez-Gómez[1], Marc Fernández-Guerrero[1], Víctor Campa[1], Juan F. Lopez-Gimenez[1,†], Alvaro Rada-Iglesias[1,2,3] and Maria A. Ros[1,4]

1) Instituto de Biomedicina y Biotecnología de Cantabria, IBBTEC (CSIC-UC-SODERCAN).
Albert Einstein 22. 39011, Santander, Spain.
2) Center for Molecular Medicine Cologne (CMMC), University of Cologne, Robert-Koch-Strasse 21, 50931 Cologne, Germany
3) Cluster of Excellence Cellular Stress Responses in Aging-Associated Diseases (CECAD), University of Cologne, Joseph-Stelzmann-Str. 2650931 Cologne, Germany
4) Departamento de Anatomía y Biología Celular, Facultad de Medicina, Universidad de Cantabria, 39011 Santander, Spain

[†] current address: Instituto de Parasitología y Biomedicina "López-Neyra", IPBLN, CSIC. Av. del Conocimiento 17, 18016 Granada, Spain.

Author for correspondence:
Marian Ros
Email: rosm@unican.es

ABSTRACT

Sp8 and Sp6 are two closely related Sp genes expressed in the limb ectoderm where they regulate proximo-distal and dorso-ventral patterning. Mouse genetics revealed that they act together in a dose-dependent manner but with Sp8 making a much greater contribution. Here, we combine ChIP-seq and RNA-seq genome-wide analyses to investigate the Sp8 regulatory network and mechanism of action. We find that Sp8 predominantly binds to putative distal enhancers to activate crucial limb patterning genes, including *Fgf8*, *En1*, *Sp6* and *Rspo2*. Sp8 exerts its regulatory function by directly binding DNA at Sp consensus sequences or indirectly through Dlx5 interaction. Overall, our work underscores Sp8 master regulatory functions and supports a model in which it cooperates with other Dlx and Sp cofactors to regulate target genes. We believe that this model could help to properly understand the molecular basis of congenital malformations.

Impact Sentence

In the limb ectoderm, Sp8 regulates master genes through a dual mechanism: directly binding DNA at Sp consensus sequences and indirectly engaging through Dlx5 interaction.

INTRODUCTION

The developing vertebrate limb has long proved as an excellent system for studying the mechanisms involved in pattern formation and morphogenesis, and more recently in transcriptional regulation. Limb development begins with the accumulation under the surface ectoderm of the proliferating limb progenitor cells. Further limb outgrowth and patterning depends on the interaction between these two limb components, epithelia-mesenchymal interactions. An initial interaction leads to the induction of *Fgf8* in the ectoderm generating the apical ectodermal ridge (AER), a crucial signaling center that provides the signals needed for the survival and proliferation of the limb progenitor cells (Fernandez-Teran and Ros, 2008; Tickle, 2015) Concomitantly with the induction of the AER, the activation of *En1* in the ventral ectoderm functions to restrict *Wnt7a*, initially expressed over the whole limb field ectoderm, to its dorsal aspect therefore defining precise dorsal and ventral domains of gene expression. Wnt7a controls the dorsalization of the limb bud by inducing the expression of the homeobox gene *Lmx1b* in the subjacent mesoderm (Fernandez-Teran and Ros, 2008; Tickle, 2015).

It is known that the induction of *Fgf8* and *En1* expression requires active Wnt/βcatenin and Bmp pathways in the limb ectoderm (Ahn et al., 2001; Barrow et al., 2003; Soshnikova et al., 2003) but the regulatory networks involved in these interactions are not completely understood. Understanding the mechanisms that lead to the induction and maintenance of the AER as well as dorsal-ventral (DV) polarity is of maximum interest not only for understanding limb development but also for tissue regeneration and repair.

Recently, the role of Sp6 and Sp8 in the early limb bud ectoderm as mediators of Wnt/βcatenin and Bmp signaling has gained attention (Haro et al., 2014). Sp6 and Sp8 are members of the Specificity protein/Krüppel-like (Sp/Klf) family of transcription factors (TFs) found across almost all metazoan species (Presnell et al., 2015; Suske, 1999). The Sp family is characterized by a highly conserved carboxyterminal DNA binding domain composed of three consecutive C2H2-type zinc finger (ZF) motifs whereas the N-terminal region is more variable (Suske et al., 2005). The buttonhead box and the Sp box are also conserved structural domains characteristic of these proteins.

Sp6 and Sp8 are specifically expressed in the limb ectoderm where they are together absolutely necessary for proximo-distal (PD) and DV patterning. The analysis

of the *Sp8;Sp6* allelic series showed that this requirement is dose dependent and that Sp8 makes a much greater contribution than Sp6, at least in part because Sp8 also controls Sp6 expression (Haro et al., 2014). The progressive reduction of *Sp6;Sp8* gene dosage results in progressively more severe limb phenotypes that go from a mild syndactyly, to Split Hand Split Foot Malformation (SHFM), to oligodactyly, to truncation and finally to amelia. These malformations are also typically presented with bidorsal limb buds and digit tips. The molecular characterization of the *Sp6;Sp8* mutant limb buds showed that a defective activation of *Fgf8* and *En1* was responsible for the phenotype. A reasonable conclusion of these studies, supported by some *in vitro* assays (Sahara et al., 2007), was that Sp8 is a direct transcriptional activator of *Fgf8* and *En1*. However, the full repertoire of Sp8 target genes as well as the molecular mechanisms and genomic context whereby Sp8 regulates them are not completely known.

Here, we have combined ChIP-seq and RNA-seq genome-wide analyses to identify Sp8 direct targets *in vivo* in the mouse limb ectoderm. These approaches revealed that Sp8 has a predominant activator role mainly executed from putative distal enhancers. The identification of the Sp8 regulatory network reveals that Sp8 is a master regulator of limb development regulating a set of crucial genes that include *Fgf8*, *En1* and *Wnt7a* and provides valuable resources for better understanding limb development. Furthermore, using *in silico* motif analyses and functional studies we show that, in addition to binding DNA directly by recognizing the Sp consensus sequence (Sahara et al., 2007), Sp8 also binds DNA indirectly, through an interaction with Dlx5. Dlx5 is one of the six members of the *Dlx* gene family (Kraus and Lufkin, 2006), homologues of the *Drosophila distal-less,* that has already been shown to mediate Sp7 transcriptional regulation in bone (Hojo et al., 2016). We propose that this dual DNA binding mechanism should be considered when evaluating the phenotypes observed both in *Sp* and *Dlx* mutants and related congenital malformations.

RESULTS

Generation of a *Sp8*-tagged knock-in mouse

The genomic targets of Sp8 have not been previously mapped in the limb ectoderm or in any other cellular context. This is probably due to the lack of commercially available ChIP-grade antibodies for Sp8. To overcome this limitation and, thus, globally identify

Sp8 binding sites in the limb ectoderm, we generated a knock-in (KI) mouse in which the endogenous *Sp8* gene was tagged with three copies of the FLAG (FL) epitope using homologous recombination (*Sp8FL*; Fig. 1A). This epitope was incorporated in frame at the C terminus of Sp8 protein, a strategy that has been widely used in ChIP-seq studies. Mice homozygous for the *Sp8FL* allele were viable and fertile and displayed no obvious phenotype indicating that the tagged protein was fully functional (Fig. 1B). Accordingly, the detection of the Sp8FL protein in immunofluorescence assays using the αFLAG M2 antibody (Ref F1804, Sigma) was similar to that of the endogenous protein detected with an αSp8 antibody (Ref 104661, Santa Cruz) (Fig. S1A). The αFLAG antibody also readily detected the Sp8FL protein in western blots (Fig. S1B). Thus, the *Sp8FL* KI model provides a useful resource for future studies requiring detection of Sp8.

Global mapping of Sp8 binding sites in the limb ectoderm

Having generated the *Sp8FL* KI mice, we used ChIP-seq technology to investigate the genome-wide Sp8 occupancy in limb ectodermal cells. For this, the ectodermal hulls of E10.5 forelimb buds of homozygous *Sp8FL* embryos were isolated by mild trypsin digestion (see M&M). The E10.5 developmental stage was selected because it corresponds to a fully matured AER and coincides with the onset of the *Sp8* null phenotype (Bell et al., 2003; Fernandez-Teran and Ros, 2008; Treichel et al., 2003). To reduce the size of the samples, and therefore the number of embryos required, we used the ChIPmentation procedure that combines ChIP and tagmentation and has proven highly efficient for low input samples (Schmidl et al., 2015). ChIPmentation was performed with 50 ectodermal hulls (approx. $7.5x10^5$ cells) using the αFLAG M2 antibody. Following next-generation sequencing, analysis of the ChIPmentation data resulted in the identification of 1,451 high-confidence Sp8 binding sites (Source Data 1).

The quality and specificity of the identified Sp8 peaks was supported by (i) their high evolutionary conservation (Fig. S1C), (ii) their enrichment in H3K4me2, a histone mark characteristic of active enhancers and promoters (Fig. S1D), (iii) the significant overrepresentation of the expected binding motifs in the Sp8 peaks (Fig. S1E-F) and (iv) their reproducibility in a second ChIPmentation biological replicate experiment (Fig. S1G). Thus, an *in-silico* analysis of the sequence motifs in the Sp8 peaks identified a CG-rich/Sp1 motif (E-value 5.9e-209) as the most over-represented one. This is in full agreement with the notion that all Sp family members share the same DNA consensus

binding sequence, frequently referred to as the GC box: 5'-CCGCCC-3' (Fig. S1E) (Suske, 1999). Interestingly, the second most enriched sequence was the AT-rich/Dlx5 motif (E-value 5e-140), previously described as the preferential motif that enables Sp7 to indirectly bind DNA through its interaction with Dlx5 (Hojo et al., 2016).

Sp8 binds proximal and distal regulatory elements with distinct functional and genetic features

Being confident about the quality of our Sp8 binding profiles, we then mapped the locations of Sp8 peaks with respect to the nearest transcription start site (TSS). Sp8 peaks similarly distributed between proximal (48%, <5kb from TSS) and distal (52%, (>5kb from TSS) locations (Fig. 1C) indicating that Sp8 is able to execute its regulatory function in two different genomic contexts, promoters and distal cis-regulatory elements. Both proximal and distal Sp8 associated regions were enriched in H3K4me2 (Fig 1D), indicative of active promoters and enhancers, respectively (Ernst et al., 2011). However, this active state was more conspicuous and widespread among proximal regions (Fig. 1D).

To investigate possible functional and genetic differences depending on whether Sp8 binding occurred in proximal promoters or distal regulatory regions, we performed separate *in silico* analyses of proximal versus distal Sp8 peaks. Firstly, functional annotation of Sp8 peaks using GREAT (McLean et al., 2010) showed that proximal binding was predominantly associated with more general biologic processes, early embryonic expression patterns and diverse developmental abnormalities (Fig. 1E, Source Data 2). In contrast, distal binding sites were predominantly associated with canonical and non-canonical Wnt signaling and were linked to limb bud expression patterns and limb abnormalities (Fig. 1F, Source Data 2). Moreover, we also examined whether Sp8 recognized different DNA sequences when binding proximal or distal regulatory sites. Similarly to what we found for all Sp8 bound peaks (Fig S1E-F), *de novo* motif analysis revealed that both proximal and distal elements were highly enriched in both the Sp family consensus CG-rich motif and the AT-rich/Dlx5 motif (Fig. 1G-H). However, while the CG-rich/Sp1 motif was the most abundant among proximal regions, distal elements were preferentially enriched in the AT-rich/Dlx5 motif. Together, these results suggest two different modes of Sp8 function: i) direct binding to DNA through the Sp consensus sequence mainly at proximal promoters to regulate general biological processes and ii) indirect binding through Dlx5 (and presumably other family members)

mainly at distal putative enhancers to regulate genes with more specific limb developmental functions.

RNA-seq based detection of differentially expressed genes in wild type and *Sp8* null limb bud ectoderm

To further understand the developmental program controlled by Sp8 in the limb ectoderm, we compared the expression profiles of wild-type (WT) versus *Sp8* deficient (*Sp8*^{CreERT2}*;14*) limb bud ectoderms. We used a robust RNA-seq protocol for low-abundance RNA (requirement of 500 pg. of total RNA) that allowed us to only use the two forelimb ectodermal hulls from one embryo for each of the three biological replicates. The samples were obtained from E10.5 WT and *Sp8*-null embryos from the same litter and of matched stage (see M&M for details). Differential gene expression revealed a total of 912 differentially expressed genes (DEG; Fig. 2A) of which 532 were downregulated and 380 upregulated in the *Sp8KO* indicating that Sp8 preferentially functions as an activator (Fig. 2A) (Source Data 3). All genes previously reported to be affected by the loss of *Sp8*, such as *Fgf8*, *En1*, *Msx2*, *Rspo2* and *Sp6* (Bell et al., 2003; Haro et al., 2014; Treichel et al., 2003) were included in the set of DEG, thus validating the quality of our RNA-seq data.

Identification of Sp8 direct targets in the limb bud ectoderm

To distinguish between direct and indirect Sp8 target genes, we intersected the RNA-seq and ChIPmentation datasets and identified 184 DEGs bound by Sp8 that were, therefore, considered as Sp8 direct target genes (Fig. 2B) (Source Data 4). Of them, 55 genes were repressed and 129 activated by Sp8, confirming that Sp8 predominantly functions as an activator in the limb ectoderm. In contrast to the equal genomic distribution of all Sp8 peaks between TSS-proximal and distal regions (Fig. 1C), the Sp8 peaks associated with direct target genes showed a predominant distal distribution, both for repressed (75%) and activated (73%) genes (Fig. 2C).

To gain additional insight into the mechanisms whereby Sp8 might regulate its direct target genes, we then performed *de novo* motif analysis separately on the set of Sp8 peaks associated with either activated or repressed direct target genes. Interestingly, peaks associated with Sp8 activated genes were highly enriched in both AT-rich/Dlx5 (E-value 9.5e-38) and GC-rich/Sp1 (E-value 1.2e-13) motifs (Fig. 2D). Sp8 peaks associated with repressed genes were enriched in a C-rich motif similar, but not identical, to the Sp1

consensus sequence (E-value 5.2e-13), while the over-representation of the AT-rich/Dlx5 motif (E-value 9e-3) was considerably less pronounced (Fig. 2E). Collectively, our data indicate that Sp8 has a predominant activator role mainly executed from putative distal enhancers and involving the cooperation with Dlx5. Since, in addition, all previously known Sp8 targets were also activated by Sp8, further analyses focused on the set of genes directly activated by Sp8.

Two modes of Sp8 action

Our data provide strong evidence for Sp8 exerting its regulatory function either directly binding to the consensus GC-box sequence (Sp1-like motif) as previously demonstrated (Sahara et al., 2007) or indirectly engaging to the AT-rich/Dlx5 binding motif through the association with Dlx5 (Fig. 2D-E). To further evaluate this dual action, the Sp8 peaks associated with activated genes were categorized into four groups according to the presence of Sp and/or Dlx motifs (Fig. 2F; Source Data 5). This analysis showed that one third (33.33%) of the Sp8 regulatory regions contained CG-rich/Sp1 but not AT-rich/Dlx5 sites whereas approximately another third (31.4%) contained AT-rich/Dlx5 but not CG-rich/Sp1 sites. The Sp8 peaks with only CG-rich/Sp1 sites were equally distributed between proximal (44%) and distal (56%) locations while the Sp8 peaks with only AT-rich/Dlx5 motifs were invariably located in distal regions. Another subset of Sp8 regulatory regions (26.9%), showing a predominant distal location (60%), contained both Sp1-like and Dlx5 binding sites. Finally, a small percentage of regulatory regions (8.3%), all of them distally located, did not contain either Sp1-like or Dlx binding sites indicating a possible indirect recruitment of Sp8 through the interaction with additional TFs or a low percentage of false positive Sp8 peaks.

Validation of selected Sp8 target genes

In order to confirm that Sp8 directly regulates genes identified through the previous global analyses, we selected a subset of these genes based on their relevance in limb development and the type of Sp8 binding sites (Fig. 2F) found within their assigned Sp8 regulatory region(s). The selected genes were subject to further analysis including *in situ* hybridization and, in some cases, examination of the enhancer activity of the assigned element in transgenic reporter assays.

Among the genes regulated by peaks containing GC-rich/Sp1 but no AT-rich/Dlx5 motifs, we selected *Fgf8* and *Sp6*. The Sp8-peak assigned to *Fgf8* was located

in the gene promoter (-326 bp) and contained several Sp1-like binding sites supporting direct regulation by Sp8. In the *Sp8* mutant limb bud *Fgf8* is initially activated but its expression decays concomitantly with the onset of the phenotype (Fig. 3A) (Bell et al., 2003; Treichel et al., 2003). We have previously suggested that the initial transient expression of *Fgf8* in the *Sp8* null mutant depends on Sp6, because *Fgf8* is never activated in double *Sp6;Sp8* mutant limb buds. In *Sp8* mutants, *Sp6* expression initiates normally but is not maintained indicating that Sp8 may be required at later stages to maintain *Sp6* expression (Fig. 3B) (Haro et al., 2014). Accordingly, our analysis assigned an intragenic regulatory element, only containing Sp1-like binding sites, to Sp6 (Fig. 3B).

Among the Sp8 direct targets whose assigned regulatory regions contained Dlx5 but not Sp binding sites, we found several components of the Wnt pathway of which we selected *Rspo2* and *Fzd1* for further analysis. *Rspo2* expression is undetectable in the *Sp8* mutant limb ectoderm by *in situ* hybridization, while other domains of expression in the mesoderm remained unaltered (Fig. 3C) (Bell et al., 2008). The putative *Rspo2* enhancer was located 217 kb upstream from the TSS and showed partial activity in the limb ectoderm when assayed in transient mouse transgenesis (6 out of 15; Fig. 3C). Similarly, *Fzd1* expression was undetectable in the *Sp8* mutant limb ectoderm confirming its dependence on Sp8 regulation, presumably through two distal Sp8 binding regions, located 72 kb and 202 kb downstream of the TSS respectively (Fig. 3D).

As Sp8 direct targets regulated by peaks containing both Sp-like and Dlx binding sites, we selected *Wnt7a* and *En1*, the major regulators of limb DV patterning. It is known that the expression of *En1* in the ventral ectoderm rapidly decays in *Sp8* mutants (*13*; Fig. 3E) and is never detected in double *Sp6;Sp8* double mutants (Haro et al., 2014). When tested in a transient transgenic reporter assay, the *En1* enhancer displayed very specific and reproducible activity in the AER (3 out of 4; Fig. 3E). In *Sp8* and compound *Sp6;Sp8* mutants, *Wnt7a* expression extends into the ventral limb ectoderm always in correlation with the reduced or absent *En1* domain. Therefore, it was considered that the extension of *Wnt7a* domain in *Sp8* mutants was an indirect effect mediated by the loss of En1 (Haro et al., 2014). However, here we show that *Wnt7a* is indeed a Sp8 direct target as its expression decreases in the mutant ectoderm according to both RNA-seq (FC=-2.2, p-value=0) and *in situ* hybridization (Fig. 3F). The expression of the above analyzed genes in the mutant and wild type limb bud ectoderm, as calculated by RNA-seq (FPKM), is summarized in Fig 3G.

Other Sp8 direct targets potentially regulated through elements containing GC-rich/Sp1, AT-rich/Dlx5 or both motifs included *Bambi*, *Msx2* and *Wnt5a* (Fig. S2 and Source Data 5). In addition, *Dlx5* and other family members were also identified as Sp8 activated direct targets, as well as Sp8 itself (Fig. S2; Source Data 5). Altogether, our study shows that Sp8 directly regulates essential limb bud ectoderm genes, thus placing this TF as a master regulator of limb development.

Sp8 as a co-factor of the Dlx family in the limb ectoderm

The finding that the AT-rich/Dlx5 motif, but not the CG-rich/Sp1 sequence, was present in a considerable number of Sp8 regulatory regions (Fig. 2F) strongly suggested that Sp8 might be indirectly recruited to some of its *cis*-regulatory targets through interactions with Dlx5, as reported for Sp7 in bone development (Hojo et al., 2016). *Dlx5* is a member of the Dlx family which is expressed in the limb ectoderm and typically co-regulated with *Dlx6* (Kraus and Lufkin, 2006). Remarkably, double *Dlx5/Dlx6* knockout mice display a Split hand foot malformation (SHFM) phenotype similar to *Sp6/Sp8* compound mutants that retain a functional copy of *Sp8* (Haro et al., 2014; Robledo et al., 2002). Thus, a functional interaction between Sp8 and Dlx5, and presumably co-expressed members of these two families of TFs, seems a reasonable hypothesis to explore.

The Dlx5-Sp8 interaction was first investigated by co-immunoprecipitation (CoIP) in heterologous HEK293 cells expressing epitope-tagged (FLAG or Myc) Sp8 and Dlx5 proteins. The results confirmed that Sp8 and Dlx5 physically interacted and the interaction required the ZF domain of Sp8 (Fig. 4A), similarly to Sp7 (Hojo et al., 2016). The functional relevance of this interaction was investigated with a luciferase reporter assay using a construct containing 12 tandem copies of the AT-rich/Dlx5 motif (kindly provided by Dr. McMahon (Hojo et al., 2016); Fig. 4B). It has been shown that co transfection with Sp7 dramatically potentiates Dlx5 driven reporter gene activity from this construct in NIH3T3 fibroblasts (3T3), indicating strong synergistic activity between Sp7 and Dlx5 (Hojo et al., 2016). However, Sp8 did not display any synergistic effect with Dlx5 in similar experiments but rather inhibited the Dlx5 dependent reporter activity suggesting that Sp8 functions as a repressor in this in vitro assay (Fig. 4B and Fig. S3A). This result is in high contrast with our genomic analysis and may reflect the inability of the cellular system to recapitulate the situation in the embryonic limb ectoderm. However, searching for an explanation for this discrepancy and given that Sp6 works together with Sp8 in the limb ectoderm (Haro et al., 2014) we decided to check the activity of Sp6 in

this experimental system. As for Sp8, CoIP experiments in heterologous HEK293 cells demonstrated that Sp6 also interacted with Dlx5 through its ZF domain (Fig. 4C). Interestingly, Sp6 transfection alone led to substantial reporter activation (Fig. 4B and Fig. S3B), which was further increased by co-transfection with Dlx5 (Fig. 4B) indicating that Sp6 and Sp8 are not functionally equivalent, contrary to the current understanding based on mouse genetics (Haro et al., 2014). The ability of Sp6 to activate alone a reporter construct that only contains copies of the AT-rich/Dlx5 motif but not of the CG-rich/Sp1 sequence could be explained by the high expression levels of Dlx family members (Dlx1 and Dlx2) in 3T3 cells (see M&M). Interestingly, co-expression of the three factors, Dlx5, Sp8 and Sp6, resulted in reporter activation comparable to that for Dlx5 alone, suggesting that the simultaneous presence of Sp8 and Sp6 might interfere with their individual interaction with Dlx5 (Fig. 4B). The specificity of the results of our luciferase assay was demonstrated by the lack of reporter expression in cells transfected with a reporter vector in which the AT-rich/Dlx5 motifs have been mutated (kindly provided by Dr. McMahon (Hojo et al., 2016); Fig. 4B).

Based on the previous results and considering that Sp family members have been shown to heterodimerize (Pascal and Tjian, 1991), we decided to evaluate in more detail the Sp8/Sp6 protein-protein interaction by Bimolecular Fluorescent Complementation (BiFC). For this, Sp8 and Sp6 were fused in frame at their C-terminus to the Yellow Fluorescent Protein (YFP) or to the YFP N-terminal (YN; residues 1-172) or C-terminal (YC; residues 173-240) moieties (see M&M). As expected, Sp8 and Sp6 showed nuclear localization (Fig. S3C) and the BiFC assays revealed that both factors can homo and heterodimerize in the nucleus and that this interaction requires the ZF domain (Fig. 4D-G). Heterodimerization between Sp8 and Sp6 was also confirmed by CoIP (Fig. 4H).

The fact that Sp8 utilizes the same domain, the ZF, to recognize its cognate DNA sequence and to mediate protein interactions clearly reflects the complexity and context-dependence of its regulatory function. This might explain how the function of Sp8 can be appraised as an activator in global genomic studies but as a repressor in individual in vitro reporter assays (see comments in Discussion). Collectively, our results support a model in which Sp8, Sp6 and Dlx5 act conjointly to regulate target genes with a final functional outcome that depends on the relative availability of the interacting TFs.

DISCUSSION

Here, to determine the molecular mechanisms underlying Sp8 transcriptional activity in the limb ectoderm, we performed ChIPmentation and RNA-seq at a genome-wide scale. To distinguish functional from non-functional binding we intersected the Sp8 chromatin binding profile with the set of DEG upon knocking out *Sp8*. This strategy identified 184 direct targets, the majority of which (70%) were activated by Sp8 mostly from distant regulatory elements (70%). In addition, by analyzing the binding motifs in the Sp8-associated DNA regions and functional assays, we identify its dual and complex mode of action. The set of Sp8 activated direct targets included the major regulators of limb patterning in the ectoderm, underscoring Sp8 role as a master transcriptional regulator of limb development.

The Sp8 regulatory network includes the major regulators of PD and DV patterning in the limb ectoderm.

The ectoderm of the emerging limb bud is patterned in specific domains of gene expression. The ventral ectoderm expresses *En1* whereas the dorsal ectoderm expresses *Wnt7a*. During normal development the AER always forms at the DV boundary evidencing the tight link between PD and DV patterning (Fernandez-Teran and Ros, 2008). Mouse genetic experiments showed that Sp8 and Sp6 may be this link by regulating *Fgf8* and *En1* expression (Haro et al., 2014). Here we show how Sp8 performs this regulatory function and confirm that *Fgf8* and *En1* are Sp8 activated direct targets.

The mouse *Fgf8* promoter is considered to span about 700 bp upstream from the TSS and is decorated by promoter-associated histone marks in *Fgf8*-expressing tissues (Shen et al., 2012). Here, we extend previous *in vitro* reports (Sahara et al., 2007) by showing that the Sp8 control over *Fgf8* is exerted through direct binding to GC boxes in its promoter. Because Sp family members share the same DNA binding sequence, it is likely that Sp6, and at later stages Sp9, also participate in the modulation of *Fgf8* expression (Haro et al., 2014; Kawakami et al., 2004) *Fgf8* is also regulated by multiple enhancer modules located in a wide region centromeric to the gene (Marinić et al., 2013). Duplications of this genomic region cause the SHFM3 (OMIM 246560), considered to result from altered *Fgf8* expression due to the disrupted genomic architecture. However, despite abundant Sp1-like binding sites, our analysis detected no Sp8 binding in this intricate regulatory ensemble. We also find that Sp8 activates *En1* from a distally located

enhancer that contains both Sp and Dlx5 binding motifs and that drives activity in the AER in mouse transgenic assays.

In addition, our study also unexpectedly reveals that *Wnt7a* is a Sp8 activated direct target. In *Sp8* mutants, *Wnt7a* expression extends to the ventral ectoderm in parallel to the loss of *En1*. However, disregarding the domain extension, the level of expression is notably reduced in *Sp8* mutants and we identify a distal element, containing both Sp and Dlx5 binding motifs, through which Sp8 presumably regulates *Wnt7a*. These findings support Sp8 involvement in the connection between PD and DV patterning and add more complexity to DV patterning regulation. A major still open question is how Sp8, that is initially expressed throughout the whole limb ectoderm participates in the generation of the defined DV expression domains.

Sp8 potentiates Wnt/βcatenin signaling in the limb ectoderm.

Sp8 is a well-recognized mediator of the Wnt/βcatenin dependent induction of *Fgf8* in the limb ectoderm and also in other systems such as the genital tubercle and the embryonic telencephalon (Haro et al., 2014; Lin et al., 2013; Sahara et al., 2007). Together with Sp5, Sp8 is also a mediator of Wnt/βcatenin dependent induction of neuromesodermal progenitors (Kennedy et al., 2016). Accordingly, our gene ontology analysis showed a significant enrichment of Wnt pathway components on genes assigned to distal Sp8 regulatory regions. We have particularly studied *Rspo2* and *Fzd1* which are regulated by Sp8 through distal located elements and very likely through Dlx5 interaction as these elements lack the typical Sp-binding sites. Rspo2 is a secreted agonist of the canonical WNT/βcatenin signaling pathway that prevents degradation of Fzd receptors by recruiting the Rnf43 and Znrf3 E3 ligases through or independently of *Lgr4/5/6* genes (Lebensohn and Rohatgi, 2018; Szenker-Ravi et al., 2018). *Fzd1* is a Wnt receptor predominantly expressed in the ventral limb ectoderm downstream of BMP signaling (Soshnikova et al., 2003). Through the transcriptional control of *Rspo2* and *Fzd1*, Sp8 potentiates Wnt signaling in the limb ectoderm, in addition of acting itself as a Wnt mediator (Haro et al., 2014; Kennedy et al., 2016; Lin et al., 2013).

Sp8 has a dual mode of action

An important conclusion of our study is the identification that Sp8 may execute its transcriptional regulation either directly binding to the consensus GC-box sequence or indirectly through association with Dlx5. All Sp family members were thought to share

the same DNA recognition site, the GC box (Suske et al., 2005), to which Sp8 has been demonstrated to bind (Sahara et al., 2007) but recently it was shown that Sp7 uses a distinct mode of action. Sp7 is unable to bind the Sp consensus motif and binds DNA though Dlx5 interaction (Hojo et al., 2016). Interestingly, *de novo* motif discovery in Sp8-bound genomic regions, besides the expected overrepresentation of the Sp consensus motif, it also identified the AT-rich/Dlx5 motif as the second top scored motif. Moreover, the AT-rich/Dlx5 motif was the top scoring motif when only Sp8 activated direct targets were considered. These results strongly suggested that Sp8 could also bind DNA through Dlx5 association, a possibility that was sustained by the overlapping expression patterns of these TFs in the limb ectoderm and validated by CoIP experiments that also showed that this interaction requires the Sp8 ZF domain. In addition, Dlx and Sp genes share mutant phenotypes. Indeed the SHFM phenotype is characteristic of both the loss of *Dlx5/6* and of the significant reduction in *Sp6;Sp8* gene dosage (Haro et al., 2014; Robledo et al., 2002).

Sp7 distinct mode of action is due to the presence of three variant amino acid residues in the ZF domain that impair the interaction with the GC-box while favoring Dlx binding (Hojo et al., 2016). Since, Sp8 does not share the Sp7 specific variants, we conclude that these variants are not required for Dlx5 interaction but rather act in impairing GC-box recognition. It is likely that all Sp family members, or at least those in the Sp6-Sp9 clade, are able to associate Dlx, as we have also demonstrated here for Sp6.

Dlx proteins bind DNA through their homeodomain and their transcriptional activity may be modulated by other TFs such as Msx1/2 in craniofacial development (Satokata and Maas, 1994) and Sp7 in bone development (Hojo et al., 2016). During osteoblast specification, Sp7 is recruited to osteoblast enhancers by Dlx5 and possibly other family members. Here we present compelling evidence of a Dlx-Sp8 regulatory complex acting in the limb ectoderm, indicating that the Dlx-Sp complex may function in diverse cellular contexts. Indeed a Dlx-Sp module is highly conserved in evolution and has been shown to function during the development of vertebrate and insect (Estella and Mann, 2010; Franch-Marro et al., 2006) appendages as well as in the regeneration of the prototypic planarian eye (Lapan and Reddien, 2011).

Complexity of Sp8 transcriptional regulation

The results of the genomic analysis indicates that Sp8 predominantly functions as an activator while the luciferase assays indicate that it decreases Dlx5 transcriptional

activity although this negative effect was modified by the presence of Sp6. This apparent discrepancy is likely due to the inability of the *in vitro* cellular system to replicate the situation in the limb ectoderm in vivo (e.g. differences in the repertoire of cofactors) and clearly reflects the complexity of Sp8 regulatory function. This complexity resides, at least in part, in Sp8 utilizing the ZF domain for both DNA and protein interactions. Therefore, the presence of cofactors may significantly influence Sp8 function favoring direct DNA binding versus interaction with other TFs or vice versa. Thus, in the limb ectoderm the interactions between Sp8, Sp6 and Dlx5 (and presumably other Sp and Dlx family members) may significantly and reciprocally impact their transcriptional regulatory activity depending on affinity and availability. An additional and critical level of complexity in Dlx-Sp interactions is given by the fact that Sp6 and all the Dlx genes are directly activated by Sp8. Thus, although in *Sp8* mutants the repressor function of Sp8 might be lost in regions containing only Dlx-binding sites, the concomitant decay of Sp6 and Dlx would result in a net downregulation of the target genes associated with these regions. This could explain why *in vivo* Sp8 seems to preferentially act as an activator, regardless of whether its target genes are associated with regulatory elements containing Dlx or Sp1-like motifs.

In summary, our study places Sp8 as a crucial regulator of limb development in the limb ectoderm and underscores the functional importance of the Dlx-Sp interactions that should be considered when evaluating the phenotypes observed both in *Sp6/8* and *Dlx5/6* mutants and related congenital malformations.

MATERIAL AND METHODS

Mouse embryos

Wild type C57BL/6 mice and the $Sp8^{FL}$ and $Sp8^{CreERT2}$ (Treichel et al., 2003) mouse strains were used in this study. Genotyping was performed using tail biopsies or embryonic membranes according to previously published reports (Haro et al., 2014). Embryos of the desired embryonic day were obtained by cesarean section. All animal procedures were conducted accordingly to the EU regulations and 3R principles and reviewed and approved by the Bioethics Committee of the University of Cantabria. All mice were maintained in a C57BL6 genetic background and genotyped by PCR protocols from tail biopsies or yolk sacs samples following standard protocols.

***Sp8:3xFLAG* knock-in mouse generation (*Sp8^{FL}*)**

Three copies of the FLAG epitope (5'-GACTACAAAGACCATGACGGTGATTATAAAGATCATGATATCGATTACAAG GATGACGATGACAAG-3') were inserted in frame at the 3'-terminal end of the *Sp8* gene with a stop codon (Fig. 1A). Mouse genomic fragments containing 5' (chr12:118845063+118849869) and 3' (chr12:118849873+118852808) homology arms were amplified from BAC clone by using high fidelity Taq polymerase and were sequentially assembled into a targeting vector together with recombination sites and selection markers (DTA and Neo cassette). Targeted ES clones were confirmed via Southern Blotting and some of them were selected for blastocyst microinjection (Cyagen Biosciences Inc., Santa Clara, California, US).

Isolation of limb bud ectodermal hulls

E10.5 forelimb buds were dissected in cold PBS and incubated in 0.25% trypsin (SV30037.01, GE Healthcare, Logan, Utah) on ice for 20 min. After a quick rinse in 10% FBS to inactivate trypsin, the limb buds were transferred to cold PBS where the ectoderm was separated from the mesoderm with fine forceps.

Skeletal preparations, *in situ* hybridization in paraffin sections and immunofluorescence

Alcian blue 8GX and Alizarin red skeletal staining was performed following standard protocols, cleared by KOH treatment and stored in glycerol. *In situ* hybridization was performed on paraffin sections with digoxigenin-labeled antisense riboprobes following standard procedures. Immunostaining was performed in cryostat sections (14 µm) using the mouse monoclonal anti FLAG M2 antibody (1:500; Ref F1804; Sigma-Aldrich, St Louis, Missouri) and the anti Sp8 C18 antibody (1:500; Ref 104661, Santa Cruz Biotechnology, Dallas, Texas) and then coupled with Alexa 488-conjugated anti-mouse IgG secondary antibody (1:500, Invitrogen, Molecular Probes, Eugene, Oregon). Slides were analyzed in a Leica Laser Scanning Confocal TCS-SP5 with a 63x 1.4 NA objective.

Transgenic mice for LacZ activity

The genomic region containing the regulatory elements of *En1* (mm10, chromosome 1:120,621,403-120,622,419) and *Rspo2* (mm10, chromosome 15:43,388,035-43,389,518) were PCR-amplified using specific primers containing XhoI restriction

enzyme sites and cloned into a vector carrying the β-globin minimal promoter and the LacZ coding sequence (pSK-LacZ; kindly provided by Denis Duboule lab). Transient transgenic embryos were harvested at E10.5 and only those embryos containing the reporter transgene were processed for detection of β-galactosidase activity according to standard procedures.

Plasmids constructions

Sp8 and Sp6 tagged with FLAG or Myc epitopes at their 5' end were amplified by PCR and cloned into pcDNA3 including the appropriate restriction enzyme site to be fused at the 3'end to the full length YFP (1-240aa) or to its N-terminal (YN; 1-172aa) or C-terminal (YC; 173-240aa) moieties. Deletion of the ZF domain of Sp8 (nt 353-486) and of the ZF domain of Sp6 (nt 254-336) was constructed by PCR strategy. The following clones were generated: FLAG-Sp8, FLAG-Sp6, HA-Sp6□ZF, Sp8-YFP, Sp8-YN, Sp8-YC, Myc-Sp6-YFP, Myc-Sp6-YN, Myc-Sp6-YC, Myc-Sp8□ZF-YFP, Myc-Sp8□ZF-YN, Myc-Sp8□ZF-YC. The Myc-Dlx5 and FLAG-Dlx5 clones were kindly provided by Dr. Andrew P. McMahon (Keck School of Medicine of the University of Southern California, Los Angeles, USA) (Hojo et al., 2016). The HA-Sp6□ZF clone was synthesized by NZYTech (Lisbon, Portugal).

Bimolecular Fluorescent Complementation (BiFC)

HEK293 cells were grown on microscope cover Glasses (Ø 18 mm) and transfected with PEI and 1 µg total DNA. To test for homodimerization, 500ng of indicated expression plasmids, each one with a complementary YFP moiety (e.g. Sp8-YN + Sp8-YC), were transfected. Single transfection of the protein fused with the complete YFP or with one of its moieties was used for positive and negative control. Transfected, cells were incubated with Hoechst33342 and confocal images (512x512 pixels; 0.15mm pixel size) were acquired sequentially on a SP5 laser-scan microscope (Leica) with a 63x 1.4 NA objective. Cells were excited sequentially with 405nm 514nm laser lines and fluorescence emission captured between 420-480nm (Hoechst) and 525-600nm (YFP). Images are presented after digital adjustment of curve levels (gamma) to maximize signal with ImageJ software. Fluorochromes and colors are as indicated in the figure legends.

Co-immunoprecipitation, SDS-PAGE, and Western Blotting

HEK293 cells were co-transfected with the desired plasmids (1 µg total DNA) and lysed (20 mM Hepes (pH 7.5), 10 mM EGTA pH 8, 40 mM Glycerol-phosphate, 1% NP-40, 2.5 mM MgCl2, 2 mM orthovanadate, 1 mM DTT, 1x PICS, 1 mM PMSF). CoIP was performed with 0.5-1mg of protein, 10 µL of Dynabeads™ Protein G (Thermo Fisher Scientific, 10004D) and 1µg of the desired antibody: FLAG-M2 (Sigma F1804) or MYC (Cell Signaling #2272) followed by SDS-PAGE analysis of protein products. FLAG-M2 and MYC antibodies were used for immunoblotting and β-Actin (Santa Cruz Ref 47778) was used as loading control when needed. Results were confirmed in at least two independent experiments.

Luciferase reporter assay.

NIH 3T3 cells were plated at a density of 25,000 cells/cm2 (22Rv1) in 24-well plates and the next day were transfected by Lipofectamine™ 3000 (L3000-008 Invitrogen by Thermo Fisher Scientific Life Technologies, Carlsbad, California) with 100 ng of firefly luciferase reporter p12xAT_luc or p12xAT_luc(Mut) (Hojo et al., 2016) 15 ng of the pRL-TK Renilla luciferase control plasmid (Promega) and 100ng (unless otherwise indicated) of Myc-Dlx5, FLAG-Sp8 or FLAG-Sp6 expression plasmids. 48 hours later, luciferase activity was measured with the DualGlo Stop&Glo luciferase assay system (Ref E2920 Promega, Madison, Wisconsin), according to manufacturer's instructions. Expression profile of NIH 3T3 cells at (http://amp.pharm.mssm.edu/Harmonizome/gene_set/nih+3T3/BioGPS+Mouse+Cell+Type+and+Tissue+Gene+Expression+Profiles).

ChIPmentation of Sp8

ChIPmentation protocol was based on (Schmidl et al., 2015). In order to reduce background and unspecific binding, only the limb bud ectoderm was used. A total of 50 ectoderms *Sp8*$^{FLAG/FLAG}$ were used for each TF ChIPmentation replicate (2 replicates). ChIPmentation (and ChIP-seq) sequencing reads were mapped to the mouse genome (mm10 assembly) using Burrows-Wheeler Aligner (BWA) (Li and Durbin, 2009). Duplicate reads were filtered, and only unique reads were considered. The resulting Binary Alignment/Map (BAM) files were then analyzed with MACS2 (Zhang et al., 2008) using the following settings to identify genomic regions significantly enriched in the investigated proteins in comparison to the total genomic input DNA: q value 0.1; Fold change >3. Pearson Correlation coefficients of the two SP8 ChIPmentation biological

replicates were performed with the bamCorrelate tool from deepTools (bins mode and a bin size of 10 kb across the whole mouse genome) (Ramírez et al., 2014). For the generation of the heatmaps, BAM files were normalized as RPGC (reads per genome coverage) and then used to visualize scores associated with genomic regions using deepTools (Ramírez et al., 2014). Because of the higher signal to noise ratio, only the dataset of the first biological ChIPmentation sample (1,451 genomic regions) was used for further analysis. Average vertebrate PhastCons score profiles around the center of enhancer sequences were generated with the Conservation Plot tool from the Cistrome Analysis pipeline (http://cistrome.dfci.harvard.edu/ap/root) (Liu et al., 2011).

Chip-seq for H3K4me2

Chromatin Immunoprecipitation (ChIP) protocol was based on (Rehimi et al., 2017). Two replicates, each using 65 ectoderms of E10.5 wild type forelimbs and the ActiveMotif 39141 antibody were performed.

Genomic Regions Enrichment of Annotations Tool (GREAT) analysis

For calculating the distribution of all the Sp8 peaks with respect to the TSS of annotated genes as well as for their *in silico* functional annotation analyses, we used the GREAT tool (GREAT 3.0.0) (McLean et al., 2010) with the "Basal Plus Extension Association Rule" settings. Proximal and Distal peaks were individually analyzed with the "Single Nearest Gene Association Rule" settings.

Motif analysis

De novo motif discovery analyses of Sp8 peaks were performed using the online tool MEME-ChIP (Version 5.0.1, http://meme-suite.org/tools/meme-chip) as described by (Ma et al., 2014). Input sequences were centered within summit regions of recovered intervals obtaining motifs that are between 6 and 20 bp wide with an E-value cut-off of >0.5 for the discovery of enriched motifs.

For the prediction of potential TFBS, we used the MatInspector (MI) tool provided by *Genomatrix* software suite (Munich, Germany) (Cartharius et al., 2005). FASTA sequences of SP8 peaks were loaded into MI tool, being scanned for matches to a library based on position weight matrices (PWM), identifying putative binding sites (matrix similarity scores not less than 0.75/optimized).

RNA-seq

Ectoderms from E10.5 wild-type and homozygous *Sp8*-null (*Sp8*$^{Cre\text{-}ERT2}$ line) embryos were individually collected. Experiments were performed in triplicates and each replicate contained the 2 forelimb ectoderms of a single embryo. RNA was extracted with RNeasy-Plus Micro Kit (Ref 74034 Qiagen GmbH, Hilden, GERMANY). Pre-amplification using the Ovation RNASeq System V2 was performed. Total RNA was used for first strand cDNA synthesis, using both poly(T) and random primers, followed by second strand synthesis and isothermal strand-displacement amplification. For library preparation, the Illumina Nextera XT DNA sample preparation protocol was used, with 1 ng cDNA input. After validation (Agilent 2200 TapeStation) and quantification (Invitrogen Qubit System) all six transcriptome libraries were pooled. The pool was quantified using the Peqlab KAPA Library Quantification Kit and the Applied Biosystems 7900HT Sequence Detection and pooled on one lane of an Illumina HiSeq4000sequencing instrument with a 2x75 bp paired-end read length.

To analyze the data, high-throughput next-generation sequencing analysis pipeline (Wagle et al., 2015) was used. Basic read quality check was performed using FastQC (Babraham Bioinformatics) and read statistics were obtained with SAMtools. Reads were mapped to the mouse reference assembly (GRCm38), using TopHat2 (Kim et al., 2013). Read count means, fold-change (FC) and values were calculated with DEseq2 (Anders and Huber, 2010) and gene expression for the individual samples was calculated with Cufflinks (Trapnell et al., 2010) as FPKMs, using in both cases genomic annotation from the Ensembl database. Differentially expressed genes (DEG) were obtained considering only genes with an Ensembl ID and an official Gene Symbol applying the following criteria:

For Sp8 wild-type vs *Sp8KO*:

• Up-regulated in *Sp8KO*: p-value<0.01, FC>1.5, FPKM in *Sp8KO* sample>0.1.

• Down-regulated in *Sp8KO*: p-value<0.01, FC<1.5, FPKM in wild-type sample>0.1

Sp6 expression was not properly calculated by the computational pipeline described above due to the presence of an antisense transcript that largely overlaps with Sp6 and that got assigned most of the RNA-seq reads (i.e. ENSMUSG00000087067). Since according to the RNA-seq data ENSMUSG00000087067 was downregulated in *Sp8KO* embryos and our RNA-seq data was not strand-specific, this indicates that Sp6 is indeed downregulated in *Sp8KO* embryos. This was further confirmed by ISH and is also supported by our previous findings (Haro et al., 2014). Therefore, Sp6 was manually

added to the list of genes identified as differentially expressed between WT and *Sp8KO* embryos by RNA-seq. Moreover, in Fig 3G, the FPKM values measured for ENSMUSG00000087067 were assigned to Sp6 instead. The original ChIP-seq and RNA-seq data of this paper have been deposited in GEO and can be accessed with the GSE144733 accession number.

Acknowledgments

We thank Laura Galán, Sandra Zunzunegui and Mar Rodriguez for excellent technical assistance. We thank Victor Sanchez Gaya for help with the management of genomic data. We are very grateful to Dr. Hojo (University of Tokyo) and Dr. McMahon (University of Southern California) for the p12xAT_luc, the p12xAT_luc and Dlx5 constructs (Hojo et al., 2016). This research was supported by the Spanish Ministry of Science, Innovation and Universities Grant BFU2017-88265-P to MAR and Grant PGC2018-095301-B-I00 and "Programa STAR-Santander Universidades, Campus Cantabria Internacional de la convocatoria CEI 2015 de Campus de Excelencia Internacional" to AR-I.

Author contributions: R.P-G performed the majority of the experiments. M.F-G., V.C. and J.F.L-G. contributed to experiments. A.R-I. contributed to the design of the project and to the writing of the manuscript. M.A.R. conceived the project, performed experiments and wrote the manuscript. All the authors edited the manuscript.

Competing interests: The authors declare that they have no competing interests.

Data and materials availability: The original ChIPmentation, ChIP-seq and RNA-seq data of this paper have been deposited in GEO and can be accessed with the GSE144733 accession number.

Source Data 1-5

Source Data 1- Sp8 ChIPmentation peaks (Excel)
Source Data 2- In silico functional annotation of Sp8 peaks using GREAT (Excel)
Source Data 3- Differentially expressed genes between WT and *Sp8*-null embryos (Excel)
Source Data 4- Sp8 direct target genes (Excel)
Source Data 5- Categorization of Sp8 peaks according to type of binding sites (Excel)

FIGURE LEGENDS

Figure 1. Generation and validation of *Sp8-3xFLAG* knock-in mice and ChIP-seq genome-wide analysis of Sp8 DNA-binding profile in the limb bud ectoderm.

A) Scheme showing the generation of the *Sp8-3xFLAG* (*Sp8FL*) allele by homologous recombination (homology arms in orange). Three copies of FLAG (dark blue) were introduced at the 3'-terminal end of the *Sp8* gene before the STOP codon TGA (middle of third exon). Numbers 1, 2, 3 indicate *Sp8* exons (white). **B**) Skeletal preparations of newborn mice failed to detect any difference between wild-type (WT), and *Sp8FL* hetero and homozygotes. **C**) Distribution of Sp8 peaks relative to the transcriptional start sites (TSSs). **D**) H3K4me3 ChIP-seq signal density around proximal and distal Sp8 peaks. **E-F**) Proximal and distal Sp8 peaks were functionally annotated using GREAT (McLean et al., 2010). Only five of the twenty most overrepresented terms belonging to the following three major gene ontologies are shown: Biological Process (red), Mouse Genome Informatics (MGI) Expression: detected (green), and Mouse Phenotype (yellow). **G-H**) Most enriched motifs identified through *de novo* motif analysis of proximal and distal Sp8 peaks.

Figure 2. Sp8 direct target genes

A) Genes were plotted according to the average normalized RNA-seq read counts in WT and *Sp8*-null ($Sp8^{Cre\text{-}ERT2}$) E10.5 limb bud ectoderms. Each gene is represented by a dot. Genes considered as significantly up and downregulated in the *Sp8*-null ectoderm are shown in red and blue, respectively. The name of some relevant genes is indicated. **B**) Integration of the Sp8 ChIP-seq dataset (pink) with the Sp8 regulated genes (RNA-seq) identified 184 direct target genes 55 repressed and 129 activated by Sp8. **C**) Comparison of the distribution of all Sp8 binding regions (pink) with those associated with repressed (green) and activated direct targets (purple). **D-E**) Most representative motifs uncovered through a *de novo* motif analysis of Sp8 peaks linked to repressed (D) and activated (E) Sp8-direct targets. **F**) Categorization of Sp8 regulatory regions according to the presence of Sp1-like and/or Dlx binding motifs.

Figure 3. Validation of Sp8 activated direct targets.

(A-F) For each of the indicated genes, the corresponding panels include a screenshot of the UCSC genome browser displaying the Sp8-associated regulatory module (green), the

Sp8 ChIPmentation track (blue), the H3K4me2 ChIP-seq track (red), the conservation track and the Sp-1 like (purple) and/or Dlx5-like binding motifs (orange) and, on the right, the *in situ* hybridization evaluation of the pattern of expression in longitudinal sections of the limb bud (transverse sections of the embryo) of *Sp8*-null versus control limb buds. The green arrow indicates the control AER and the red arrow indicates altered expression in the mutant AER and limb bud ectoderm. In addition, for *Rspo2* (C) and *En1* (E), the associated Sp8-regulatory regions (highlighted in blue) were evaluated in transient transgenic reporter assays. The *Rspo2* associated regions showed activity in the ventral limb bud ectoderm (arrowheads), while *En*1 associated region showed activity in the limb bud ectoderm including the AER (arrow). **G**) Bar chart of average FPKM values (RNA-seq) in *Sp8*-null and wild-type limb bud ectoderm. At least three mutant specimens per probe were examined. Error bars show standard errors. FL: fore limb; HL: hind limb.

Figure 4. Characterization and functional analysis of Dlx5, Sp8 and Sp6 interactions.

A) Western blots after co-immunoprecipitation (Co-IP) show that Sp8 interacts with Dlx5 in co-transfected HEK293 cells (left panel) and that this interaction requires the Sp8 ZF domain (right panel). **B**) Luciferase activity of the luciferase reporter construct containing 12 tandem copies of the AT-rich/Dlx5 motif or this reporter with all the 12 copies mutated (Hojo et al., 2016) after co-transfection with Dlx5, Sp8 or Sp6 as indicated. Luciferase activity was normalized to Dlx5 and is shown as relative fold activation. Each experiment was performed in duplicate and was repeated at least three times. Error bars represent standard deviations calculated from the biological triplicates. *P<0.05 ***P<0.005. **C**) Western blots after co-immunoprecipitation (Co-IP) show that Sp6 interacts with Dlx5 in co-transfected HEK293 cells (left panel) and that this interaction requires the Sp6 ZF domain (right panel). **D-G**) Bimolecular fluorescence complementation (BiFC) showed that Sp8 and Sp6 form homo and heterodimers in the nucleus of HEK293 cells. Co-transfected constructs indicated at the top. HEK293 cells were stained with Hoechst before analysis. For each experiment, the YFP channel (left panel) and Hoechst plus YFP channels (right panel) are shown. Scale bar: 10 μm. BiFC data are representative of at least three independent experiments. **H**) CoIP showing heterodimerization between Sp6 and Sp8 (bottom). TL (Total Lysate), C- (Negative control), IP (Immunoprecipitation). Structural domains indicated at the bottom right.

Supplementary Figure 1. Validation of *Sp8-3xFLAG* knock-in mice and of the Sp8 ChIP-seq.

A) Immunofluorescence with anti-FLAG, and with anti Sp8-C18 antibodies in cryostat sections of in cryostat transverse sections of an E10.5 $Sp8^{FL/FL}$ embryo showing similar expression pattern in the neural tube (NT) and in the limb ectoderm. **B**) Western blot showing detection of Sp8FL in $Sp8^{FL/FL}$ but not in WT E10.5 embryos. **C**) Average vertebrate PhastCons score profiles around the center of Sp8 binding regions showing evolutionarily conservation. D) Heatmaps of H3K4me2 chromatin regulatory mark around Sp8 binding regions (+3 kb and -3 kb from the peak center). **E)** Sp family consensus binding motif. **F**) The three most significant sequence motifs enriched at Sp8 ChIP-seq peaks identified by *de novo* motif analysis. **G**) To assess the reproducibility of our Sp8 ChIP-seq data, Sp8 peaks were identified in two independent Sp8 ChIP- seq experiments. The overlapping peaks in these two experiments were then used to visualize the Sp8 ChIP-seq signals in the two independent replicates as density heaptmaps (top) as well as to calculate the around the Pearson Correlation in ChIP-seq signals between the two replicates (bottom).

Supplementary Figure 2. Validation of *Sp8* direct target genes.

A) Comparison by *in situ* hybridization in tissue sections of the pattern of expression of the genes indicated in *Sp8*-null versus control limb buds. All are longitudinal sections of the limb (transverse sections of the embryo). **B**) Bar chart comparing average FPKM values (RNA-seq) in *Sp8*-null and control limb bud ectoderms. Error bars show standard errors.

Supplementary Figure 3. Functional analysis of Dlx5, Sp8 and Sp6 interactions and nuclear localization of Sp8 and Sp6.

A) Luciferase activity of the luciferase reporter construct containing 12 tandem copies of the AT-rich/Dlx5 motif (Hojo et al., 2016) after co-transfection with Dlx5 (50ng) and growing concentrations of Sp8 as indicated. **B**) Luciferase activity of the luciferase reporter construct containing 12 tandem copies of the AT-rich/Dlx5 motif (Hojo et al., 2016) after co-transfection with Dlx5 (50ng) and growing concentrations of Sp6 as indicated. Each experiment was performed in triplicate and was repeated at least three times. Error bars represent standard deviations calculated from the biological triplicates.

C) Bimolecular fluorescence complementation showing Sp8 and Sp6 nuclear localization in HEK293. Transfected construct indicated at the top. Scale bar: 10 µm

REFERENCES

Ahn K, Mishina Y, Hanks MC, Behringer RR, Crenshaw EB. 2001. BMPR-IA signaling is required for the formation of the apical ectodermal ridge and dorsal-ventral patterning of the limb. *Development* **128**:4449–4461.

Anders S, Huber W. 2010. Differential expression analysis for sequence count data. *Genome Biol* **11**:R106.

Barrow JR, Thomas KR, Boussadia-Zahui O, Moore R, Kemler R, Capecchi MR, McMahon AP. 2003. Ectodermal Wnt3/beta-catenin signaling is required for the establishment and maintenance of the apical ectodermal ridge. *Genes Dev* **17**:394–409.

Bell SM, Schreiner CM, Waclaw RR, Campbell K, Potter SS, Scott WJ. 2003. Sp8 is crucial for limb outgrowth and neuropore closure. *Proc Natl Acad Sci U S A* **100**:12195–12200.

Bell SM, Schreiner CM, Wert SE, Mucenski ML, Scott WJ, Whitsett JA. 2008. R-spondin 2 is required for normal laryngeal-tracheal, lung and limb morphogenesis. *Development* **135**:1049–1058.

Cartharius K, Frech K, Grote K, Klocke B, Haltmeier M, Klingenhoff A, Frisch M, Bayerlein M, Werner T. 2005. MatInspector and beyond: promoter analysis based on transcription factor binding sites. *Bioinformatics* **21**:2933–2942.

Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, Epstein CB, Zhang X, Wang L, Issner R, Coyne M, Ku M, Durham T, Kellis M, Bernstein BE. 2011. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473**:43–49. doi:10.1038/nature09906

Estella C, Mann RS. 2010. Non-redundant selector and growth-promoting functions of two sister genes, buttonhead and Sp1, in Drosophila leg development. *PLoS Genet* **6**:1–13. doi:10.1371/journal.pgen.1001001

Fernandez-Teran M, Ros MA. 2008. The Apical Ectodermal Ridge: Morphological aspects and signaling pathways. *Int J Dev Biol* **52**:857–871. doi:10.1387/ijdb.072416mf

Franch-Marro X, Martín N, Averof M, Casanova J. 2006. Association of tracheal placodes with leg primordia in Drosophila and implications for the origin of insect tracheal systems. *Development* **133**:785–790.

Haro E, Delgado I, Junco M, Yamada Y, Mansouri A, Oberg KC, Ros MA. 2014. Sp6

and Sp8 Transcription Factors Control AER Formation and Dorsal-Ventral Patterning in Limb Development. *PLoS Genet* **10**. doi:10.1371/journal.pgen.1004468

Hojo H, Ohba S, He X, Lai LP, McMahon AP. 2016. Sp7/Osterix Is Restricted to Bone-Forming Vertebrates where It Acts as a Dlx Co-factor in Osteoblast Specification. *Dev Cell* **37**:238–253. doi:10.1016/j.devcel.2016.04.002

Kawakami Y, Rodríguez-Esteban C, Matsui T, Rodríguez-León J, Kato S, Izpisúa Belmonte JC. 2004. Sp8 and Sp9, two closely related buttonhead-like transcription factors, regulate Fgf8 expression and limb outgrowth in vertebrate embryos. *Development* **131**:4763–4774. doi:10.1242/dev.01331

Kennedy MW, Chalamalasetty RB, Thomas S, Garriock RJ, Jailwala P, Yamaguchi TP. 2016. Sp5 and Sp8 recruit β-catenin and Tcf1-Lef1 to select enhancers to activate Wnt target gene transcription. *Proc Natl Acad Sci* **113**:3545–3550. doi:10.1073/pnas.1519994113

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* **14**:R36. doi:10.1186/gb-2013-14-4-r36

Kraus P, Lufkin T. 2006. Dlx Homeobox Gene Control of Mammalian Limb and Craniofacial Development. *Am J Hum Genet* **140**:1366–1374. doi:a

Lapan SW, Reddien PW. 2011. dlx and sp6-9 Control optic cup regeneration in a prototypic eye. *PLoS Genet* **7**:e1002226.

Lebensohn AM, Rohatgi R. 2018. R-spondins can potentiate WNT signaling without LGRs. *Elife* **7**.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**:1754–1760. doi:10.1093/bioinformatics/btp324

Lin C, Yin Y, Bell SM, Veith GM, Chen H, Huh SH, Ornitz DM, Ma L. 2013. Delineating a Conserved Genetic Cassette Promoting Outgrowth of Body Appendages. *PLoS Genet* **9**:1–12. doi:10.1371/journal.pgen.1003231

Liu T, Ortiz JA, Taing L, Meyer CA, Lee B, Zhang Y, Shin H, Wong SS, Ma J, Lei Y, Pape UJ, Poidinger M, Chen Y, Yeung K, Brown M, Turpaz Y, Liu XS. 2011. Cistrome: An integrative platform for transcriptional regulation studies. *Genome Biol* **12**:1–10. doi:10.1186/gb-2011-12-8-r83

Ma W, Noble WS, Bailey TL. 2014. Motif-based analysis of large nucleotide data sets using MEME-ChIP. *Nat Protoc* **9**:1428–1450.

Marinić M, Aktas T, Ruf S, Spitz F. 2013. An Integrated Holo-Enhancer Unit Defines Tissue and Gene Specificity of the Fgf8 Regulatory Landscape. *Dev Cell* **24**:530–542.

McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, Wenger AM, Bejerano G. 2010. GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol* **28**:495–501.

Pascal E, Tjian R. 1991. Different activation domains of Sp1 govern formation of multimers and mediate transcriptional synergism. *Genes Dev* **5**:1646–1656.

Presnell JS, Schnitzler CE, Browne WE. 2015. KLF/SP transcription factor family evolution: Expansion, diversification, and innovation in eukaryotes. *Genome Biol Evol* **7**:2289–2309. doi:10.1093/gbe/evv141

Ramírez F, Dündar F, Diehl S, Grüning BA, Manke T. 2014. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res* **42**:W187-91.

Rehimi R, Bartusel M, Solinas F, Altmüller J, Rada-Iglesias A. 2017. Chromatin Immunoprecipitation (ChIP) Protocol for Low-abundance Embryonic Samples. *J Vis Exp*.

Robledo RF, Rajan L, Li X, Lufkin T. 2002. The Dlx5 and Dlx6 homeobox genes are essential for craniofacial, axial, and appendicular skeletal development. *Genes Dev* **16**:1089–1101.

Sahara S, Kawakami Y, Izpisua Belmonte JC, O'Leary DDM. 2007. Sp8 exhibits reciprocal induction with Fgf8 but has an opposing effect on anterior-posterior cortical area patterning. *Neural Dev* **2**:10.

Satokata I, Maas R. 1994. Msx1 deficient mice exhibit cleft palate and abnormalities of craniofacial and tooth development. *Nat Genet* **6**:348–356.

Schmidl C, Rendeiro AF, Sheffield NC, Bock C. 2015. ChIPmentation: fast, robust, low-input ChIP-seq for histones and transcription factors. *Nat Methods* **12**:963–965.

Shen Y, Yue F, McCleary DF, Ye Z, Edsall L, Kuan S, Wagner U, Dixon J, Lee L, Lobanenkov V V, Ren B. 2012. A map of the cis-regulatory sequences in the mouse genome. *Nature* **488**:116–120.

Soshnikova N, Zechner D, Huelsken J, Mishina Y, Behringer RR, Taketo MM, Crenshaw EB, Birchmeier W. 2003. Genetic interaction between Wnt/beta-catenin and BMP receptor signaling during formation of the AER and the dorsal-ventral axis in the limb. *Genes Dev* **17**:1963–1968.

Suske G. 1999. The Sp-family of transcription factors. *Gene* **238**:291–300. doi:10.1016/S0378-1119(99)00357-1

Suske G, Bruford E, Philipsen S. 2005. Mammalian SP/KLF transcription factors: bring in the family. *Genomics* **85**:551–556.

Szenker-Ravi E, Altunoglu U, Leushacke M, Bosso-Lefèvre C, Khatoo M, Thi Tran H, Naert T, Noelanders R, Hajamohideen A, Beneteau C, de Sousa SB, Karaman B, Latypova X, Başaran S, Yücel EB, Tan TT, Vlaminck L, Nayak SS, Shukla A, Girisha KM, Le Caignec C, Soshnikova N, Uyguner ZO, Vleminckx K, Barker N, Kayserili H, Reversade B. 2018. RSPO2 inhibition of RNF43 and ZNRF3 governs limb development independently of LGR4/5/6. *Nature* **557**:564–569.

Tickle C. 2015. How the embryo makes a limb: Determination, polarity and identity. *J Anat* **227**:418–430. doi:10.1111/joa.12361

Trapnell C, Williams B a, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. 2010. Transcript assembly and abundance estimation from RNA-Seq reveals thousands of new transcripts and switching among isoforms. *Nat Biotechnol* **28**:511–515. doi:10.1038/nbt.1621.Transcript

Treichel D, Schöck F, Jäckle H, Gruss P, Mansouri A. 2003. mBtd is required to maintain signaling during murine limb development. *Genes Dev* **17**:2630–2635.

Wagle P, Nikolić M, Frommolt P. 2015. QuickNGS elevates Next-Generation Sequencing data analysis to a new level of automation. *BMC Genomics* **16**:487.

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nussbaum C, Myers RM, Brown M, Li W, Shirley XS. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**. doi:10.1186/gb-2008-9-9-r137

FIGURE 1

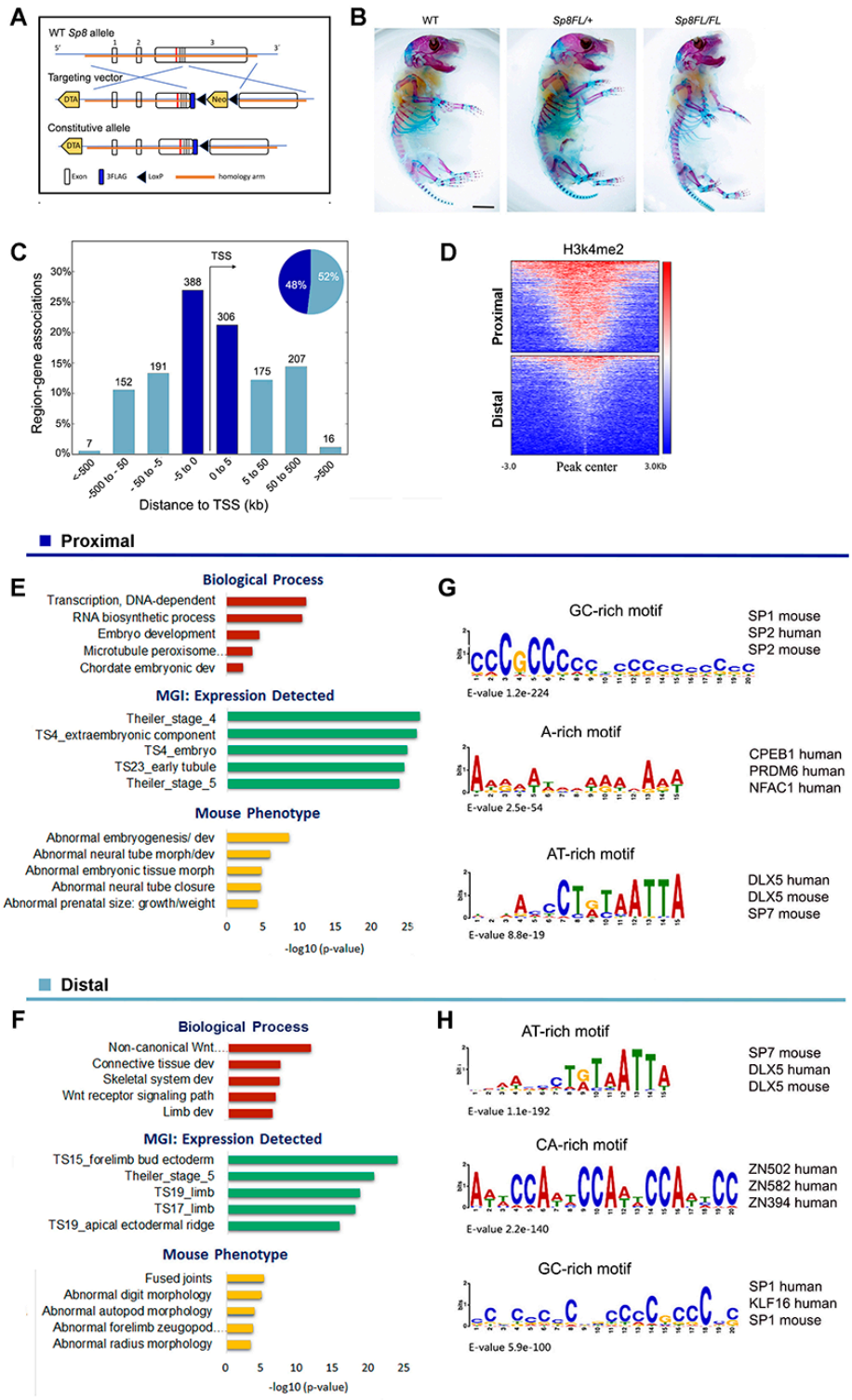

A
WT Sp8 allele
Targeting vector
Constitutive allele
DTA
Neo
Exon
3FLAG
LoxP
homology arm
B
WT
Sp8FL/+
Sp8FL/FL
C
TSS
Region-gene associations
48%
52%
7
152
191
388
306
175
207
16
<-500
-500 to -50
-50 to -5
-5 to 0
0 to 5
5 to 50
50 to 500
>500
Distance to TSS (kb)
D
H3k4me2
Proximal
Distal
-3.0
Peak center
3.0Kb
Proximal
E
Biological Process
Transcription, DNA-dependent
RNA biosynthetic process
Embryo development
Microtubule peroxisome...
Chordate embryonic dev
MGI: Expression Detected
Theiler_stage_4
TS4_extraembryonic component
TS4_embryo
TS23_early tubule
Theiler_stage_5
Mouse Phenotype
Abnormal embryogenesis/ dev
Abnormal neural tube morph/dev
Abnormal embryonic tissue morph
Abnormal neural tube closure
Abnormal prenatal size: growth/weight
0 5 10 15 20 25
-log10 (p-value)
G
GC-rich motif
SP1 mouse
SP2 human
SP2 mouse
E-value 1.2e-224
A-rich motif
CPEB1 human
PRDM6 human
NFAC1 human
E-value 2.5e-54
AT-rich motif
DLX5 human
DLX5 mouse
SP7 mouse
E-value 8.8e-19
Distal
F
Biological Process
Non-canonical Wnt...
Connective tissue dev
Skeletal system dev
Wnt receptor signaling path
Limb dev
MGI: Expression Detected
TS15_forelimb bud ectoderm
Theiler_stage_5
TS19_limb
TS17_limb
TS19_apical ectodermal ridge
Mouse Phenotype
Fused joints
Abnormal digit morphology
Abnormal autopod morphology
Abnormal forelimb zeugopod...
Abnormal radius morphology
0 5 10 15 20 25
-log10 (p-value)
H
AT-rich motif
SP7 mouse
DLX5 human
DLX5 mouse
E-value 1.1e-192
CA-rich motif
ZN502 human
ZN582 human
ZN394 human
E-value 2.2e-140
GC-rich motif
SP1 human
KLF16 human
SP1 mouse
E-value 5.9e-100

FIGURE 2

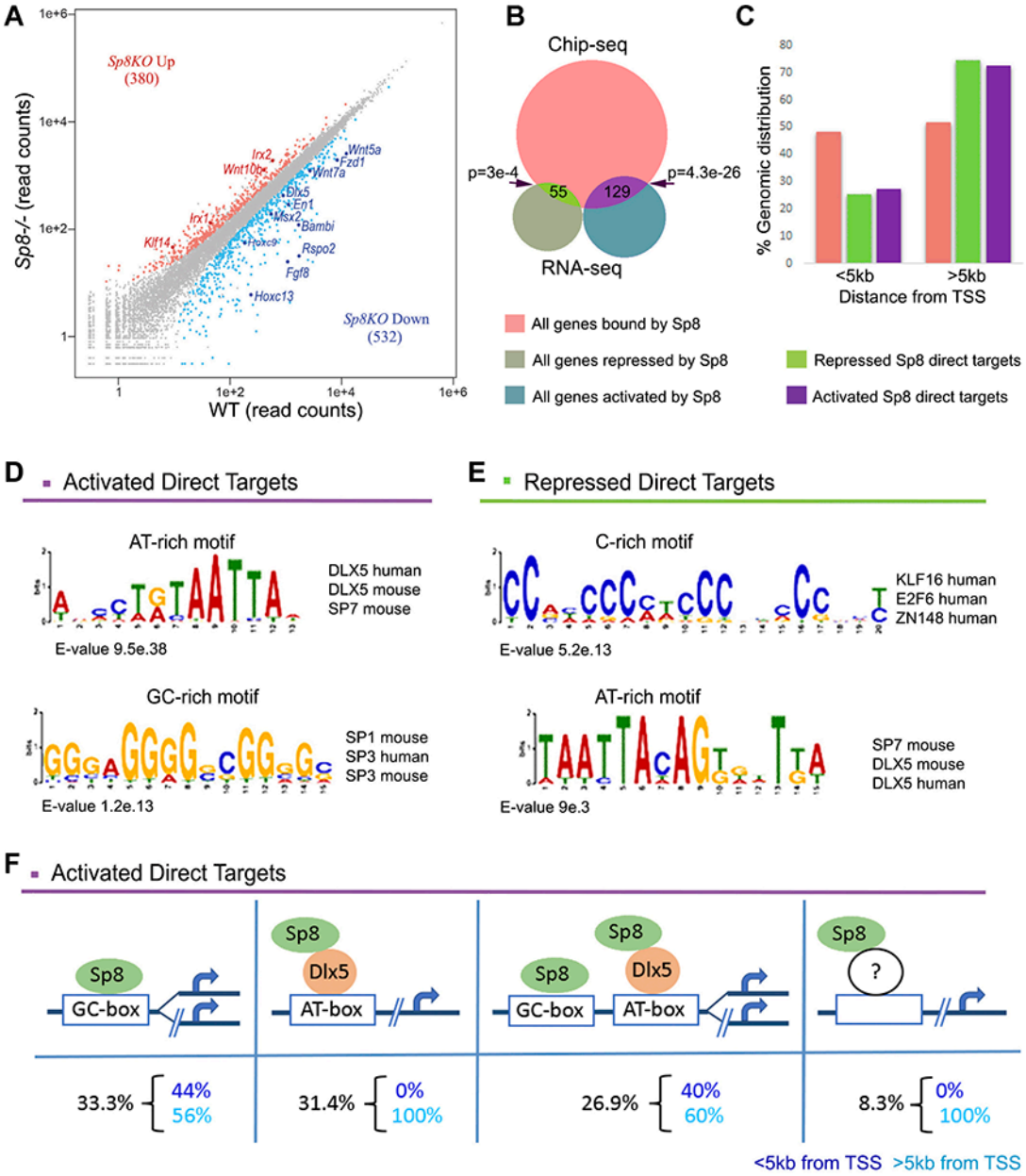

A
Sp8KO Up
(380)
Sp8KO Down
(532)
Sp8-/- (read counts)
WT (read counts)
Irx2
Wnt10b
Irx1
Klf14
Wnt5a
Fzd1
Wnt7a
Dlx5
En1
Msx2
Bambi
Hoxc9
Rspo2
Fgf8
Hoxc13
B
Chip-seq
RNA-seq
p=3e-4
p=4.3e-26
55
129
All genes bound by Sp8
All genes repressed by Sp8
All genes activated by Sp8
C
% Genomic distribution
<5kb
>5kb
Distance from TSS
Repressed Sp8 direct targets
Activated Sp8 direct targets
D
Activated Direct Targets
AT-rich motif
DLX5 human
DLX5 mouse
SP7 mouse
E-value 9.5e.38
GC-rich motif
SP1 mouse
SP3 human
SP3 mouse
E-value 1.2e.13
E
Repressed Direct Targets
C-rich motif
KLF16 human
E2F6 human
ZN148 human
E-value 5.2e.13
AT-rich motif
SP7 mouse
DLX5 mouse
DLX5 human
E-value 9e.3
F
Activated Direct Targets
Sp8
GC-box
Dlx5
AT-box
?
33.3%
44%
56%
31.4%
0%
100%
26.9%
40%
60%
8.3%
0%
100%
<5kb from TSS
>5kb from TSS

FIGURE 3

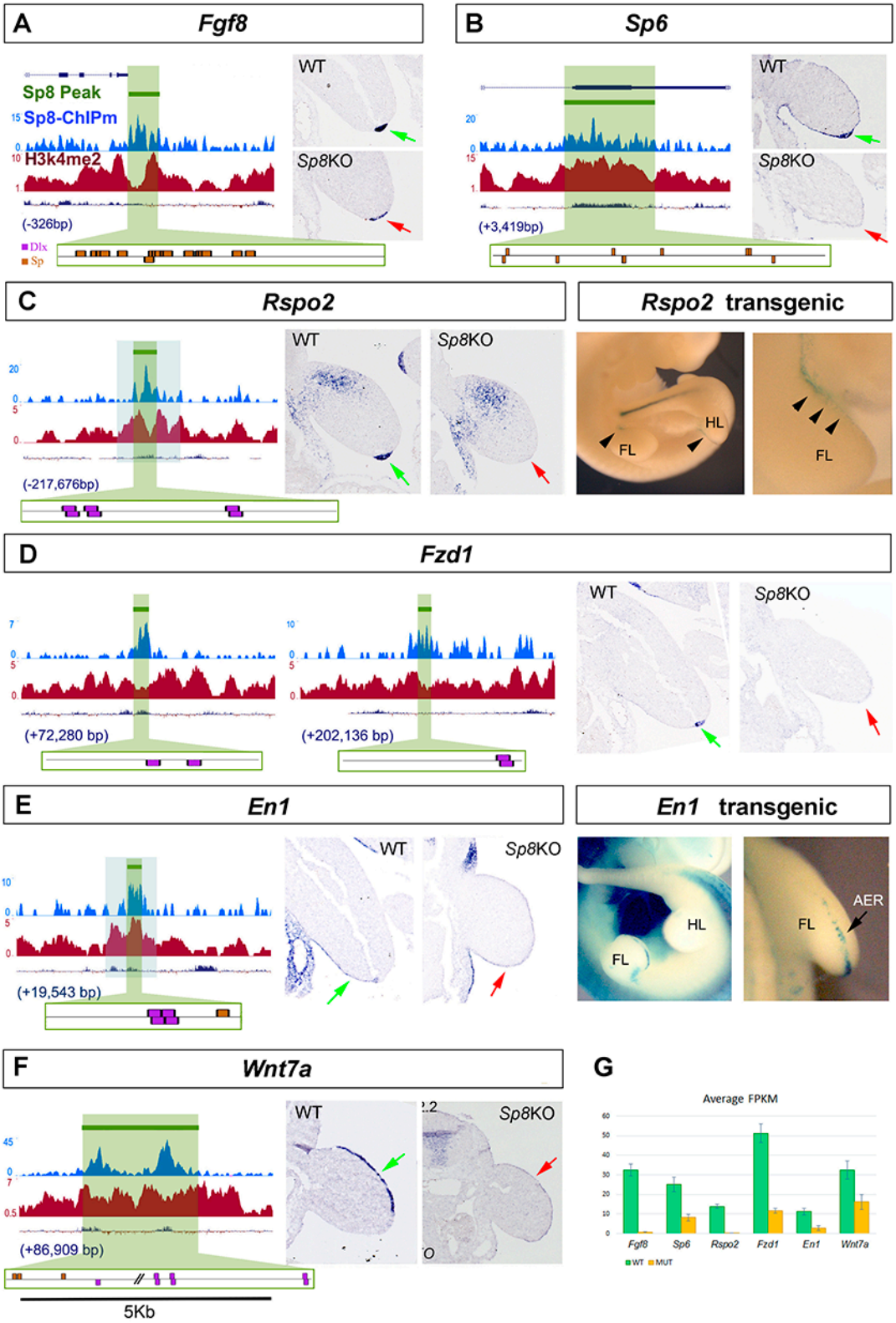
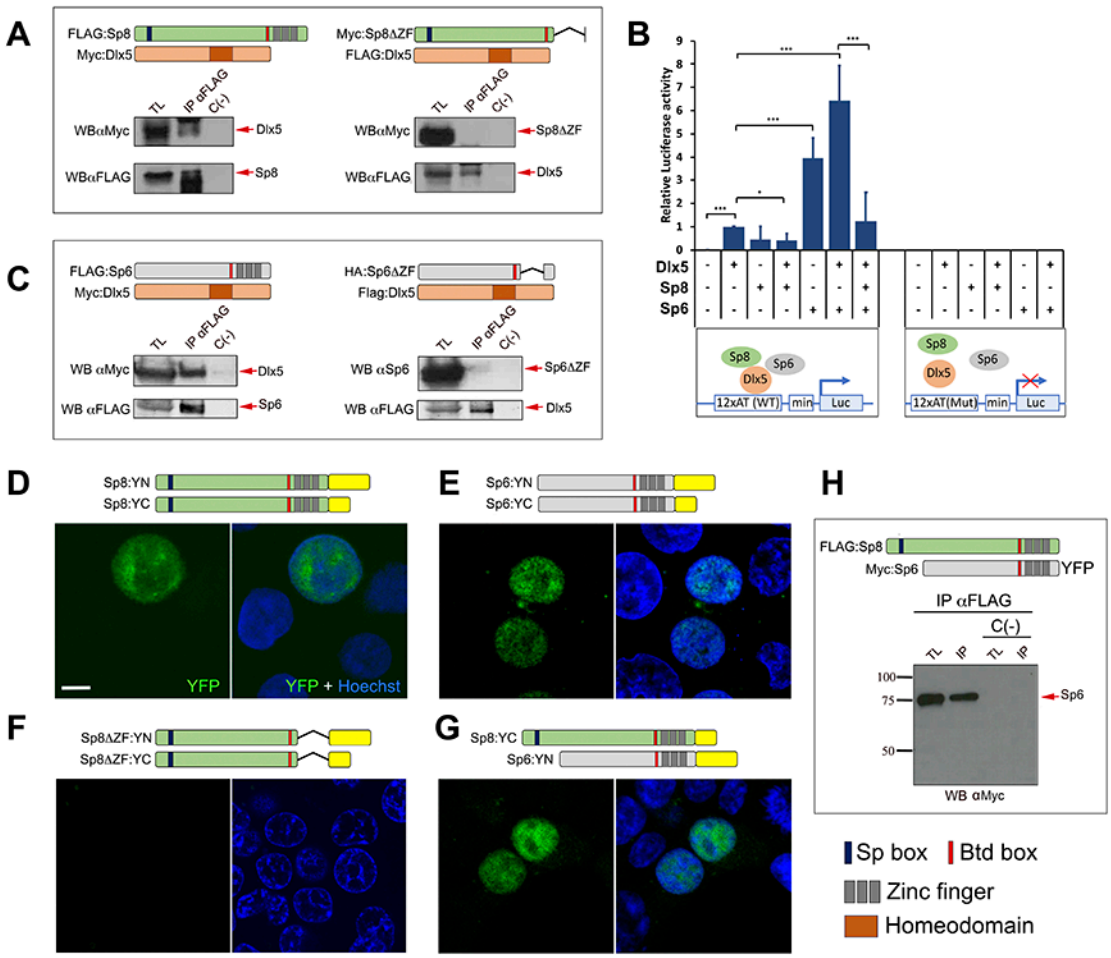
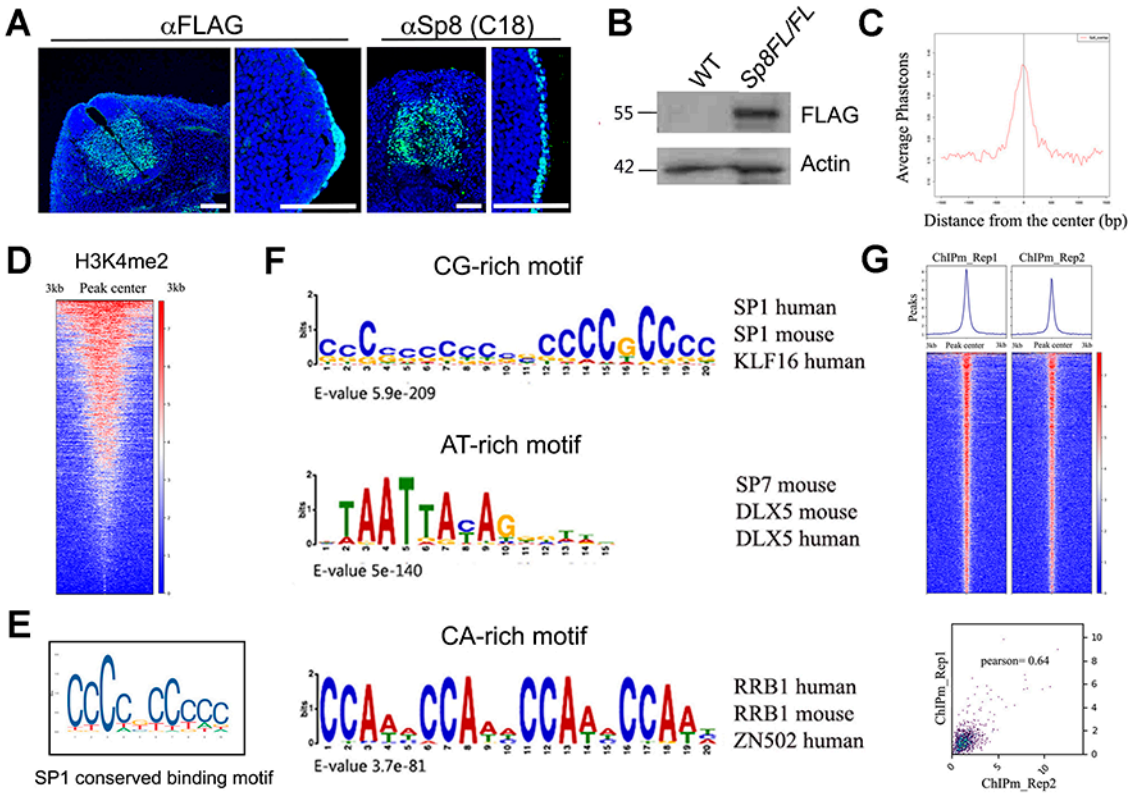

A
Fgf8
Sp8 Peak
Sp8-ChIPm
H3k4me2
(-326bp)
Dlx
Sp
WT
Sp8KO
B
Sp6
(+3,419bp)
WT
Sp8KO
C
Rspo2
(-217,676bp)
WT
Sp8KO
Rspo2 transgenic
HL
FL
FL
D
Fzd1
(+72,280 bp)
(+202,136 bp)
WT
Sp8KO
E
En1
(+19,543 bp)
WT
Sp8KO
En1 transgenic
HL
FL
FL
AER
F
Wnt7a
(+86,909 bp)
WT
Sp8KO
5Kb
G
Average FPKM
Fgf8
Sp6
Rspo2
Fzd1
En1
Wnt7a
WT
MUT

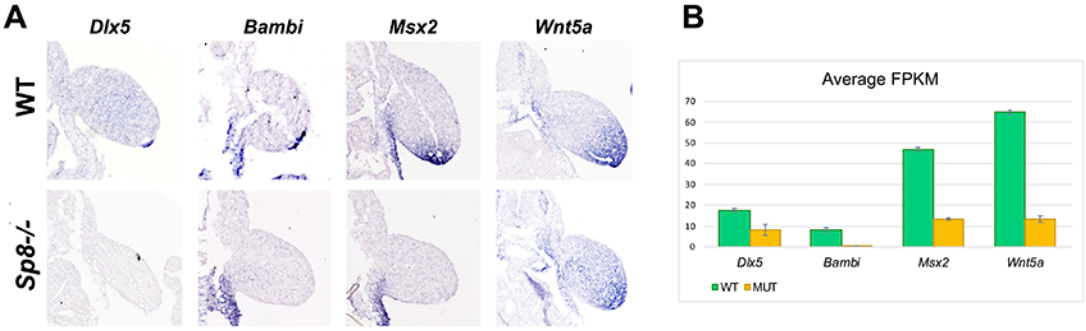
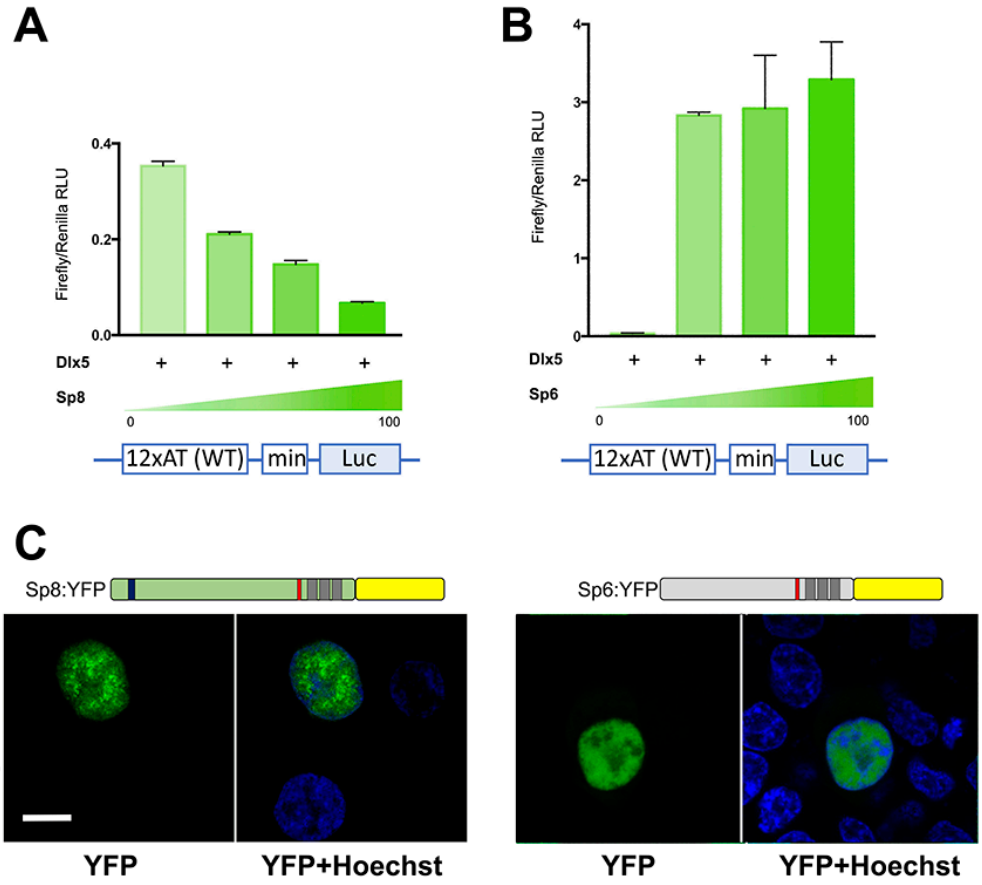
FIGURE 4


A
FLAG:Sp8
Myc:Dlx5
Myc:Sp8ΔZF
FLAG:Dlx5
TL
IP αFLAG
C(-)
WBαMyc
WBαFLAG
Dlx5
Sp8
Sp8ΔZF
B
Relative Luciferase activity
Dlx5
Sp8
Sp6
12xAT (WT)
min
Luc
12xAT(Mut)
C
FLAG:Sp6
Myc:Dlx5
HA:Sp6ΔZF
Flag:Dlx5
WB αMyc
WB αFLAG
WB αSp6
Sp6
Sp6ΔZF
D
Sp8:YN
Sp8:YC
YFP
YFP + Hoechst
E
Sp6:YN
Sp6:YC
F
Sp8ΔZF:YN
Sp8ΔZF:YC
G
Sp8:YC
Sp6:YN
H
FLAG:Sp8
Myc:Sp6
YFP
IP αFLAG
C(-)
TL
IP
100
75
50
Sp6
WB αMyc
Sp box
Btd box
Zinc finger
Homeodomain

SUPPLEMENTARY FIGURE 1



SUPPLEMENTARY FIGURE 2



SUPPLEMENTARY FIGURE 3