© 2021. Published by The Company of Biologists Ltd | Development (2021) 148, dev197202. doi:10.1242/dev.197202

RESEARCH ARTICLE

Sox2-*Evf2* lncRNA-mediated mechanisms of chromosome topological control in developing forebrain

Ivelisse Cajigas[1], Abhijit Chakraborty[2], Madison Lynam[1], Kelsey R. Swyter[1], Monique Bastidas[1], Linden Collens[1], Hao Luo[1], Ferhat Ay[2,3] and Jhumku D. Kohtz[1,*]

ABSTRACT

The *Evf2* long non-coding RNA directs *Dlx5/6* ultraconserved enhancer(*UCE*)-intrachromosomal interactions, regulating genes across a 27 Mb region on chromosome 6 in mouse developing forebrain. Here, we show that *Evf2* long-range gene repression occurs through multi-step mechanisms involving the transcription factor Sox2. *Evf2* directly interacts with Sox2, antagonizing Sox2 activation of *Dlx5/6UCE*, and recruits Sox2 to the *Dlx5/6eii* shadow enhancer and key *Dlx5/6UCE* interaction sites. Sox2 directly interacts with Dlx1 and Smarca4, as part of the *Evf2* ribonucleoprotein complex, forming spherical subnuclear domains (protein pools, PPs). *Evf2* targets Sox2 PPs to one long-range repressed target gene (*Rbm28*), at the expense of another (*Akr1b8*). *Evf2* and Sox2 shift *Dlx5/6UCE* interactions towards *Rbm28*, linking *Evf2*/Sox2 co-regulated topological control and gene repression. We propose a model that distinguishes *Evf2* gene repression mechanisms at Rbm28 (*Dlx5/6UCE* position) and Akr1b8 (limited Sox2 availability). Genome-wide control of RNPs (Sox2, Dlx and Smarca4) shows that co-recruitment influences Sox2 DNA binding. Together, these data suggest that *Evf2* organizes a Sox2 PP subnuclear domain and, through Sox2-RNP sequestration and recruitment, regulates chromosome 6 long-range *UCE* targeting and activity with genome-wide consequences.

KEY WORDS: lncRNA, Chromosome 3D structure, Architectural proteins, Enhancers, Epigenetics, Forebrain development, Ribonucleoprotein complex, Transcription factor binding, Gene repression

INTRODUCTION

Ultraconserved elements (UCEs) were identified as 200 bp (or greater) segments of 100% DNA conservation between humans, mice and rats, many associated with key developmental regulators (Bejerano et al., 2004; Sandelin et al., 2004; Woolfe et al., 2005). Removal of a select few UCEs in mice initially suggested that UCEs are dispensable (Ahituv et al., 2007). However, removal of UCE sequences near developmental regulators Arx, Gli and Shox2 causes neurological and growth defects (Dickel et al., 2018; Osterwalder et al., 2018), and limb defects (Nolte et al., 2014), revealing specific developmental roles. Transcription of UCE sequences and enhancer-regulating activity of UCE transcripts (Calin et al., 2007; Feng et al., 2006) was followed by the identification of genome-wide scale enhancer transcripts with enhancer-like activities (Ørom et al., 2010; Ørom and Shiekhattar, 2011). Together, these data support mechanistic and functional diversity of RNA regulatory roles (Rinn and Chang, 2020).

Our studies on the *Evf2* ultraconserved enhancer lncRNA (*Dlx5/6UCE*-lncRNA, overlapping with *Dlx6OS1*) support complex RNA regulatory roles for UCE sequences during embryonic forebrain development, specifically at sites of GABAergic interneuron birth in E13.5 mouse ganglionic eminences (E13.5 GEs) (Berghoff et al., 2013; Bond et al., 2009; Cajigas et al., 2018, 2015; Feng et al., 2006). *Evf2* is a 3.7 kb, spliced and polyadenylated lncRNA, containing *Dlx5/6UCE* sequences responsible for enhancer-regulating activities (Feng et al., 2006). *Evf2* controls a mouse embryonic brain interneuron gene regulatory network (GRN), adult hippocampal and cortical circuitry, and seizure susceptibility (Bond et al., 2009; Cajigas et al., 2018; Feng et al., 2006). *Evf2* positively and negatively regulates gene expression through *cis* (same chromosome as *Evf2* expression site) and *trans* (different chromosome) mechanisms (Berghoff et al., 2013; Cajigas et al., 2018). Mechanisms of *Evf2* gene activation and repression are distinguished by different regional requirements of the RNA, with the 5′-UCE containing region of the lncRNA controlling gene repression and the 3′ end controlling gene activation (Cajigas et al., 2018).

Evf2 RNA cloud formation is similar to lncRNAs that regulate dosage compensation (*Xist*; Brockdorff et al., 1992; Brown et al., 1992) and imprinting (*Kcnq1ot1*; Pandey et al., 2008; Redrup et al., 2009). *Evf2* assembles a ribonucleoprotein complex (RNP^{87}) containing at least 87 functionally diverse proteins, including transcription factors (TFs; Dlx1 and Sox2), chromatin remodelers (Smarca4, Smarcc2 and Smarcb1), regulators of chromosome topology (Smc1a and Smc3) and lamin B1 (Cajigas et al., 2015) (Fig. 1A). RNP^{87} was previously identified by comparing proteomic profiles from anti-DLX, affinity-purified $Evf2^{+/+}$ and $Evf2^{TS/TS}$ E13.5GE complexes, showing that the number of Dlx-associated proteins is 87 in the presence of *Evf2*, and 15 in the absence of *Evf2*. A description of the site of transcription stop (TS) insertion that generates mice lacking *Evf2* ($Evf2^{TS/TS}$) is shown in Fig. S1. *Evf2*-dependent gene regulation across a 27 Mb region of mouse chromosome 6 (chr6) is characterized by recruitment of individual RNPs and regulation of histone modifications at key DNA regulatory sites, including the *Dlx5/6UCE* (Cajigas et al., 2018).

The identification of the TF Sox2, as a component of *Evf2*-RNP^{87} (Cajigas et al., 2015), raised questions about its role in this lncRNA-mediated gene regulation. Sox2 is a well-characterized

[1]Department of Pediatrics, Northwestern University, Feinberg School of Medicine, Department of Human Molecular Genetics, Stanley Manne Children's Research Institute 2430 N Halsted, Chicago, IL 60614, USA. [2]Centers for Autoimmunity and Cancer Immunotherapy, La Jolla Institute for Immunology, 9420 Athena Circle, La Jolla, CA, 92037, USA. [3]School of Medicine, University of California San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA.

*Author for correspondence (j-kohtz@northwestern.edu)

M.B., 0000-0003-1562-9576; L.C., 0000-0003-1897-6629; F.A., 0000-0002-0708-6914; J.D.K., 0000-0002-3483-5140

Handling Editor: Haruhiko Koseki
Received 22 September 2020; Accepted 7 February 2021

Development (2021) 148, dev197202. doi:10.1242/dev.197202

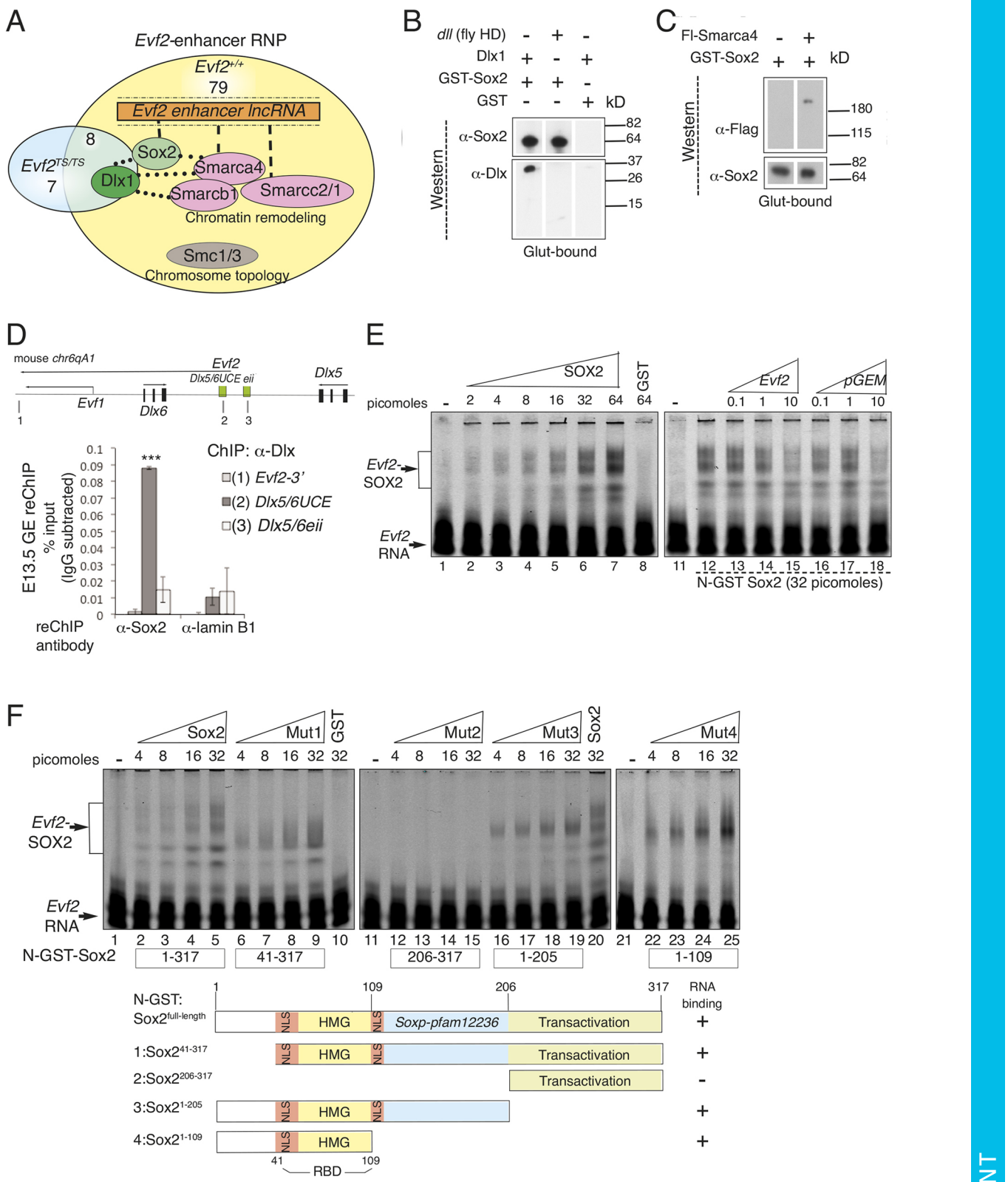


Fig. 1. *Evf2* repression through Sox2 antagonism in mouse developing forebrain. (A) The *Evf2* UCE lncRNA and Sox2 protein are RNA/protein scaffolds for multiple RNPs, including chromatin remodelers (Smarca4/c2/c1) and cohesin (Smc1a/3) in E13.5 mouse ganglionic eminences (GEs). Of the 87 proteins in the Evf2-Dlx RNP complex, Sox2 is one of 79 proteins complexed with Dlx1 in the presence of *Evf2* (*Evf2*$^{+/+}$), and not detected in GEs lacking *Evf2* (*Evf2*$^{TS/TS}$). (B) Western analysis shows direct binding of GST-Sox2 to mouse Dlx1, but not to GST or Dll (fly homeodomain fragment). (C) Western analysis shows that GST-Sox2 directly binds flag-tagged Smarca4. (D) Dlx-Sox2 complexes are detected at the *Dlx5/6UCE* by ChIP-reChIP (first anti-Dlx, second anti-Sox2) from GE crosslinked chromatin. (E,F) RNA electrophoretic mobility shift assays (REMSAs) using infrared-labelled *Evf2* RNA as a probe. (E) Sox2 binding to *Evf2* RNA at picomolar concentrations is promiscuous, as indicated by pGEM RNA competition with *Evf2* RNA for binding to Sox2. (F) The Sox2 RNA-binding domain is narrowed to the Sox2 high mobility group domain (HMG) and adjacent nuclear localization (NLS) regions.

Development (2021) 148, dev197202. doi:10.1242/dev.197202

pioneer TF (Dodonova et al., 2020) that maintains pluripotency through lineage-specific gene repression (Avilion et al., 2003; Takahashi and Yamanaka, 2006). Sox2 associates with lncRNAs involved in pluripotency and neuronal differentiation (Guo et al., 2018; Ng et al., 2013, 2012), and binds both DNA and RNA through its high mobility group domain (HMG) (Holmes et al., 2020). Crystal structures of HMG-POU-DNA ternary complexes support Sox2 multivalency and concentration-dependent enhancer regulation (Remenyi et al., 2003, 2004; Williams et al., 2004). In this report, we show that *Evf2* regulates *Dlx5/6UCE* targeting and activity through mechanisms involving the *Evf2*-RNP Sox2 complex, revealing multi-step contributions of Sox2 TF-RNA interactions. Sox2 colocalizes with *Evf2* RNA clouds in subnuclear domains that we have termed protein pools (PPs), detectable both in the presence and absence of *Evf2* RNA. *Evf2* controls Sox2 PP targeting and sizes at repressed genes *Rbm28* and *Akr1b8*, and recruits Sox2 to key DNA regulatory sites, including *Dlx5/6* intergenic enhancers and enhancer-chromosome interaction sites. At the genome-wide level, *Evf2* co-recruitment of Sox2 with the RNPs Smarca4 and Dlx affects Sox2-DNA recognition. We propose that the *Evf2* lncRNA functions as a Sox2 subnuclear domain organizer, controlling *Dlx5/6UCE* targeting and activity by distributing Sox2 and the associated RNPs Smarca4 and Dlx to key DNA regulatory sites on chr6, with genome-wide effects.

RESULTS

Evf2 gene repression through Sox2 antagonism

Evf2 activates and represses genes across a 27 Mb region on mouse chromosome 6, raising questions regarding the mechanistic basis for *Evf2*-dependent differential gene regulation (Cajigas et al., 2018). The *Evf2* RNA cloud is a scaffold for the assembly of the *Evf2*-RNP (Fig. 1A) (Cajigas et al., 2015). *Evf2* directly binds chromatin remodelers Smarca4 and Smarcc2/1 through promiscuous RNA-protein interactions, and indirectly to the Dlx homeodomain TF. Previous work showed that Smarca4 bridges the *Evf2* RNA with the protein Dlx1, and other RNA-binding proteins within the RNP (Cajigas et al., 2015). In order to investigate the role of individual *Evf2-RNP*[87] proteins in gene regulation, we further studied the role of the pioneer TF Sox2 in *Evf2*-regulated gene expression. First, we studied Sox2 interactions with *Evf2*-RNP components. In the absence of *Evf2*, there is a ∼25% decrease in total Sox2 protein levels (Fig. S1A,B) and a ∼50% increase in Sox2 RNA (Fig. S1C), supporting the involvement of both transcriptional and post-transcriptional control mechanisms. Sox2 directly binds Dlx1 (Fig. 1B) and Smarca4 (Fig. 1C), supporting multiple protein partners within the *Evf2*-RNP. We next used ChIP-reChIP to show that Sox2 and Dlx simultaneously bind *Dlx5/6UCE* (Fig. 1D). RNA electrophoretic mobility shift assays (REMSAs) show that *Evf2* RNA binding to Sox2 has low sequence specificity (Fig. 1E), requiring Sox2 amino acids 41-109 [containing the high mobility group (HMG) DNA-binding domain and N-terminal nuclear localization signal (NLS); Fig. 1F]. These data are consistent with a recent report showing the requirement for the HMG domain in high-affinity/low-specificity Sox2-*ES2* lncRNA interactions (Holmes et al., 2020). Together, these data support that Sox2 is similar to Smarca4, forming multivalent interactions and potentially functioning as a protein bridge between non-RNA binding proteins in the *Evf2-RNP* and *Evf2* RNA (Fig. 1A).

In a previous report, we showed that *Evf2* represses adjacent genes *Dlx6* and *Dlx5*, and long-range target genes *Rbm28* and *Akr1b8*, and activates long-range target genes *Umad1* and *Lsm8* on mouse chr6 (Cajigas et al., 2018). In order to determine whether Sox2 contributes to *Evf2*-dependent gene activation or repression, we analyzed gene expression in E13.5 GEs (mouse embryonic GABAergic interneuron progenitors) from $Sox2^{fl/fl}$;*Dlx5/6cre*+, a genetic model in which floxed *Sox2* (Shaham et al., 2009) removes Sox2 from *Dlx5/6*+ GABAergic progenitors (Monory et al., 2006). *Dlx5/6cre*-mediated removal of Sox2 in E13.5GE decreases the expression of *Evf2*-repressed target genes (*Dlx6*, *Dlx5*, *Rbm28* and *Akr1b8*), but does not affect *Evf2*-activated target gene expression (*Umad1* and *Lsm8*) (Fig. 2A). Loss of one copy of *Evf2* from $Sox2^{fl/fl}$;*Dlx5/6cre*+ E13.5 GEs ($Evf2^{TS/+}$;$Sox2^{fl/fl}$:*Dlx5/6cre*) rescues the effects of Sox2 loss on repressed target genes (Fig. 2A). *Evf2* transcripts resulting from $Evf2^{TS}$ insertion have been previously reported (Bond et al., 2009) and are schematized in Fig. S1A. Sox2 expression in $Sox2^{fl/fl}$;*Dlx5/6cre*+ $Sox2^{+}$/Dlx5/6^{-} subpopulations persists at ∼40% of wild-type levels (Fig. 2A). Despite heterogeneity, we are able to detect gene expression effects. Furthermore, *Dlx5/6UCE*-luciferase reporter assays show that *Evf2* antagonizes Sox2 activation of *Dlx5/6UCE* activity in E13.5 GEs (Fig. 2B), supporting a mechanism of *Evf2*-Sox2 antagonism during gene repression.

Evf2-5′ end-mediated regulation of Sox2 binding to the *Dlx5/6eii* shadow enhancer

In order to determine whether *Evf2*/*Sox2*-mediated regulation of *Dlx5/6UCE* activity involves Sox2 recruitment to the *Dlx5/6* intergenic enhancers, we used the native ChIPseq method CUT&RUN (Fig. 2C) (Meers et al., 2019a,c; Skene and Henikoff, 2017). In the CUT&RUN method, sequencing of <120 bp and >150 bp fragments distinguishes between proteins directly bound to DNA (less than 120 bp) and indirect binding through protein-protein interactions (more than 150 bp) (Meers et al., 2019a,b,c). Analysis of >150 bp CUT&RUN peaks has the potential to detect proteins associated with the large *Evf2-RNP*, which was not previously possible using crosslinked ChIPseq (X-ChIP) methods. CUT&RUN analysis shows that whereas *Evf2* recruits *Evf2*-RNPs Dlx and Smarca4 to *Dlx5/6UCE*, Sox2 binding to Dlx5/6UCE is *Evf2* independent (Fig. 2C). However, *Evf2* recruits Sox2 to *Dlx5/6eii*, a shadow enhancer (Furlong and Levine, 2018; Zerucha et al., 2000) located adjacent to *Dlx5/6UCE* and regulated by both Dlx and *Evf2* in *trans* assays, similarly to *Dlx5/6UCE* (Feng et al., 2006; Zerucha et al., 2000). The significance of *Evf2*-mediated Sox2-*Dlx5/6eii* recruitment with respect to gene repression is supported by rescue in $Evf1^{TS/TS}$, a genetic model in which *Evf2* repression is also rescued (Cajigas et al., 2018) (Fig. 2C). In $Evf1^{TS/TS}$, the transcription stop sequence is inserted into exon 3, preventing expression of *Evf1* (and also the *Evf2*-3′ region), but producing a truncated *Evf2*-5′ transcript (Fig. S1A). Loss of Dlx and Smarca4 binding to *Dlx5/6UCE* in $Evf2^{TS/TS}$ (expressing only *Evf1*-3′, overlapping with the *Evf2*-3′ region) is not rescued in $Evf1^{TS/TS}$ (expressing only *Evf2*-5′); transcripts resulting from $Evf1^{TS}$ and $Evf2^{TS}$ insertion are schematized in Fig. 2D and Fig. S1A. These data suggest that *Evf2*-5′ and -3′ regions are required for Dlx and Smarca4 recruitment to *Dlx5/6UCE*, linking these events to gene activation. Furthermore, a role for *Evf2*-5′-regulated Sox2 binding to the *Dlx5/6eii* shadow enhancer supports a functional role for Sox2 recruitment in gene repression, building on previous work showing that the *Evf2*-5′ region is sufficient for repression (Cajigas et al., 2018). Thus, *Evf2*-5′ and 3′ differentially contribute to recruitment in a site-specific and RNP-dependent manner, linking individual recruitment events to gene repression and activation.

Development (2021) 148, dev197202. doi:10.1242/dev.197202

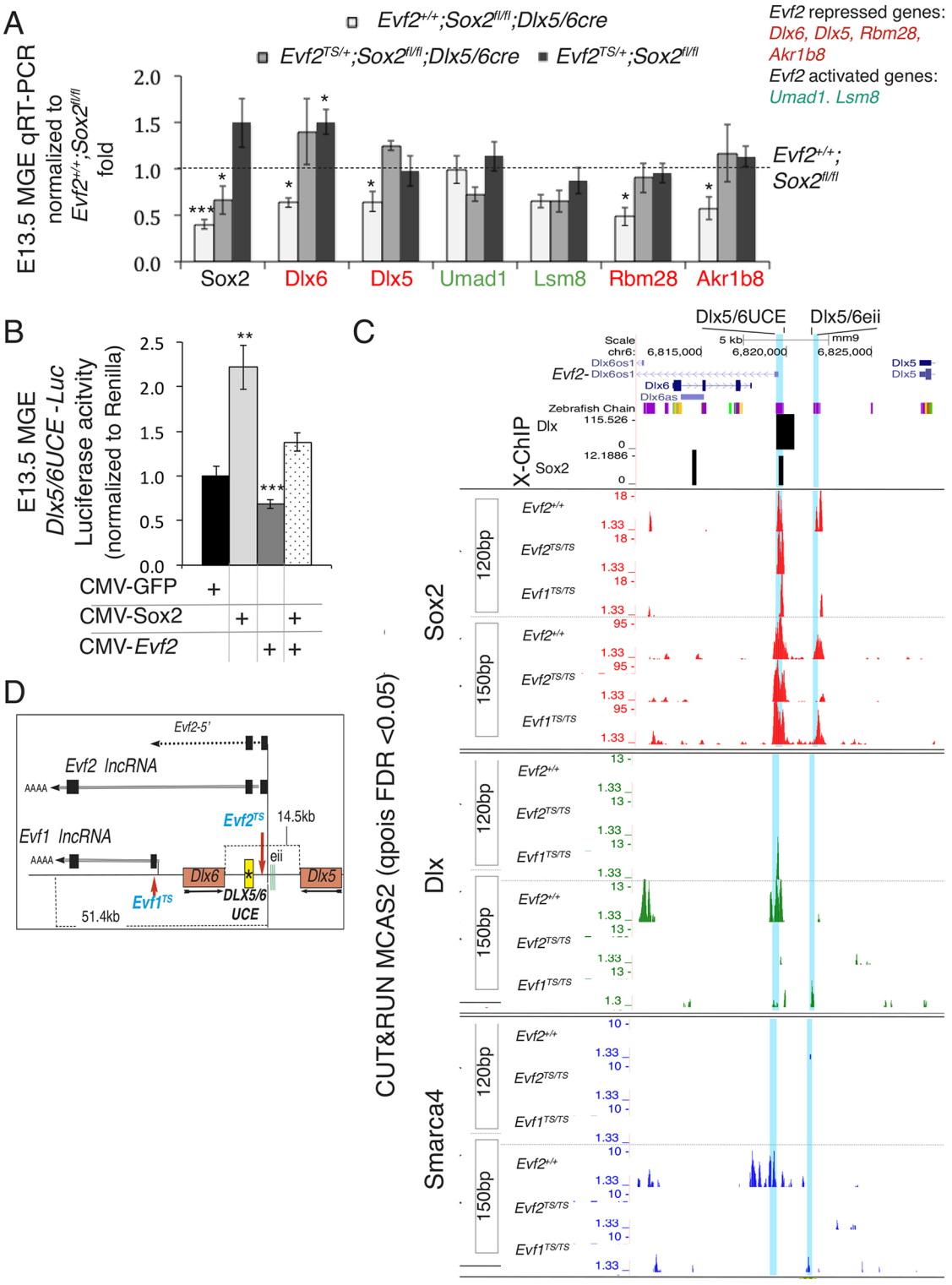


Fig. 2. *Evf2* gene repression through Sox2 antagonism. (A) TAQman qRT-PCR analysis of E13.5 medial ganglionic eminences (MGEs) from mice lacking Sox2 in interneuron progenitors (*Evf2*$^{+/+}$;*Sox2*$^{fl/fl}$;*Dlx5/6cre*). Additional loss of one copy of *Evf2*$^{TS/+}$;*Sox2*$^{fl/fl}$;*Dlx5/6cre* and loss of one copy of *Evf2* (*Evf2*$^{TS/+}$; *Sox2*$^{fl/fl}$) are normalized to wild type (*Evf2*$^{+/+}$;*Sox2*$^{fl/fl}$). Sox2 loss reduces *Evf2*-repressed target genes (red; *Dlx6, Dlx5*, *Rbm28* and *Akr1b8*) (white bars), which is rescued by loss of one copy of *Evf2* (gray bars). *n*=3-6/genotype. (B) Luciferase assays of E13.5 MGEs using a *Dlx5/6UCE*-luciferase reporter (*Dlx5/6UCE*-Luc) shows that Sox2-mediated activation of *Dlx5/6UCE* is inhibited by *Evf2* repressor activity. *n*=6/condition (each experiment replicated three times). (C) CUT&RUN native ChIPseq binding profiles of *Evf2*-RNPs (Sox2, Dlx and Smarca4) in *Evf2*$^{+/+}$, *Evf2*$^{TS/TS}$ and *Evf1*$^{TS/TS}$ GEs. *Dlx5/6* intergenic enhancers (*Dlx5/6UCE* and *Dlx5/6eii* shadow enhancer) are highlighted in pale blue. 120 bp and 150 bp profiles show MACS2-validated peaks (FDR<0.05), *n*=2-4/genotype [MACS2 is a peak-calling method for ChIPseq (Zhang et al., 2008)]. Comparisons between *Evf2* loss (*Evf2*$^{TS/TS}$) and truncation (*Evf1*$^{TS/TS}$) mutants identify *Evf2* 5′ versus 3′ Sox2 differentially regulated sites. Student's *t*-test, **P*<0.05, ***P*<0.01, ****P*<0.001. Data are mean±s.e.m. (D) On mouse chr6, the sites of Evf1TS and Evf2TS transcription stop insertions (TS, pale blue) are shown with respect to repressed genes *Dlx5* and *Dlx6* (orange boxes), and Evf1, Evf2 and Evf2-5′ transcripts (exons 1-4 in black). Dlx5/6UCE is indicated by a yellow rectangle and star.

Development (2021) 148, dev197202. doi:10.1242/dev.197202

Linking *Evf2* and Sox2-regulated *Dlx5/6*UCE-chr6 targeting, and RNP recruitment

Previous work showed that *Evf2* regulated *Dlx5/6UCE* targeting near long-range gene targets involves cohesin binding, specifically Smc1a and Smc3 (Cajigas et al., 2018), raising the possibility of roles of additional *Evf2*-RNPs in topological control. We performed ChIPseq using crosslinked E13.5GE chromatin to analyze *Evf2*-regulated Sox2 and Smc3 binding across chr6 (Fig. S1E). Analysis of overlapping regulatory sites identifies antagonistic sites of *Evf2* positively (+) regulated Sox2 binding and *Evf2* negatively (−) regulated Smc3 binding (Fig. S1E). In order to explore the possibility that Sox2 contributes to *Dlx5/6UCE* targeting, we used chromosome conformation capture (4Cseq) to compare *Dlx5/6UCE* interaction (*Dlx5/6UCEin*) profiles across chr6 in the presence ($Sox2^{fl/fl}$;*Dlx5/6*cre-) and absence ($Sox2^{fl/fl}$;*Dlx5/6*cre+) of Sox2 in E13.5GE GABAergic progenitors. A subset of *Evf2*-regulated *Dlx5/6UCEins* across chr6 originally reported by Cajigas et al. (2018) overlaps with Sox2 regulated *Dlx5/6UCEins* (Fig. 3), including the Rbm28-5′ *Dlx5/5UCEins* (Fig. 3A). CUT&RUN analysis compares *Evf2*-5′ (rescued in $Evf1^{TS/TS}$) and -3′ (lost in $Evf2^{TS/TS}$, not rescued in $Evf1^{TS/TS}$) -mediated recruitment of Sox2 and Smarca4 at *Rbm28*-5′ and *Akr1b8*-3′ *Dlx5/6UCEins* (Fig. 3A). The complete Sox2- and *Evf2*-regulated 4Cseq-*Dlx5/6UCEin* counts across chr6 are shown in Fig. S2. Complete CUT&RUN profiles for *Rbm28*-5′ are shown in Fig. S3A.

Sox2 binding at the *Rbm28*-5′-*Dlx5/6UCEin* is rescued in $Evf1^{TS/TS}$, linking Sox2 recruitment to gene repression. In contrast, Sox2 does not regulate *Dlx5/6UCE* interaction 3′ of the long-range repressed target gene *Akr1b8* [Fig. 3A, green arrow at *Akr1b10* (*Akr1b8*-3′-*Dlx5/6UCEin*), complete profiles shown in Fig. S3B]. In contrast to the *Rbm28*-5′-*Dlx5/6UCEin*, Sox2 binding to the *Akr1b8*-3′-*Dlx5/6UCEin* is not detected in $Evf2^{+/+}$ or $Evf2^{TS/TS}$, and a small peak is detected only when *Evf2* is truncated (in $Evf1^{TS/TS}$). While *Evf2* recruits Smarca4 to the *Akr1b8*-3′-*Dlx5/6UCEin*, Smarca4 recruitment is not rescued in $Evf1^{TS/TS}$, decoupling *Evf2*-RNP recruitment at the *Akr1b8*-3′-*Dlx5/6UCEin* from gene repression. Together, these data support the theory that distinct *Evf2*-Sox2 mechanisms contribute to long-range repression of target genes *Rbm28* and *Akr1b8*, with *Evf2*-Sox2 interactions at the *Rbm28*-5′-*Dlx5/6UCEin*, but not the *Akr1b8*-3′-*Dlx5/6UCEin*, contributing to long-range repression.

Across chr6, *Evf2* and Sox2 co-regulate *Dlx5/6UCEins* near specific genes, as shown in graphs (Fig. 3B-D) and categorized into synergistic positive (green +/+, Fig. 3E), synergistic negative (red−/−) and antagonistic sites (red/green +/− and −/+) (Fig. 3F). *Dlx5/6UCEins* found in both $Evf2^{+/+}$ and $Evf2^{TS/TS}$ are categorized as independent sites (I, Fig. 3E,F). *Evf2*-Sox2 co-regulated *Dlx5/6UCEins* frequently overlap with combinations of *Evf2*-regulated histone marks and RNP binding (Sox2, Dlx and Smarca4; Figs S3 and S4). For example, at the *Evf2*-Sox2 positively regulated *Ing3-Dlx5/6UCEin*, *Evf2* regulates H3K4me3, H3K27Ac and *Evf2*-RNP recruitment (Smc1a, Sox2, Dlx and Smarca4) (Fig. S3C). Similar to profiles at the *Rbm28*-5′ *Dlx5/6UCEin*, Sox2 loss in $Evf2^{TS/TS}$ is rescued in $Evf1^{TS/TS}$ at additional *Dlx5/6UCEins*, including *Ing3*, *Ezh2* and *Rmnd5* (Figs S3C and S4A), supporting a role for the *Evf2*-5′ region in RNP recruitment at these sites. However, at the *Evf2*-Sox2 co-regulated Umad1 and Cc8b1 *Dlx5/6UCEins*, Sox2 recruitment is not rescued in $Evf1^{TS/TS}$ (Figs S3D and S4A), suggesting that the *Evf2*-5′ is not sufficient for recruitment at these sites.

In order to determine the relationships between *Evf2*-regulated RNP binding (Sox2, Dlx and Smarca4), and *Evf2* (+) and (−) regulated *Dlx5/6UCEins* across chr6, we combined RNP CUT&RUN results and *Dlx5/6UCE*-4Cseq data from $Evf2^{+/+}$ and $Evf2^{TS/TS}$ E13.5 GEs (Cajigas et al., 2018; Fig. 4A). *Evf2* increases Sox2 binding [120 bp and 150 bp fragments at both *Evf2* (+) and (−) regulated *Dlx5/6UCEins* (Fig. 4A, Fig. S4B,C)]. Dlx recruitment differs from Sox2 and Smarca4, as *Evf2* decreases Dlx binding overall, with significant differences at *Evf2* (−), but not at *Evf2* (+) regulated *Dlx5/6UCEins* (Fig. 4A). Consistent with site-specific analysis, $Evf1^{TS/TS}$ rescues a subset of the overall effects, distinguishing roles of the *Evf2*-5′ and -3′ regions in RNP recruitment at *Dlx5/6UCEins* (Fig. S4C).

We next asked whether *Evf2* co-regulated RNP recruitment occurs at *Dlx5/6UCEins* and/or at the genome-wide level. *Evf2* co-regulated Sox2/Dlx/Smarca4 (RNP^{co}) binding sites overlap *Dlx5/6UCEins* (150 bp CUT&RUN fragments, Venn diagram in Fig. 4B). *Evf2* (+) regulates RNP^{co} binding at 22 *Dlx5/6UCEins*, where 14/22 coincide with *Evf2* (+) regulated *Dlx5/6UCEins*, and 8/22 coincide with *Evf2* (−) regulated *Dlx5/6UCEins*. *Evf2* (−) regulates RNP^{co} binding at fewer *Dlx5/6UCEins* (5), where 3/5 coincide with *Evf2* (+) regulated *Dlx5/6UCEins* and 2/5 coincide with *Evf2* (−) regulated *Dlx5/6UCEins*. Although the significance of chromosome-wide effects remains to be determined, correlations between *Evf2*-RNP^{co} sites and *Evf2*-regulated *Dlx5/6UCEins* raises the possibility of a wider role for *Evf2*-Sox2 and RNP co-recruitment in topological control.

We next analyzed enrichment of Sox2 DNA motifs at *Evf2*-regulated Sox2 peaks (Fig. 4B). At *Evf2* (+)-regulated Sox2 peaks, 38% of 120 bp fragments contain Sox2 motifs, when compared with 8% of 150 bp fragments, while at *Evf2* (−)-regulated Sox2 peaks, the numbers are 43% for 120 bp and 7% for 150 bp. These data are consistent with CUT&RUN analysis of Sox2-binding profiles where Sox2 120 bp fragments are motif enriched, while 150 bp fragments are motif depleted, and reflect nucleosomal binding (Meers et al., 2019b). Comparison of Sox2 motif enrichment at *Evf2*-regulated RNP^{co} sites compared with Sox2 singly bound sites shows reductions at both *Evf2* (+)-regulated and *Evf2* (−)-regulated sites (∼3-fold and ∼4.4-fold, respectively) (Fig. 4B, 150 bp analysis). Even greater differences are identified in the 120 bp analysis, where Sox2 motifs are not detected at *Evf2* RNP^{co} sites, but are detected in 38% *Evf2* (+)-regulated and 43% *Evf2* (−)-regulated Sox2 singly bound sites. Together, these findings suggest that co-recruitment of Sox2 with Dlx and Smarca4 decreases genome-wide Sox2-DNA motif binding.

Evf2 regulates the Sox2 RNP protein pool targeting to *Dlx5/6* and repressed genes

We previously found that the *Evf2*-5′ expressed in $Evf1^{TS/TS}$ is sufficient for both RNA cloud formation and colocalization with *Dlx5/6UCE* (Cajigas et al., 2018). Here, we find that *Evf1*-3′ RNA (overlapping *Evf2*-3′) continues to be expressed in E13.5GE $Evf2^{TS/TS}$ nuclei, forming clouds that are properly targeted to repressed genes (*Akr1b8* and *Rbm28*) (Fig. 5A-D). These data suggest that *Evf2*-5′ and *Evf1*-3′ RNA clouds localize to regulated target genes, but that RNA cloud targeting is not sufficient for gene activation or repression. As *Evf1*-3′ lacks the UCE containing *Evf2*-5′, but overlaps with the *Evf2*-3′, $Evf2^{TS/TS}$ and $Evf1^{TS/TS}$ provide ideal models to compare *Evf2*-5′ and -3′ end functions *in vivo*.

Confocal microscopy in wild-type E13.5 GE nuclei previously showed that *Evf2*-RNPs are enriched within *Evf2* RNA clouds [Dlx (Feng et al., 2006), Smarca4 (Cajigas et al., 2015) and Smc1a

Development (2021) 148, dev197202. doi:10.1242/dev.197202

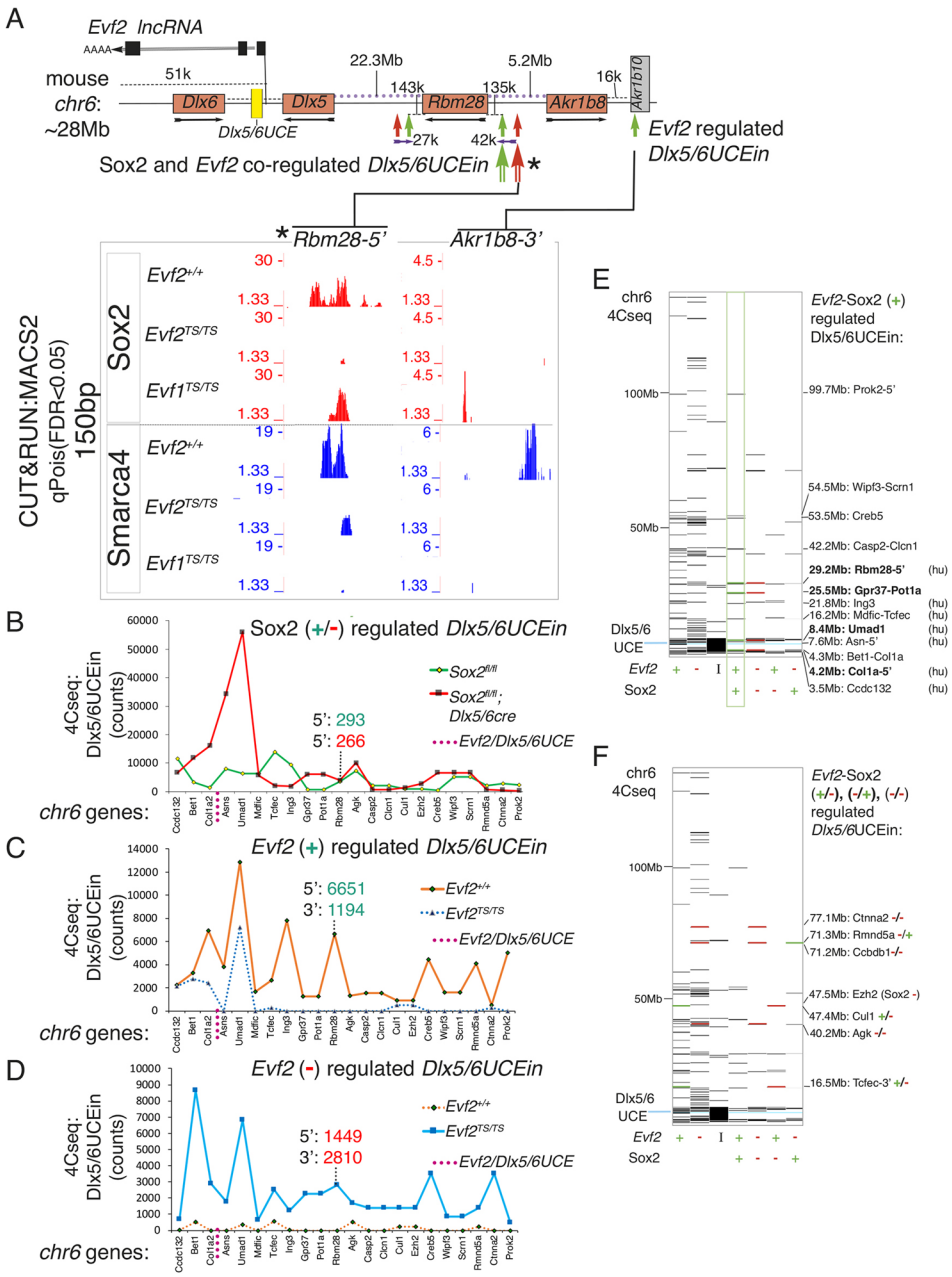


Fig. 3. *Evf2* and Sox2 co-regulate *Dlx5/6UCE* positioning at the Rbm28-5′, overlapping with Sox2 and Smarca4 recruitment. 4Cseq using *Dlx5/6UCE* as bait in E13.5 ganglionic eminences (GEs) from *Sox2*$^{fl/fl}$;*Dlx5/6cre*+, *Sox2*$^{fl/fl}$;*Dlx5/6cre*-, *Evf2*$^{+/+}$ and *Evf2*$^{TS/TS}$, and CUT&RUN ChIPseq (E13.5 GE from *Evf2*$^{+/+}$, *Evf2*$^{TS/TS}$ and *Evf1*$^{TS/TS}$) show profiles of *Evf2* regulated *Evf2* RNP binding (Sox2, Dlx and Smarca4) across chr6. (A) Schematic showing the relationship between *Evf2* repressed target genes (orange boxes), *Evf2* regulated *Dlx5/6UCE* interaction sites (*Dlx5/6UCEin*; solid red and green arrows) and *Evf2*/Sox2 co-regulated *Dlx5/6UCEin* shifting 42 kb towards repressed target gene *Rbm28*-5′ (asterisk, double-lined red and green arrows). CUT&RUN profiles of Sox2 and Smarca4 binding identifies *Evf2*-5′-mediated regulation of Sox2 binding (absent in *Evf2*$^{TS/TS}$ and rescued in *Evf1*$^{TS/TS}$) at the *Rbm28*-5′ *Dlx5/6UCE* targeted site. *Evf2* regulation of Smarca4 binding is not rescued in *Evf1*$^{TS/TS}$. (B-D) 4Cseq *Dlx5/6UCEin* counts (*y*-axis) at genes on mouse chromosome 6 (*x*-axis) where *Evf2* and Sox2 co-regulate *Dlx5/6UCEin*. Vertical dotted purple lines on the *x*-axis indicate the location of the *Evf2* and *Dlx5/6UCE* bait. (B) Counts obtained from Sox2 positively regulated sites (*Sox2*$^{fl/fl}$, green) and negatively regulated sites (*Sox2*$^{fl/fl}$ *;Dlx5/6cre*, red). Counts from the *Rbm28*-5′ *Dlx5/6UCEin* are indicated (green, Sox2+ regulated; red, Sox2− regulated). 4Cseq analysis using *Dlx5/6UCE* bait in *Sox*$^{fl/fl}$;*Dlx5/6cre*+/−, *n*=4 each genotype. Intersection of two computational methods (FourCseq and DESeq) was reported to identify *Dlx5/6UCEin* (Cajigas et al., 2018). (C,D) 4Cseq-*Dlx5/6UCEin* counts based on previously reported data (Cajigas et al., 2018). (C) Counts obtained from *Evf2* positively regulated sites (*Evf2*$^{+/+}$, orange solid lines) and negatively regulated sites (*Evf2*$^{TS/TS}$, blue dotted lines). Counts from the *Rbm28*-5′ and 3′ *Dlx5/6UCEin* are indicated (green, *Evf2*+ regulated). (D) Counts obtained from *Evf2* negatively regulated sites (*Evf2*$^{TS/TS}$, blue solid lines) and positively regulated sites (*Evf2*$^{+/+}$, orange dotted lines). Counts from the *Rbm28*-5′ and 3′ *Dlx5/6UCEin* are indicated (red, *Evf2* regulated). (E,F) *Evf2*-Sox2 regulated *Dlx5/6UCEin* across chr6 (~150 Mb), where FDR<0.05 sites are identified by bars. Genes closest to Dlx5/6UCEin are indicated with chr6 position in Mb listed on the right. Evf2 and Sox2 positively regulated sites are indicated in columns marked by (+), while negatively regulated sites are indicated in columns marked by (−). I indicates *Dlx5/6UCEin* that are independent of *Evf2* regulation (found in both *Evf2*$^{+/+}$ and *Evf2*$^{TS/TS}$ by 4Cseq analysis, as previously reported by Cajigas et al., 2018). Conserved organization of genes on mouse chr6 and human chr7 (hu) are indicated. (E) Positively co-regulated *Dlx5/6UCEin* sites are boxed in green and nearby genes listed. Green (+) and red (−) bars indicate overlapping Evf2 and Sox2 co-regulated Dlx5/6UCEin, representing shifts near four genes (bold). (F) Genes near antagonistic +/−, −/+ and negatively co-regulated *Dlx5/6UCEin* sites are listed. Green (+) and red (−) bars indicate overlapping Evf2- and Sox2-regulated Dlx5/6UCEin. At Ezh2, Sox2-negative regulation is near the antagonistic Cul1 site.

Development (2021) 148, dev197202. doi:10.1242/dev.197202

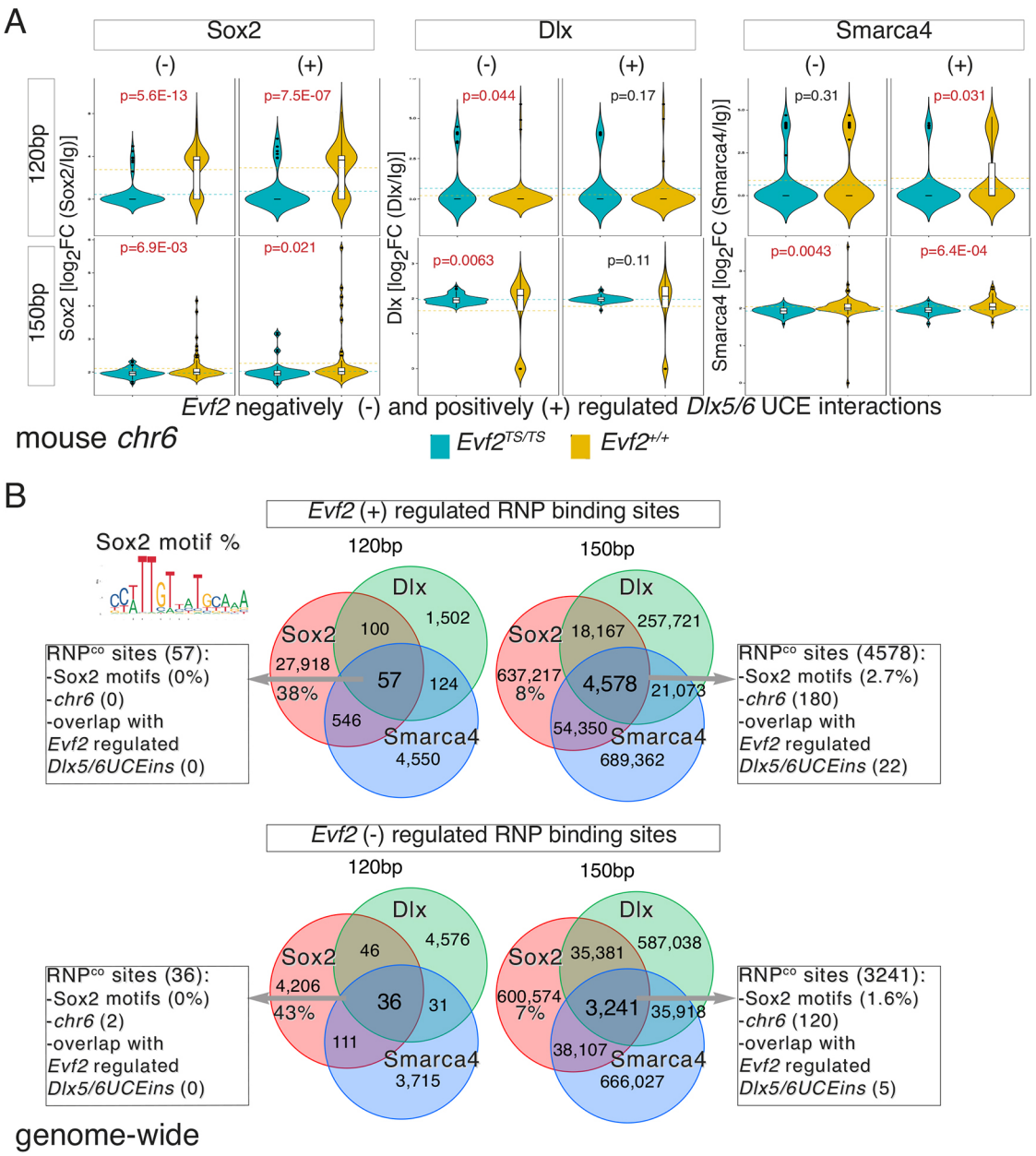


Fig. 4. *Evf2*-RNP recruitment at *Dlx5/6UCEin* and genome-wide effects of co-recruitment on Sox2 DNA binding. (A) Violin plots comparing *Evf2* regulated RNP binding (CUT&RUN, Sox2, Dlx and Smarca4) at *Evf2* positively (+) and negatively (–) regulated *Dlx5/6UCE* interaction sites (*Dlx5/6UCEin*) (4Cseq) across chr6 in E13.5 ganglionic eminences (GEs). Student's *t*-test, $P<0.05$ in red; $Evf2^{TS/TS}$ (blue); $Evf2^{+/+}$ (yellow). Reads are normalized to IgG. (B) Venn diagrams of the number of *Evf2* positively (+) and negatively (–) regulated Sox2, Dlx and Smarca4 peaks identified by CUT&RUN in E13.5 GEs. The percentage of Sox2 peaks that contain Sox2-DNA motifs is indicated (Sox2 DNA motif defined by JASPAR, using FDR<0.01 cut-off). Sox2 peaks from 120 bp fragment sequencing are enriched for Sox2 DNA motifs compared with 150 bp [*Evf2* (+) regulated, 120 bp/38% versus 150bp/8%; *Evf2* (–) regulated, 120bp/43% versus 150bp/7%]. Analysis of 150 bp fragments shows that the percentage of co-regulated sites (arrow, bound by all three RNPs: Sox2, Dlx and Smarca4; RNP^{co}) with Sox2 DNA motifs decreases compared with Sox2 singly bound sites [*Evf2* (+) regulated, RNP^{co} 2.7% versus Sox2 8%; *Evf2* (–) regulated, RNP^{co} 1.6% versus Sox2 7%]. RNP^{co} sites identified in the 120 bp analysis do not contain known Sox2 DNA motifs (0%) compared with singly bound Sox2 *Evf2* (+) regulated (38%) or *Evf2* (–) regulated (43%). RNP^{co} sites are associated with 27 *Dlx5/6UCEin*, 22 at *Evf2* (+)-regulated RNP^{co} sites and 5 at *Evf2* (–) RNP^{co}-regulated sites.

(Cajigas et al., 2018)]. Therefore, we used RNA fluorescent *in situ* hybridization combined with immunofluorescence to investigate whether *Evf2* affects Sox2 localization. Unlike diffusely distributed Smarca4 protein enriched in *Evf2*-RNA clouds (Cajigas et al., 2015), Sox2 forms heterogeneously sized nuclear condensates in E13.5GE nuclei (Fig. 5A,G). Here, we name *Evf2*-RNP spherical subnuclear domains as protein pools (PPs), to distinguish these from spherical *Evf2* RNA subnuclear 'clouds', a term originally used to describe nuclear domains formed by *Kcnq1ot1* lncRNA (Redrup et al., 2009). Sox2 PPs colocalize with *Evf2* RNA clouds and with *Evf2*-repressed target genes *Akr1b8* and *Rbm28* (Fig. 5A, additional examples in Fig. S5). Examples of monoallelic, biallelic and non-colocalized Sox2 PPs-*Evf2* RNA clouds are detected (Fig. S5). Sox2 PP colocalization with Dlx5/6 and

Development (2021) 148, dev197202. doi:10.1242/dev.197202

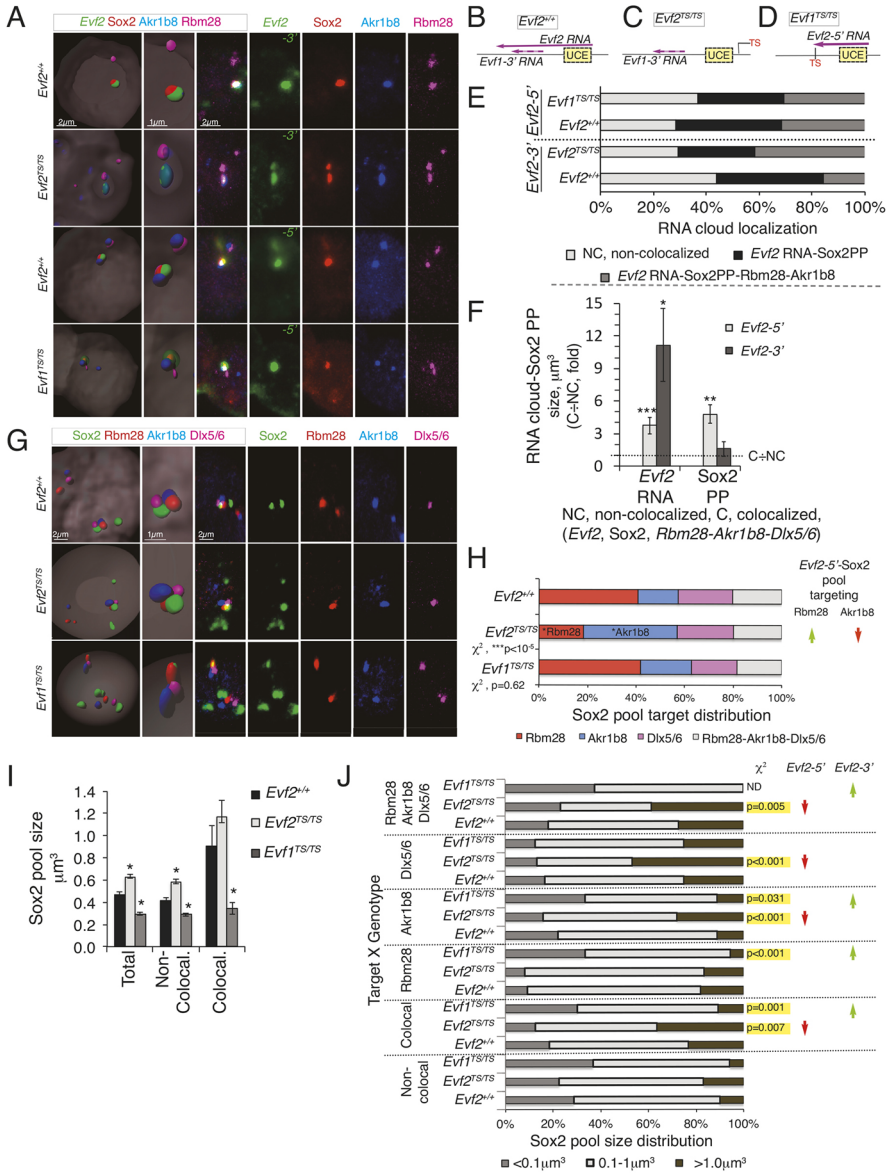


Fig. 5. *Evf2* regulates Sox2 protein pool targeting and size. (A) Fluorescence *in situ* hybridization confocal analysis of *Evf2* RNA clouds (green), Sox2 protein pools PPs (red), *Akr1b8* DNA (blue) and *Rbm28* DNA (pink) in $Evf2^{+/+}$, $Evf2^{TS/TS}$ and $Evf1^{TS/TS}$ nuclei. Left two panels: IMARIS 3D reconstructions. *Evf2*-3′ end or -5′ end antisense probes differentiate between transcripts. (B-D) Schematics of *Evf2* full-length, *Evf2* truncated 5′ and *Evf1*-3′ transcripts in different genotypes. TS, triple polyadenylation signal insertion to stop transcription; UCE, *Dlx5/6UCE*. (B) $Evf2^{+/+}$. (C) $Evf2^{TS/TS}$. (D) $Evf1^{TS/TS}$. (E) Percentages of RNA clouds colocalized with repressed target genes *Akr1b8* and *Rbm28* or with Sox2PP are not significantly changed in $Evf1^{TS/TS}$ or $Evf2^{TS/TS}$ mutants. $\chi 2$ analysis, $P>0.05$. $n=144$ for *Evf2*-5′ probe, $n=73$ for *Evf2*-3′ probe. (F) In $Evf2^{+/+}$ nuclei, sizes of *Evf2*-5′ and -3′ RNA clouds colocalized with Sox2, *Rbm28* or *Akr1b8* are significantly larger than non-colocalized clouds; Sox2 protein pool sizes colocalized with *Evf2*-5′, *Rbm28* or *Akr1b8* are larger than non-colocalized. Student's *t*-test (*$P<0.05$, **$P<0.01$, ***$P<0.001$). Data are mean±s.e.m. $n=32$ ($Evf2^{+/+}$: *Evf2*-3′ probe), $n=41$ ($Evf2^{TS/TS}$: *Evf2*-3′ probe), $n=95$ ($Evf2^{+/+}$: *Evf2*-5′ probe) and $n=54$ ($Evf1^{TS/TS}$: *Evf2*-5′ probe). (G) Confocal analysis in $Evf2^{+/+}$, $Evf2^{TS/TS}$ and $Evf1^{TS/TS}$ E13.5 ganglionic eminence (GE) nuclei: Sox2 protein pools (green), *Rbm28* DNA (red), *Akr1b8* DNA (blue) and *Dlx5/6* DNA (pink). (H) The number of Sox2 PPs located at *Rbm28*, *Akr1b8* and *Dlx5/6*, and simultaneously at *Rbm28*, *Akr1b8* and *Dlx5/6* in $Evf2^{+/+}$ ($n=65$), $Evf2^{TS/TS}$ ($n=88$) and $Evf1^{TS/TS}$ ($n=56$) nuclei were determined; $\chi 2$ analysis indicates that Sox2 PP distribution among targets in $Evf2^{TS/TS}$ ($P<10^{-5}$) differs from distribution in $Evf2^{+/+}$. *Evf2*-5′, but not -3′, increases the numbers of Sox2 protein pools targeted to *Rbm28* (green arrow) and decreases targeting to *Akr1b8* (red arrow). (I,J) IMARIS volumetric measurements of Sox2 PP sizes performed in E13.5 GE nuclei from $Evf2^{+/+}$, $Evf2^{TS/TS}$ and $Evf1^{TS/TS}$. (I) Sox2 PPs are larger in $Evf2^{+/+}$ compared with $Evf1^{TS/TS}$, but smaller compared with $Evf2^{TS/TS}$ (total and non-colocalized; colocalization is defined by fluorescence *in situ* hybridization colocalization with *Rbm28*, *Akr1b8* and/or *Dlx5/6UCE*, simultaneously or in any combination), supporting the theory that *Evf2*-5′ reduces Sox2 PP size, while *Evf2*-3′ increases Sox2 PP size. Sox2 PPs: $Evf2^{+/+}$ (non-colocalized, $n=605$; colocalized, $n=65$), $Evf2^{TS/TS}$ (non-colocalized, $n=993$; colocalized, $n=88$) and $Evf1^{TS/TS}$ (non-colocalized, $n=611$; colocalized, $n=56$). ANOVA, post-hoc Dunnett's C test (*$P<0.05$). (J) Analysis of Sox2 PP sizes categorized into three groups (<0.1 µm, 0.1-1.0 µm and >1 µm) for each genotype and target gene location (non-colocalized, colocalized and specific target gene). Arrows on the right summarize conclusions regarding *Evf2*-5′ (red arrows indicate a decrease in size) and *Evf2*-3′ (green arrows indicate an increase in size) -mediated control of Sox2 PP populations. In $Evf2^{TS/TS}$, the percentage of Sox2 PPs increases in the >1 µm category (colocalized, *Akr1b8* and *Dlx5/6*), whereas profiles are unchanged at *Rbm28*. In $Evf1^{TS/TS}$, the percentage of Sox2 PPs increase in <0.1 µm and/or 0.1-1.0 µm categories at the expense of larger sizes (>1 µm category) at all colocalized targets, except *Dlx5/6*. Non-colocalized Sox2 PP size profiles are not altered in $Evf2^{TS/TS}$ and $Evf1^{TS/TS}$. *P* values from $\chi 2$ analysis (compared with $Evf2^{+/+}$) are shown on the right. *n* values are equivalent to those in I. Colocalization of Evf2 RNA clouds and Sox2 PPs was based on 3D analysis of serial confocal *z*-stacks obtained from single nuclei. *Evf2* RNA clouds and Sox2 PP sizes were determined based on 3D reconstructions of confocal *z*-stacks using IMARIS software.

Development (2021) 148, dev197202. doi:10.1242/dev.197202

repressed target genes is shown in Fig. 5G, with additional examples in Fig. S6.

Colocalization and size analysis of Sox2PPs and *Evf2* RNA clouds using IMARIS software after 3D reconstruction indicates that a subset of Sox2 PPs colocalizes with *Evf2*-5′ and -3′ RNA clouds (Fig. 5E,F). While the percentages of *Evf2*-5′ and -3′ RNA clouds colocalized with Sox2 PPs does not significantly change in $Evf1^{TS/TS}$ and $Evf2^{TS/TS}$ nuclei, the sizes of *Evf2*-5′ and -3′ RNA clouds are significantly larger when colocalized with Sox2-PPs or with *Evf2*-repressed target genes (Fig. 5F). Furthermore, *Evf2*-5′ RNA clouds colocalizing with Sox2 PPs and repressed target genes are larger than non-colocalized (Fig. 5F).

Visualization of Sox2 PPs in $Evf1^{TS/TS}$ and $Evf2^{TS/TS}$ nuclei indicates that Sox2 PPs continue to be targeted to *Rbm28*, *Akr1b8* and *Dlx5/6* (Fig. 5G). However, in $Evf2^{TS/TS}$ the percentage of Sox2-PPs colocalized with *Akr1b8* increases at the expense of *Rbm28*, a shift that is rescued in $Evf1^{TS/TS}$ (Fig. 5H). Total and non-colocalized Sox2-PPs are larger in $Evf2^{TS/TS}$, and smaller in $Evf1^{TS/TS}$ nuclei (Fig. 5I), where non-colocalization is defined as non-overlapping with *Rbm28*, *Akr1b8* or *Dlx5/6*, using IMARIS software parameters normalized according to $Evf2^{+/+}$ values. Increased variability of $Evf2^{TS/TS}$ colocalized Sox2-PPs (Fig. 5I) led to binning Sox2-PPs into three size groups (<0.1 µm^3, 0.1-1 µm^3 and >1 µm^3) for comparisons of size distributions at specific targets in $Evf2^{+/+}$, $Evf2^{TS/TS}$ and $Evf1^{TS/TS}$ nuclei (Fig. 5J). Colocalized Sox2PP sizes are increased in $Evf2^{TS/TS}$ (red arrows) and decreased in $Evf1^{TS/TS}$ (green arrows), with larger effects observed at *Akr1b8* (alone or complexed with *Rbm28* and *Dlx5/6*) (Fig. 5J). At *Dlx5/6*, the percentage of $Evf2^{TS/TS}$ Sox2-PPs with a volume greater than 1 µm^3 increases at the expense of 0.1-1 µm^3 volumes (Fig. 5J, red arrow). This shift in Sox2PP sizes is rescued in $Evf1^{TS/TS}$, supporting the significance of Sox2PP size regulation at *Dlx5/6UCE* during gene repression. Together, these data support the observations that *Evf2*-5′ and -3′ differentially regulate Sox2-PP gene targeting and size distributions in a site-specific manner.

We next used fluorescent *in situ* hybridization analysis of N-terminally tagged, mCherry-Sox2 (mch-Sox2) transfected into E13.5 GEs to determine whether ectopically expressed Sox2 associates with endogenous *Evf2* RNA clouds and/or *Dlx5/6*. Transfected mch-Sox2 forms PPs that colocalize with endogenous *Evf2* RNA clouds and/or *Dlx5/6* (Fig. 6A), through a minimal region spanning the HMG nucleic acid binding domains and adjacent NLSs (orange) (Sox2$^{40\text{-}120}$, Fig. 6A-C). In Sox2 mutant 3 (Sox2$^{40\text{-}67,\ 98\text{-}317}$), crucial RNA/DNA-binding amino acids within the HMG domain are deleted: Δ66-97 deletes W79, K80, K87 and K95, defined as RNA and/or DNA binding (Holmes et al., 2020). However, Sox2 mut3 colocalization with endogenous *Evf2* RNA clouds and/or *Dlx5/6* increases, supporting the theory that nucleic acid binding is dispensable, while amino acids in the NLSs are crucial for PP formation and RNA/DNA localization. Future experiments that distinguish between Sox2 nuclear localization and RNA cloud/*Dlx5/6* localization and define the role of individual Sox2-RNP interactions in localization will be important for understanding *Evf2* RNP assembly and targeting *in vivo*.

DISCUSSION

Understanding lncRNA-dependent chromosome topological control requires mechanistic experiments that define individual contributions of lncRNA-RNPs. In addition to RNA cloud formation, *Evf2* shares functional characteristics with *Xist*, one of the most well-studied lncRNAs (gene repression, chromatin remodeling effects and topological control) (Giorgetti et al., 2016; Jégu et al., 2019; Nora et al., 2012), and formation of a similar sized RNP (*Xist*-RNP^{85}; Chu et al., 2015), where 16/85 proteins are shared with the *Evf2*-RNP^{87} (Cajigas et al., 2015). The *Xist-RNP* is among the most well-characterized lncRNA regulatory complexes, with validated functions on individual proteins (Chen et al., 2016; Chu et al., 2015; Dossin et al., 2020; Minajigi et al., 2015; Yi et al., 2020). Cohesin recruitment is a shared function between *Evf2* (Cajigas et al., 2018), *Xist* (Minajigi et al., 2015) and ThymoD lncRNAs (Isoda et al., 2017), raising the possibility that topological control is a shared function of cloud-forming lncRNAs. Given that components of the *Evf2*-RNP interact with other lncRNAs, the study of gene regulation by the *Evf2*-RNP provides important insight into general mechanisms of gene regulation by lncRNAs, as well as potentially unique characteristics in developing interneurons.

Multi-step *Evf2*-Sox2 interactions support distinct gene repression mechanisms at long-range target genes

Although Sox2 transcriptional activities have been extensively characterized, new roles have emerged for Sox2 RNA-binding activities. The Sox2 HMG region contains both DNA- and RNA-binding domains (Holmes et al., 2020), which are also necessary for direct *Evf2* binding (Fig. 1F). Together with genetic epistasis experiments and luciferase reporter assays (Fig. 2A,B), these experiments support the observation that *Evf2* gene repression occurs through direct antagonism in which *Evf2* lncRNA binds to Sox2, reduces the binding of Sox2 to DNA regulatory elements and decreases enhancer activity.

However, evidence in this work also demonstrates a role for Sox2-*Evf2* lncRNA interactions in regulating enhancer targeting and *Evf2*-Sox2 protein recruitment, in events that likely precede direct effects on enhancer activity (models in Fig. 6D-F). We propose that key to this multi-step model of regulation is the formation of the *Evf2*-RNP, which is assembled on both lncRNA and protein scaffolds. lncRNA scaffolds can bridge individual RNA-binding proteins through low sequence-specific and/or promiscuous RNA-binding properties, as shown in this report for Sox2 (*Evf2*-Sox2; Fig. 1E) and previous work for chromatin remodelers [*Evf2*-Smarca4, *Evf2*-Smarcc2 and *Evf2*-Smarcc1 (Cajigas et al., 2015)] (Fig. 1A). Multivalent RNA-binding proteins Sox2 (Fig. 1B) and Smarca4 (Cajigas et al., 2015) can bridge non-RNA binding RNPs (Dlx, Smarcb1) to the lncRNA. Low sequence-specific RNA-binding explains the ability of the lncRNA to act as an organizer, with ‘glue’ like properties, controlling the availability of specific RNPs in a site-specific manner. In such a model, DNA site specificity is determined by enhancer transcription and sequence as follows: (1) enhancer transcription produces an lncRNA that is retained and provides a scaffold where the RNP grows; and (2) enhancer sequences recruit and stabilize TF binding (Sox2 and Dlx).

Evf2 RNA clouds (yellow circles) and *Evf2*-RNP-Sox2 PP (green stars) sizes are larger when colocalized with *Dlx5/6UCE* and/or specific DNA target genes than non-localized (Fig. 6D), supporting the idea that the *Evf2*-RNP grows at key DNA regulatory sites. As revealed by analysis of *Evf2* mutants, *Evf2* shifts Sox2 PPs from Akr1b8 towards Rbm28, and also limits Sox2PP size at Akr1b8, linking PP targeting and size regulation to functionality (gene repression).

Incorporation of *Evf2*-Sox2 regulated events (chromosome topology, TF recruitment, PP size and targeting, and genetic and biochemical data) leads to a multi-step model of gene repression (Fig. 6E, steps 1-6). Linear organization of *Evf2* gene repression on chr6 shows relationships between the following: the *Evf2*

Development (2021) 148, dev197202. doi:10.1242/dev.197202

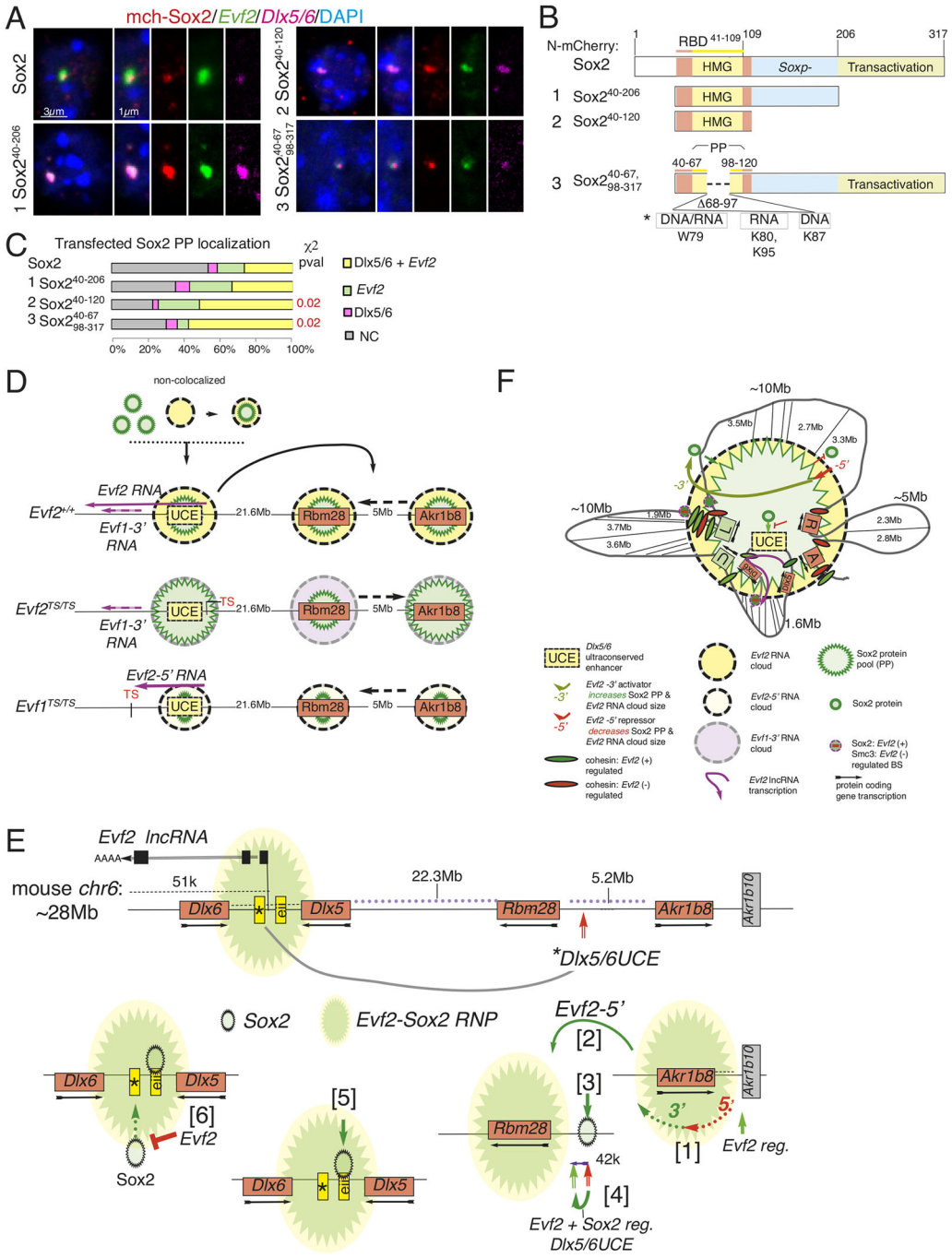


Fig. 6. Multiple contributions *Evf2*-Sox2 interactions to gene repression. (A) Sox2 nuclear localization sequences overlap with *Evf2-Dlx5/6* colocalization in transfected ganglionic eminences (GEs). Fluorescence *in situ* hybridization confocal analysis of E13.5 GE nuclei transfected with mCherry-Sox2 fusion constructs [Sox2, full length; (1) Sox2 40-206, (2) Sox2 40-120 and (3) Sox2 deletion 68-97]: mCherry (red), *Evf2* RNA clouds (green), *Dlx5/6UCE* (pink) and DAPI (blue). (B) Schematic of Sox2 mCherry fusion proteins used for transfection into E13.5 GE cells. Orange bars flanking the HMG domain contain nuclear localization signal sequences, HMG (high mobility group domain, DNA/RNA binding), RBD (RNA-binding domain defined from *in vitro* binding Fig. 1), Soxp-pfam12236 (domain shared by Sox protein family members) and PP (protein pool). Sox2-HMG domain flanked by NLS is sufficient for Sox2 PP formation, colocalization with *Evf2* RNA clouds and colocalization with the *Dlx5/6UCE*. The PP co-localization domain is further narrowed to two regions overlapping the NLS (40-67 and 98-120) outside critical DNA- (K87), RNA- (K80, K95) and DNA/RNA- (W79) binding amino acids identified by Holmes et al. (2020) (asterisk). (C) Regions outside the two PP/NLS regions (40-67 and 98-120) decrease Sox2 colocalization with *Evf2* RNA clouds and *Dlx5/6*. The percentage of Sox2 PP colocalized with *Evf2* RNA and *Dlx5/6UCE* is shown for Sox2 mutants 1-3 (yellow bars), P=0.02 from χ2 analysis are shown on the right (χ2 analysis of distributions), n=60 (mchSox2), n=51 (mchSox2$^{40\text{-}206}$ mutant 1), n=35 (mchSox2$^{40\text{-}120}$ mutant 2) and n=33 (mchSox2$^{\Delta 68\text{-}97}$ mutant 3). Colocalization was based on 3D analysis of serial confocal z-stacks obtained from single nuclei. (D-F) Modelling multi-step contributions of *Evf2*-Sox2 interactions during gene repression. In the absence of Evf2, the Dlx5/6UCE binds to Rbm28-5' (red arrow). [1] Upon Evf2 expression, the Evf2 RNP containing Sox2 PPs (green star) and Evf2 RNA cloud (yellow) forms and associates with Akr1b8, directing Dlx5/6UCE to bind Akr1b8-3' (green arrow). Evf2-5' (dotted red arrow) decreases Sox2 PP size while Evf2-3' increases Sox2 PP size (dotted green arrow). [2] Evf2-5' shifts the position of Sox2 PPs from Akr1b8 to Rbm28. [3] Evf2-5' recruits Sox2 to the Evf2-Sox2 negatively regulated Dlx5/6UCEin (red arrow), causing [4] Dlx5/6UCE shifting ~42kb closer to Rbm28. [5] Evf2-5' recruits Sox2 protein to the shadow enhancer (Dlx5/6eii). [6] Evf2 directly inhibits Sox2 activation of Dlx5/6UCE.

Development (2021) 148, dev197202. doi:10.1242/dev.197202

transcription site across *Dlx5/6UCE*(*); the *Evf2* RNP assembly on *Dlx5/6UCE* containing the *Evf2* RNA cloud (yellow) and Sox2PP (green); repressed target genes (red boxes); and the *Evf2*/Sox2 negatively regulated *Dlx5/6UCE* interaction site 5′ of Rbm28 (double red arrow) associated with gene repression. At Akr1b8, the *Evf2*-5′ (dotted red arrow) reduces the size of *Evf2*-Sox2 RNPs (step 1) and shifts *Evf2*-Sox2 RNPs towards *Rbm28* (step 2). In this model, *Evf2*-5′ repression of *Akr1b8* occurs by limiting the availability of Sox2 (an activator of *Dlx5/6UCE*) through RNP targeting and size regulation, thereby sequestering Sox2 PPs within the *Evf2*-RNP. Furthermore, *Evf2*-5′ (which is sufficient for *Akr1b8* repression) properly balances the numbers of *Evf2*-Sox2RNPs at *Akr1b8* and *Rbm28* in $Evf1^{TS/TS}$, linking rescue of *Evf2*-Sox2RNP targeting to rescue of gene repression.

In step 3, *Evf2*-5′ recruits Sox2 to the *Rbm28*-5′-*Dlx5/6UCEin* (double red arrow), shifting *Rbm28*-5′-*Dlx5/6UCEin* towards Rbm28-5′ (step 4) (double green arrow). One caveat to *Evf2* gene repression through limiting Sox2 PP activator is that in $Evf2^{TS/TS}$, Sox2-PPs targeted to *Rbm28* (also a repressed target gene) decrease, while Rbm28 gene expression increases. One possibility is that the extent of repression is affected by a combination of Sox2 targeting and size regulation at the two repressed target genes: loss of *Evf2* causes a ∼15-fold increase in *Akr1b8*, but a ∼2-fold increase in *Rbm28* (Cajigas et al., 2018). Another possibility is that mechanisms of *Evf2*-Sox2 antagonism differ between *Akr1b8* and *Rbm28*, as reflected by target gene dependence. This is supported by 4Cseq analysis showing that Sox2 shifts *Dlx5/6UCE* towards *Rbm28*-5′ (overlapping with *Evf2*-*Dlx5/6UCE* regulation), but does not affect *Dlx5/6UCEin* near *Akr1b8*. Site-specific *Evf2*-Sox2 synergistic and antagonistic regulation of *Dlx5/6UCEins* across chr6 further supports variable combinations of events during topological control.

In step 5, *Evf2* recruits RNPs (Dlx and Smarca4) to *Dlx5/6UCE* and Sox2 to *Dlx5/6eii* (Fig. 2C), consistent with previous reports of *Evf2*-regulated recruitment (Bond et al., 2009; Cajigas et al., 2015). Although *Dlx5/6eii* lacks ultraconserved sequences, *Dlx5/6* intergenic enhancers are functionally similar, both are regulated by Dlx and *Evf2* (Feng et al., 2006; Zerucha et al., 2000). Deletion of *Dlx5/6eii* in mice alters gene expression in developing and adult interneurons (Fazel Darbandi et al., 2016), supporting overlapping and distinct functions of the *Dlx5/6UCE*/*Dlx5/6eii* enhancer pair *in vivo*. A key distinguishing feature of *Dlx5/6UCE* is transcription into a spliced polyadenylated lncRNA (*Evf2*), whereas stable transcripts from *Dlx5/6eii* are not detected. With respect to *Evf2*-RNP binding, Sox2 binds both enhancers, while Dlx and Smarca4 binding is limited to *Dlx5/6UCE*. *Evf2*-RNP recruitment also differs: *Evf2* regulates Sox2 binding to *Dlx5/6eii*, but not to *Dlx5/6UCE*, whereas *Evf2* regulates Dlx, Smarca4 binding to *Dlx5/6UCE*, but not to *Dlx5/6eii*. Such differential recruitment is consistent with a 'separation of inputs' hypothesis that enables shadow enhancer pairs to buffer noise better than duplicate enhancers (Waymack et al., 2020). Thus, *Evf2* recruitment of Sox2 to *Dlx5/6eii* may reflect the need to precisely regulate levels of Sox2, beyond that necessary for factors that are recruited to *Dlx5/6UCE*, where on/off decisions are made.

In step 6, *Evf2* decreases *Dlx5/6UCE* activity by directly binding to the Sox2-HMG DNA-binding domain, antagonizing activation. This model is supported by REMSAs (Fig. 1E,F) and luciferase reporter assays (Fig. 2B), and by competition between RNA and DNA at the Sox2 HMG domain (Holmes et al., 2020). The ability of *Evf2* lncRNA to directly inhibit chromatin remodeling by inhibiting ATPase activity (Cajigas et al., 2015), and extension of this mechanism to X-inactivation (Jégu et al., 2019), adds an extra dimension to lncRNA-mediated transcriptional regulation that is not schematized in the model.

Together, these data suggest that *Evf2* long-range gene repression involves multiple levels of *Evf2*-RNP regulation, including Sox2 PP targeting and size regulation, and site-specific recruitment and sequestration that ultimately affect both *Dlx5/6UCE* targeting and activity (shown in the predicted arrangement in Fig. 6F).

Predicting functional significance

Although the majority of this work focuses on *Evf2*-Sox2 differential gene repression of long-range target genes *Rbm28* and *Akr1b8*, chr6-wide and genome-wide effects are also detected. Consistent with our previous reports of *Evf2* topological and histone modification control (Cajigas et al., 2018), *Evf2* regulated RNP-binding sites and *Evf2*-Sox2 regulated *Dlx5/6UCEins* are also detected outside the *Evf2* 27Mb GRN region, revealing chromosome-wide effects. A subset of overlapping *Evf2*-Sox2 regulated *Dlx5/6UCEins* map across chr6, and are located within 50 kb of *Evf2* RNP recruited sites and histone modifications. Notably, analysis at Ing3 and Rmnd5a (Figs S3C and S4A) reveal the combinatorial nature of *Evf2*-RNP regulation. Interestingly, Sox2-negative regulation of *Dlx5/6UCEin* at the *Ezh2*-5′ promoter overlaps with *Evf2* (+)-regulated Sox2 and Smarca4 and *Evf2* (−)-regulated Dlx, suggesting topological control of *Ezh2*, which has been identified as a crucial component of the Polycomb group complex responsible for histone methylation and gene silencing (Cao et al., 2002) (Fig. S4A). Recently, Ezh2 inhibitors have been approved to treat cancer (Richart and Margueron, 2020), but also potentially increase seizure susceptibility in mice (Wang et al., 2020). Thus, it will be important to address whether *Evf2*-mediated regulation of RNPs in E13.5 GE is predictive and, specifically, whether *Evf2*-Sox2-controlled *Dlx5/6UCEins* near *Ezh2*-5′, contribute to the seizure susceptibility phenotype observed in adult mice lacking *Evf2* (Cajigas et al., 2018).

In addition to *Evf2*-regulated RNP recruitment at *Dlx5/6UCEins*, genome-wide analysis identifies thousands of *Evf2*-regulated Sox2-binding sites, a subset that are co-regulated with Dlx and Smarca4 (RNP^{co}, Fig. 4B). *Evf2* recruits Sox2 to 714,312 sites, but also inhibits Sox2 binding to 677,303 sites, supporting a role in balancing Sox2 recruitment and sequestration on a genome-wide scale. Surprisingly, we find that *Evf2* (+) and (−)-regulated RNP^{co} sites contain significantly fewer Sox2 DNA motifs compared with singly bound Sox2 sites (Fig. 4B). Given that CUT&RUN 120 bp fragments are associated with direct Sox2 DNA binding, and 150 bp fragments are associated with nucleosomal binding (Meers et al., 2019b), loss of Sox2 motifs at RNP^{co} sites supports a role for Dlx and Smarca4 co-recruitment in Sox2 DNA interactions. Similar effects are observed at both *Evf2* (+) and *Evf2* (−)-regulated sites, suggesting that individual *Evf2* RNPs influence Sox2-DNA interactions, even in the absence of the *Evf2* RNA, *in vivo*. Although the significance of RNP^{co} sites and *Evf2* genome-wide effects is unknown, these results highlight the importance of designing *in vivo* experiments to determine whether lncRNA-RNPs influence TF-DNA motif recognition as a general mechanism.

How individual *Evf2*-RNPs regulate *Evf2* RNA cloud assembly at specific DNA sites is not known. Identification of Nono in the *Evf2* RNP (Cajigas et al., 2015) and the known role for Nono in NEAT lncRNA assembly into paraspeckles raise the possibility that Nono plays a similar role in the *Evf2*-RNP (Yamazaki et al., 2018). Experiments that tag lncRNA clouds and follow assembly through high-resolution microscopy will be important in validating static

Development (2021) 148, dev197202. doi:10.1242/dev.197202

studies of *Evf2* RNA clouds and *Evf2*-RNPs. As only one or two *Evf2* RNA clouds are detected in the nucleus, either the *Evf2*-RNP moves with *Dlx5/6UCE* to repressed target genes, or DNA looping (Rao et al., 2017) brings *Evf2*-RNP-bound *Dlx5/6UCE* closer to target genes (Cajigas et al., 2018). It will be important to determine whether additional *Evf2*-RNPs form PPs and are regulated similarly to Sox2 PPs with respect to size and targeting.

Experiments in this report reveal a complex multi-step role for *Evf2*-Sox2 interactions in enhancer targeting and direct antagonism of *Dlx5/6UCE* activity during gene repression. How each of the individual *Evf2* RNPs (of the 87 identified) contributes to *Dlx5/6UCE* targeting and activity regulation, and the relationship of these functions to RNP assembly and to common mechanisms of lncRNA transcriptional control remain unresolved. Our data suggest that *Evf2* selectively represses genes across megabase distances by coupling recruitment and sequestration of Sox2, a crucial pioneer transcription factor, affecting key steps of enhancer targeting and activity, with genome-wide effects. We propose that *Evf2* RNA clouds function as both chromosome and protein organizers, contributing to dynamic regulation at enhancers.

MATERIALS AND METHODS

Mice were housed and treated according to approved IACUC guidelines (Northwestern University IACUC). Embryonic brain E13.5 GEs contained mixtures of males and females. The following mouse strains were used: $Evf2^{+/+}$ $Evf2^{TS/TS}$ (Bond et al., 2009), $Evf1^{TS/TS}$ (Cajigas et al., 2018), $Sox2^{fl/+}$ (Jackson laboratory 013093) and *Dlx5/6cre* (Jackson laboratory 008199).

Recombinant protein pulldown

Co-immunoprecipitation experiments on 6×-his tagged proteins purified from *E.coli* or flag-tagged baculovirus proteins were performed as described (Cajigas et al., 2015). Proteins were incubated in 200 µl NETN buffer [100 mM NaCl, 0.5 mM EDTA, 20 mM Tris-HCl (pH 8.0), 0.5% NP-40] for 1 h at 4°C with rotation. Glutathione agarose beads (30 µl) were washed with NETN buffer and added to each sample. Samples were incubated for 1 h at 4°C with rotation. Beads were pelleted by centrifugation and washed three times with 1 ml NETN buffer and once with 1×PBS. Proteins were eluted by adding protein loading buffer [62.5 mM Tris (pH 6.8), 10% glycerol, 2% SDS, 5% β-mercaptoethanol, 0.002% Bromophenol Blue] and incubating at 95°C for 5 min. Samples were analyzed by western blot.

RNA electrophoretic mobility assay

The NIR probe was generated by *in vitro* transcription of *pGEM-Evf2(UCR)* (Cajigas et al., 2015), a plasmid containing 115 nt of *Evf2*, including the ultraconserved sequence. 1 µg of SalI linearized DNA template, 5 mM DTT, 0.6 µl RNasin (Promega), 0.5 mM ATP, 0.5 mM CTP, 0.5 mM GTP, 12.5 µM UTP, 20 mM Aminoallyl-UTP-Atto680, 1 µg BSA and 2 µl (100 U) T7 RNA polymerase were incubated in 20 µl 1×RNA polymerase buffer for 1 h at 37°C. 2 µl Turbo DNase (Life Technologies) and 2 µl 10×Turbo DNase Buffer were added to the reaction and incubated at 37°C for 15 min. The RNA was denatured and separated on a 6% urea-polyacrylamide gel, cast on a Hoeffer miniVE apparatus and pre-run 20 min before loading. Full-length probe was excised, eluted overnight at 4°C in 0.5 M ammonium acetate/1 mM EDTA and ethanol precipitated. The concentration of the NIR labeled RNA probe was measured by absorption at 260 nm using the NanoDrop 1000 (Thermo Scientific). The RNA competitors were generated by *in vitro* transcription. *pGEM-T Easy* was linearized with BsrBI to generate a 209 bp RNA competitor. *pGEM-Evf2(UCR)* was linearized with SalI to generate a 206 bp RNA competitor. The linearized templates were treated with proteinase K and ethanol precipitated. RNA was transcribed as follows: 1.25 µg DNA template, 10 mM DTT, 1.5 µl (80 U) RNasin (Promega), 2 mM A, C, G and UTP (Roche), and 2 µl (100 U) T7 RNA polymerase (NEB) in 50 µl 1×RNA polymerase buffer were incubated at 37°C for 1 h. Samples were incubated with 2 µl Turbo DNase (Life Technologies) in 1×Turbo DNase Buffer for 15 min at 37°C. Samples were treated with Proteinase K (Roche), ethanol precipitated and quantified using the Quantifluor RNA System (Promega).

To generate GST-tagged Sox2 proteins, full-length Sox2 and Sox2 truncations ($Sox2^{1\text{-}205}$, $Sox2^{1\text{-}109}$, $Sox2^{206\text{-}317}$ and $Sox2^{41\text{-}317}$) were subcloned into pGEX4T1. GST fusion proteins were purified from bacteria using standard protocols. The recombinant proteins were incubated with 0.15 pmoles *Evf2* NIR-labeled probe in 10 µl reactions for 30 min at room temperature. For all competition experiments, protein and competitor RNA were pre-incubated for 10 min at room temperature before adding probe. 5 µg tRNA and 0.5 µl RNasin (Promega) were included in all the RNA electrophoretic mobility assay (REMSA) reactions. Pre-electrophoresis of 4% native polyacrylamide gels was performed for 20 min, REMSA reactions loaded and electrophoresed at 200 V for 40 min, and data visualized in the Odyssey Infrared Imager (LI-COR Biosciences).

Primary embryonic brain MGE transfections

For all transfections, E13.5 MGE tissues were dissected from Swiss Webster mouse embryos, dissociated in L15 media by pipetting several times, and spun through a cell strainer for single cell preparations. Cells were seeded at a density of 2.5×10^5 cells per cm^2 (Flandin et al., 2011) in neurobasal medium [DMEM/F-12 supplemented with L-glutamate, B-27 (Gibco), N2 supplement (Gibco), bovine pituitary extract (35 µg/ml; Life Technologies), mito+ serum extender (BD Biosciences), penicillin (100 U/ml; Gibco), streptomycin (100 µg/ml; Gibco) and glutamax (0.8 mM; Gibco)]. One day before seeding cells, plate wells were coated with poly-L-lysine (Sigma) and laminin (Sigma).

For luciferase experiments, 78,300 cells per well were cultured in a 96-well microplate treated for tissue culture. Cells were allowed to attach for 24 h before changing the medium to neurobasal media without antibiotics. Transfections using Fugene 6 were performed as recommended. Cells were harvested 48 h after transfection with 1×passive lysis buffer (Promega) supplemented with 0.1% digitonin for cell lysis. To ensure thorough cell lysis, lysates were subjected to two freeze-thaw cycles prior to performing Dual Luciferase Reporter assays. All transfections were normalized to the internal control expressing *Renilla luciferase*, performed at least in triplicate and a minimum of two times. For transfection of mCherry-Sox2 fusions, 850,000 cells per well were cultured in a 12-well tissue culture plate. mCherry-Sox2 plasmids were generated by subcloning Sox2 or Sox2 truncations ($Sox2^{40\text{-}206}$ and $Sox2^{40\text{-}120}$) into the mCherry2-C1 plasmid using SacI and KpnI. Quick Change Site Directed Mutagenesis (Agilent) mutagenesis was used to generate $Sox2^{\Delta68\text{-}97}$. Plasmid (1 µg) was transfected using Fugene 6, according to the manufacturer's instructions. Cell were harvested by scrapping after 72 h of incubation and nuclei were isolated for combined RNA/DNA fluorescence *in situ* hybridization and immunofluorescence using an anti-mCherry antibody (see method below).

Combined DNA-RNA fluorescence *in situ* hybridization with immunofluorescence

DNA fluorescence *in situ* hybridization probes were generated by nick translation using the fluorescence *in situ* hybridization Tag DNA Kit following manufacturer's recommendations. The templates for the nick translation reactions were obtained from the BACPAC Resources Center (Children's Hospital Oakland Research Institute): Dlx5/6 region, WI1-1693G2; Akr1b8 region, RP23-120B14; Rbm28 region, RP23-276H18. The digoxigenin-labeled RNA probe was generated as described previously (Feng et al., 2006).

E13.5 whole ganglionic eminences were dissected in L15. Tissues were pooled for each genotype, triturated by pipetting and filtered through a cell-strainer capped 5 ml polystyrene round-bottomed tube (BD Falcon) to make single-cell suspensions. Cells were pelleted by centrifugation at 100 ***g*** for 5 min at 4°C. The supernatant was removed and cells were gently resuspended in 500 µl Nuclear Extraction Buffer [0.32 M sucrose, 5 mM $CaCl_2$, 3 mM $Mg(Ac)_2$, 0.1 mM EDTA, 20 mM Tris-HCl (pH 8.0), 0.1% TritonX-100] and incubated on ice for 10 min. Cells were centrifuged at 100 ***g*** for 2.5 min at 4°C and the supernatant was removed. Cells were washed gently with ice-cold 1×PBS with 2 mM EGTA. Cells were centrifuged at 100 ***g*** for 2.5 min at 4°C. The supernatant was removed and cells were gently resuspended in 500 µl of ice-cold fixative (3:1, methanol:glacial acetic acid). The cells were fixed for

Development (2021) 148, dev197202. doi:10.1242/dev.197202

10 min on ice. 5 µl of cells in fixative were transferred to Superfrost Plus microscope slides (Fisher Scientific) and allowed to air dry. The slides were transferred to a slide holder, vacuum sealed and stored at −80°C.

Slides were incubated with 50 µg/ml pepsin in 0.01 M HCl at 37°C for 7 min, and washed twice with 2×SSC. Cells were fixed in 4% paraformaldehyde for 5 min at room temperature and washed three times with 2×SSC for 5 min. The slides were incubated in 1×PBS with 1% hydrogen peroxide for 30 min at room temperature and rinsed twice with 2×SSC. The slides were dehydrated by incubation for 2 min in 70%, 80% and 100% ethanol. 200 µl denaturation solution (70% formamide in 2×SSC) were added and the slides were incubated at 85°C for 10 min. Slides were dehydrated in ice-cold 70%, 80% and 100% ethanol for 2 min and allowed to air dry. 150 µl pre-hybridization buffer (50% formamide, 0.1% SDS, 300 ng/ml Salmon Sperm DNA and 2×SSC) were added and the slides were incubated overnight at 37°C.

DNA probes and RNA probe in hybridization buffer (50% formamide, 10% dextran sulfate, 0.1% SDS, 300 ng/ml Salmon Sperm DNA and 2×SSC) were denatured in the presence of 2 µg mouse Hybloc DNA (Applied Genetics Laboratories) at 80°C for 7 min and re-annealed at 37°C for 1 h. Slides were incubated for 5 min in 2×SSC with 50% formamide, 2 min in 4×SSC with 0.1% Tween-20 and 2 min in 2×SSC at 45°C. The slides were dehydrated in ethanol and denatured as described above. 10 µl of fluorescence *in situ* hybridization probe solution was added, coverslips were sealed with rubber cement and the slides were incubated overnight at 37°C.

Slides were incubated in 2×SSC with 50% formamide for 10 min (three times), in 2×SSC for 10 min and in 2×SSC with 0.1% NP40 for 5 min at 45°C. The slides were rinsed with 1×PBS and incubated in 1% blocking solution (Tyramide Signal Amplification Kit) for 1 h. The appropriate antibody was diluted 1:500 in blocking reagent, along with a mouse monoclonal anti-digoxigenin (DIG, 1:500), added to the slides and incubated at 4°C overnight. Slides were washed three times in 1×PBS for 3 min at room temperature, incubated with 1:100 HRP-goat anti-mouse IgG in blocking solution for 1 h at room temperature and tyramide labeled according to manufacturer's instructions (TSA Kit). A second tyramide labeling step was performed for immunostaining. Slides were washed three times in 1×PBS for 3 min at room temperature after the first round of labeling. The slides were then incubated with 1:100 HRP-goat anti-rabbit IgG for 1 h at room temperature and tyramide labeled. The slides were washed three times with 1×PBS for 3 min and incubated with 5 mg/ml DAPI for 5 min, rinsed with 1×PBS and mounted using SlowFade Gold antifade reagent (Thermo Fisher Scientific).

Nuclei were visualized using a Zeiss Laser Scanning Microscope 880 using the Zen 2.1 software. A 100× immersion oil objective was used to generate *z*-stacks of 0.3 µm (2D colocalization for Sox2 PP transfections) and a 63× immersion oil objective used to generate *z*-stacks of 0.1 µm intervals (3D colocalization for volume analysis). Colocalization of transfected mCherry-fused SoxPPs with Evf2 RNA clouds and Dlx5/6UCE was determined by Zen2.1 software and manual inspection of *z*-stacks through each nucleus. Imaris software was used for 3D reconstruction, colocalization analysis and size measurements. Colocalization was defined as any regional overlap in one of four channels. The numbers of overlapping clouds and/or Sox2PPs with DNA target genes were determined by IMARIS software following 3D reconstruction, and in conjunction with size determination. Threshold settings were determined in pilot experiments performed in *Evf2*$^{+/+}$, and remained constant throughout the analysis between genotypes.

ChIPseq

Ten E13.5 whole ganglionic eminences were dissected for X-ChIP, native ChIP and ChIP-reChIP, as previously reported (Cajigas et al., 2018). For ChIP-reChIP, the pre-cleared chromatin from Swiss Webster E13.5GE was preadsorbed with rabbit IgG (5 µg), followed by 1st round incubations with anti-DLX (5 µg) and 2nd round incubations with anti-Sox2 or anti-LaminB1 (1 µg).

CUT&RUN

CUT&RUN was performed as previously described (Skene et al., 2018), with some modifications. The number of samples was determined according to the number of antibodies to be tested and the number of replicates (two replicates per antibody and two replicates for IgG). The appropriate volume of Concanavalin A magnetic beads (10 µl per sample) was mixed into 1.5 ml of ice-cold binding buffer [20 mM HEPES-KOH (pH 7.9), 10 mM KCl, 1 mM $CaCl_2$, 1 mM $MnCl_2$] and placed on a magnetic stand for 2 min. The supernatant was removed, and the beads were washed with 1.5 ml binding buffer. After magnetic separation of the beads, the supernatant was removed and a volume of binding buffer equal to the initial beads volume (10 µl per sample) was added to the beads. The beads were placed on ice. E13.5 ganglionic eminences were isolated from embryos in L15 medium (six embryos per genotype). Tissues were pooled for each genotype, triturated by pipetting and filtered through a cell-strainer capped 5 ml polystyrene round-bottomed tube (BD Falcon) to generate single-cell suspensions. Cells were counted using the Luna Automated Cell Counter (Logos Biosystems). The appropriate volume of cells to obtain 250,000 cells per sample was centrifuged at 600 ***g*** for 3 min at 4°C. The supernatant was removed and the cell pellet was gently resuspended in ice-cold wash buffer [20 mM HEPES (pH 7.5), 150 mM NaCl, 0.5 mM spermidine and EDTA-free protease inhibitor cocktail]. The cells were centrifuged at 600 ***g*** for 3 min at 4°C. The supernatant was removed and cells were gently resuspended in 1 ml of wash buffer. The Concanavalin A bead suspension was added to the cells, while gently vortexing (∼100 ***g***), and the tube was incubated with rotation for 10 min at 4°C. The cells/beads suspension was split into aliquots, according to the number of samples previously determined. Tubes were placed on a magnetic stand and the supernatant removed. 50 µl of antibody buffer [20 mM HEPES (pH 7.5), 150 mM NaCl, 0.5 mM spermidine, 2 mM EDTA, 0.02% digitonin, EDTA-free protease inhibitor cocktail] containing 0.5 µg (Sox2, Smarca4, 1:100 dilution) or 0.07 µg (Dlx, 1:75 dilution) of antibody was added to each tube, while gently vortexing. The samples were incubated with rotation for 2 h at 4°C. The samples were centrifuged for 5 s at 100 ***g***, and placed on the magnetic stand. The supernatant was removed, and the pellet resuspended gently in 1 ml ice-cold digitonin buffer [20 mM HEPES (pH 7.5), 150 mM NaCl, 0.5 mM spermidine, 0.02% digitonin and EDTA-free protease inhibitor cocktail]. The digitonin buffer wash was repeated once. Tubes were placed on the magnetic stand, the supernatant removed and 50 µl of pA-Mnase solution (final concentration 700 ng/ml in digitonin buffer) was added to each tube while gently vortexing. The samples were incubated with rotation for 1 h at 4°C. The samples were centrifuged for 5 s at 100 ***g*** and placed on the magnetic stand. The supernatant was removed and the pellets were washed in ice-cold digitonin buffer twice, as described above. After removal of supernatant from washes, 150 µl of digitonin buffer were added to each sample while gently vortexing. Tubes were placed on a metal block on ice (0°C) for 5 min. Tubes were removed from ice briefly to add 3 µl 100 mM $CaCl_2$ while gently vortexing, then tubes were returned to 0°C for 30 min. 100 µl of stop buffer (340 mM NaCl, 20 mM EDTA, 4 mM EGTA, 0.02% digitonin, 0.05 mg/ml RNAseA, 2 pg/ml heterologous spike-in DNA) were added to each sample and mixed with gentle vortexing. Samples were incubated for 10 min at 37°C, then centrifuged for 5 min at 4°C at 16,000 ***g***. The tubes were placed on the magnetic stand and the supernatant was transferred to a clean 1.5 ml microcentrifuge tube. DNA extraction was performed using standard phenol chloroform and ethanol precipitation methods as described (Skene et al., 2018). Samples were ethanol precipitated overnight at −20°C. DNA pellets were dissolved in 20 µl 0.1×TE buffer [1 mM Tris-HCl (pH 8), 0.1 mM EDTA]. The Qubit High-Sensitivity Assay was used for DNA quantification. CUT&RUN libraries were prepared using the KAPA Hyper Prep Kit protocol, with some modifications. The total volume of CUT&RUN DNA was used for library construction. For adapter ligation, 5 µl of 3 µM adapter stock from the KAPA Dual-Indexed Adapter Kit was used. The ligation was incubated at 20°C for 15 min. The library was amplified using the following cycling conditions: 14 cycles of 98°C for 45 s, 98°C for 15 s and 60°C for 10 s; 72°C for 1 min. After library amplification, the libraries were purified using 50 µl of KAPA Pure Beads and eluted in 20 µl of water. CUT&RUN sample quality was analyzed by TapeStation prior to sequencing on a NovaSeq 6000 (SP 100 cycles).

CUT&RUN data processing

Paired-end CUT&RUN reads of different transcription factors and their respective Ig-controls from Evf2+/+ and Evf2TS/TS Dlx5/6UCE samples

Development (2021) 148, dev197202. doi:10.1242/dev.197202

were first mapped on mm9 genome using Bowtie2 v2.1.0 (Langmead and Salzberg, 2012) with options '–local –very-sensitive-local –no-unal –no-mixed –no-discordant –phred33 -I 10 -X 700'. We used the Picard toolkit command 'MarkDuplicates' to mark PCR duplicates and remove them from the final mm9 genome mapped bam files. Next, we separated the sequence fragments into ≤120 and ≥150 bp classes that provided the mapping of the local vicinity of a DNA-binding protein. The base-pair sizes can vary depending on the steric access to the DNA by the tethered MNase (Skene et al., 2018). Fragments mapping to repeat elements were removed, and replicates were joined before peak calling. The peak calling was performed using MACS2 (Zhang et al., 2008) callpeak options '-t -c –f BED -g mm –keep-dup all –bdg –nomodel –slocal 500 –llocal 5000 –-extsize 120/150'. An FDR cutoff of 0.05 was used to call the final set of peaks (Janssens et al., 2018). Differential CUT&RUN analysis of transcription factor-binding peaks between two Dlx5/6UCE conditions was performed using MACS2 program by treating one of the samples as the 'control' for the other.

Multi-overlap of CUT&RUN peaks

The multi-overlapping differential and continuous CUT&RUN peaks (FDR<0.05) of Sox2, Dlx and Smarca4 transcription factors belonging to ≤120 and ≥150 bp classes from *Evf2* (+) and *Evf2* (−) conditions were quantified using bedtools 'multintersect' function (Quinlan and Hall, 2010) separately. The Venn diagram of the respective overlapping set was plotted using 'ggpubr' R package (Kassambara).

CUT&RUN Sox2 peaks containing Sox2 DNA-binding motifs

Sox2 frequency matrices (MA0143.1 and MA0143.2) from the JASPAR database (Fornes et al., 2020) were used to scan the respective CUT&RUN peak sequences using the PWMEnrich Bioconductor package (Stojnic and Diez, 2020, PWMEnrich: PWM enrichment analysis, R package version 4.26.0.). An FDR cutoff of 0.01 was used to determine significant motif enrichment.

4C-Seq

4C using Dlx5/6UCE as bait was performed as previously described (van de Werken et al., 2012), with some modifications. E13.5 GEs from $Sox2^{fl/fl}$; Dlx5/6 cre+ and $Sox2^{fl/fl}$;Dlx5/6 cre-single embryos were dissected in L15 and kept in separate tubes. A single-cell suspension was obtained through gentle pipetting of the tissue in 250 µl of L15. Cells (~2×10^6 cells per embryo) were transferred to a tube containing 5 ml 2% paraformaldehyde/10% fetal bovine serum (FBS) and incubated with rotation for 10 min at room temperature. 710 µl of 1 M glycine was added to quench the formaldehyde and tubes were placed on ice. Cells were pelleted by centrifugation at 400 ***g*** for 8 min at 4°C. The supernatant was removed and cells were gently resuspended in 2.5 ml ice-cold 4°C lysis buffer [50 mM Tris (pH 7.5), 150 mM NaCl, 5 mM EDTA, 0.5% NP-40, 1% Triton X-100 and protease inhibitors] and incubated for 10 min on ice. Cells were pelleted by centrifugation at 750 ***g*** for 5 min at 4°C. The supernatant was removed, and the cells were washed with 1 ml ice-cold 1×PBS by gentle resuspension. Cells were transferred to 1.5 ml microcentrifuge tubes and centrifuged at 600 ***g*** for 2 min at 4°C. The supernatant was removed and the cell pellets were flash frozen in liquid nitrogen and stored at −80°C.

The cell pellets were resuspended in 440 µl of molecular grade water and 60 µl of cutsmart buffer (New England Biolabs) were added. Tubes were incubated at 37°C and 15 µl of 10% SDS were added. Samples were incubated for 1 h at 37°C while shaking at 900 rpm. 75 µl of 20% Triton X-100 was added to the samples and incubated at 1 h at 37°C while shaking at 900 rpm. 200 U of EcoRI-HF were added to the samples and incubated overnight at 37°C while shaking. The next day, 200 U of EcoRI-HF were added and incubated overnight at 37°C with shaking. Complete digestion was confirmed by agarose gel electrophoresis. If undigested DNA was still present, another 200 U of restriction enzyme was added and incubated overnight. The enzyme was inactivated at 65°C for 20 min. Samples were transferred to a 15 ml conical tube. 100 U of T4 DNA Ligase was added (2 ml reaction volume) and incubated overnight at 16°C. The next day, 500 µl of molecular grade water, 50 µl fresh T4 ligase buffer and 100 U of T4 DNA ligase were added to the samples and incubated overnight at 16°C. Complete ligation was confirmed by agarose gel electrophoresis. The DNA was extracted using standard ethanol precipitation procedures (van de Werken et al., 2012).

The DNA pellet was resuspended in 450 µl of molecular grade water, then 50 µl of DpnII buffer and 50 U of DpnII were added and incubated at 37°C overnight. Complete digestion was confirmed by agarose gel electrophoresis. The enzyme was inactivated at 65°C for 20 min. Samples were transferred to a 50 ml conical tube and ligation was performed using 100 U of T4 DNA ligase in a total volume of 3.5 ml at 16°C overnight. The DNA was precipitated using standard ethanol precipitation procedures. The samples were purified using the Qiaquick PCR purification kit (two columns per sample). The following steps were performed to generate the 4C library for sequencing. First, overhangs were added to the 4C template using PCR amplification with primers containing the bait sequence, as follows: 200 ng 4C template, 0.2 mM dNTPs, 35 pmol primer *Dlx5/6UCE*-Fwd, 35 pmol primer *Dlx5/6UCE*-Rev, 1.75 U Expand Long Template Enzyme Mix (Roche) and 1×Buffer I underwent 29 cycles of 94°C for 2 min, 94°C for 10 s, 55°C for 1 min and 68°C for 3 min; followed by 68°C for 5 min. The PCR product was purified using the High Pure PCR Product Purification Kit (Roche). The 4C DNA containing the overhangs was then used as a template for a second PCR that added index sequences and Illumina sequencing adapters to generate the 4C library for sequencing. PCR reaction (50 µl) was made up of 225 ng DNA template, 0.5 mM dNTPs, 5 µl Nextera XT Index1 primer (N7XX, Illumina), 5 µl Nextera Index 2 primer (S5XX, Illumina), 3.5 U Expand Long Template Enzyme Mix (Roche) and 1×Buffer I. The reaction was 8 cycles of 94°C for 5 min, 94°C for 10 s, 55°C for 30 s and 68°C for 1 min; followed by 68°C for 7 min. The PCR product was purified using the High Pure PCR Product Purification Kit (Roche).

Dlx5/6UCE bait 4C-seq differential data analysis

Analysis has previously been described for $Evf2^{+/+}$ and $Evf2^{TS/TS}$ (Cajigas et al., 2018). 4C reads were first mapped at the EcoRI restriction enzyme cut sites on chromosome 6 of mm9 reference genome using Bowtie2 v2.1.0 (Langmead and Salzberg, 2012). The mapped reads were further filtered based on their reproducibility between the pair of replicates. An EcoRI cut-site was deemed to reproducibly interact or not interact with the 4C bait if the two replicates in a given condition ($Evf2^{+/+}$ and $Evf2^{TS/TS}$) both have non-zero or zero counts, respectively. We identified 1108 and 1266 non-zero count EcoRI restriction cut sites that are reproducible in both replicates of $Evf2^{+/+}$ and $Evf2^{TS/TS}$, respectively. Across the two conditions ($Evf2^{+/+}$ and $Evf2^{TS/TS}$), we retained a total of 997 reproducible 4C sites that have reproducible interactions in the two replicates of either one condition or in both conditions. We then performed a DESeq2 (Love et al., 2014) -based differential contact count analysis on these sites to identify Evf2-regulated sites [*P*-adjusted<0.05 and a log2 fold change≥2 for positively regulated (+) or ≤−2 for negatively regulated (−)] and Evf2-independent (I) (*P*-adjusted>0.05 and an absolute log2 fold change<2) 4C interaction sites. We also performed the 4Cseq analysis using the FourCSeq program (Klein et al., 2015). In FourCSeq program models, the overall decreasing interaction frequency was related to genomic distance by fitting a smooth monotonically decreasing function to suitably transformed count data. With this transformed and normalized count data, FourCSeq performs differential analysis between conditions to obtain significant differential interactions. We applied FourCSeq on our $Evf2^{+/+}$ and $Evf2^{TS/TS}$ 4Cseq samples and retrieved *Evf2* (+) and (−)-regulated Dlx5/6UCE interactions (*P*-adjusted<0.05 and an absolute log2 fold change≥2) and Evf2-independent interactions (I) (*P*-adjusted>0.05 and an absolute log2 fold change<2). To avoid method-specific biases, interaction sites that were assigned the same label (+ 50,−73, I, 167) by the two different approaches (DESeq2 and FourCSeq) were called 4Cseq-intersectional computational method sites.

$Sox2^{fl/fl}$; Dlx5/6+cre/-cre 4C-seq differential data analysis

Sox2 regulated Dlx5/6UCE interactions were determined as described above for *Evf2*+/+ and $Evf2^{TS/TS}$, except that sites were identified based on non-zero counts in at least two out of four of the replicates in each genotype and DESeq2 (Love et al., 2014) [*P*-adjusted<0.05 and a log2 fold change≥2 for positively regulated or ≤−2 for negatively regulated (+) 4C interaction

Development (2021) 148, dev197202. doi:10.1242/dev.197202

sites]. We applied FourCSeq on *Sox2*$^{fl/fl}$ +*Dlx5/6cre* and *Sox2*$^{fl/fl}$-*cre* 4Cseq samples, and retrieved Sox2-regulated 4C interactions (*P*-adjusted<0.05 and an absolute log2 fold change≥2). To avoid method-specific biases, we retained a common set of 244 Sox2 (+) and (−)-regulated Dlx5/6UCE interaction sites.

Dlx5/6UCE 4C counts for Sox2 (+)-regulated (enriched in Sox2$^{fl/fl}$;Dlx5/6 cre-) and Sox2 (−)-regulated (enriched in Sox2$^{fl/fl}$;Dlx5/6 cre-), and those previously reported for *Evf2* (+)-regulated (enriched in *Evf2*$^{+/+}$), *Evf2* (−)-regulated (enriched in *Evf2*$^{TS/TS}$) and *Evf2* (I) independently regulated (detected in both *Evf2*$^{+/+}$ and *Evf2*$^{TS/TS}$) are included in Table S1.

Dlx5/6UCE and Sox2$^{fl/fl}$ +/– cre site overlap

Evf2$^{+/+}$ and *Evf2*$^{TS/TS}$ *Dlx5/6UCE* and *Sox2*$^{fl/fl}$ +/cre 4C-peak overlap was measured using 'bedtools window' function (Quinlan and Hall, 2010) with a window span of 50 kb.

Dlx5/6UCE 4C-seq and CUT&RUN signal overlap

Evf2$^{+/+}$ and *Evf2*$^{TS/TS}$ *Dlx5/6UCE* 4C-seq peaks were first mapped on CUT&RUN differential transcription factor peaks using bedtools 'intersect' function (Quinlan and Hall, 2010). The overlapping set of 4C-seq peaks from a *Dlx5/6UCE* condition were then mapped on the respective CUT&RUN Ig-normalized transcription factor signal data. The log2 fold-enrichment of Ig-normalized signal was generated using MACS2 'bdgcmp' command (Zhang et al., 2008). The violin plots were made using 'ggpubr' R package (Kassambara).

Quantification and statistical analysis

Quantification and statistical analysis were performed using R. Significance levels are **P*<0.05, ***P*<0.01 and ****P*<0.001. In violin plots, the colored dotted line represents the mean of the respective class. An unpaired *t*-test was used to measure the significance. For ChIP-seq and CUT&RUN, a peak is defined as a region with q<0.05; for a 4C-seq experiment, a significant peak is defined with FDR<0.05 and an absolute log2 fold enrichment≥2. Addition statistical details can be found in the figure legends.

Antibodies and reagents

The following antibodies and reagents were used: anti-DLX (Kohtz Lab, Northwestern University, Evanston, IL, USA; Bond et al., 2009; Cajigas et al., 2015; Feng et al., 2006), anti-Smarca4 (Wang Lab, NIH; Wang et al., 1996), anti-Sox2 (A301-740A; RRID:AB_1211355, Bethyl Laboratories), anti-Lamin B1 (ab16048; RRID:AB_1010782, Abcam), anti-SMC3 (ab9263; RRID:AB_307122, Abcam), anti-mCherry (ab205402; RRID:AB_2722769, Abcam), anti-Flag M2 (F1804; RRID:AB_262044, Sigma), UTP-Atto680 (NU-821-680, Jena Bioscience), (pA-MNase; Henikoff Lab; Skene et al., 2018), Concanavalin A magnetic beads (BP531, Bangs Laboratories), Protein G Agarose (11719416001, Roche), Expand Long Template Enzyme Mix (11681834001, Roche), Fugene 6 (E2691, Promega), Dual Luciferase reporter assay (E1910, Promega), FISH TAG DNA kit (F32951, Thermo Fisher), Nextera XT Index Kit (FC-131-1001, Illumina), TruSeq Nano DNA Library Prep Kit (FC-121–4003, Illumina) KAPA Dual Indexed Adapter Kit (KK8722, KAPA Biosystems), KAPA hyper prep kit (KK8502, KAPA Biosystems), pcDNA3-EGFP (Addgene, 13031), pGL3-mDlx5/6 (Feng et al., 2006), pcDNA-Evf2 (Addgene, 99478), pcDNA3.3-Sox2 (Addgene, 26817), mCherry2-C1 (54563, Addgene) and pGEM-Evf2(UCR) (Kohtz Lab).

Oligonucleotides for TAQman PCR were obtained from Life Technologies: Dlx6 Mm01166201_m1, Dlx5 Mm00438430_m1, Umad1 AJWR2X8, Lsm8 AJX004G, Rbm28 Mm01137037_m1, Akr1b8 Mm00484314_m1, ActB Mm00607939_s1, Sox2 Mm03053810_s and Sox2 ot Mm01291217_m1. 4C bait sequences were as follows: *Dlx5/6UCE*-Fwd, 5′TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGATGCCAAACCACTGTGAGTGTA3′; *Dlx5/6UCE*-Rev, 5′GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGTCCCCAATGTCTGCTTCAA3′.

Acknowledgements

We thank S. Henikoff for pA-MNase (Fred Hutchinson Cancer Research Center), and R. Kingston and J. Cochrane (Harvard) for flag-tagged Smarca4 protein. We also thank C. DiDonato and J. Topczewski for critically reviewing the manuscript.

Competing interests

The authors declare no competing or financial interests.

Author contributions

Conceptualization: I.C., J.D.K.; Methodology: I.C., J.D.K.; Software: A.C., F.A.; Validation: I.C., M.L., M.B., L.C.; Formal analysis: J.D.K., F.A., A.C.; Investigation: I.C., M.L., K.R.S., M.B., L.C., H.L.; Data curation: J.D.K., A.C.; Writing-original draft: J.D.K.; Writing review & editing: J.D.K., I.C., F.A.; Visualization: J.D.K.; Supervision: J.D.K.; Project Administration: J.D.K.; Funding acquisition: J.D.K., F.A.

Funding

This work was funded by the National Institutes of Health (NIMH R01MH111267 to J.D.K. and NIGMS R35GM128938 to F.A.). Deposited in PMC for release after 12 months.

Data availability

The *Sox2*$^{fl/fl}$ +/-*Dlx5*/*6cre* 4C-seq and CUT&RUN datasets generated and analyzed in this study are available in GEO under accession number GSE164301. The published article (Cajigas et al., 2018) includes the crosslinked *X*-ChIPseq and native ChIPSeq datasets, and *Evf2* regulated and independent Dlx5/6UCE-4C-seq datasets (GSE117184) used in this study. Confocal images and Imaris data are available on Mendeley Data, V1 (http://dx.doi.org/10.17632/pfkj9d5fk7.1).

Supplementary information

Supplementary information available online at
https://dev.biologists.org/lookup/doi/10.1242/dev.197202.supplemental

Peer review history

The peer review history is available online at
https://dev.biologists.org/lookup/doi/10.1242/dev.197202.reviewer-comments.pdf

References

Ahituv, N., Zhu, Y., Visel, A., Holt, A., Afzal, V., Pennacchio, L. A. and Rubin, E. M. (2007). Deletion of ultraconserved elements yields viable mice. *PLoS Biol.* **5**, e234. doi:10.1371/journal.pbio.0050234

Avilion, A. A., Nicolis, S. K., Pevny, L. H., Perez, L., Vivian, N. and Lovell-Badge, R. (2003). Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes Dev.* **17**, 126-140. doi:10.1101/gad.224503

Bejerano, G., Pheasant, M., Makunin, I., Stephen, S., Kent, W. J., Mattick, J. S. and Haussler, D. (2004). Ultraconserved elements in the human genome. *Science* **304**, 1321-1325. doi:10.1126/science.1098119

Berghoff, E. G., Clark, M. F., Chen, S., Cajigas, I., Leib, D. E. and Kohtz, J. D. (2013). Evf2 (Dlx6as) lncRNA regulates ultraconserved enhancer methylation and the differential transcriptional control of adjacent genes. *Development* **140**, 4407-4416. doi:10.1242/dev.099390

Bond, A. M., Vangompel, M. J., Sametsky, E. A., Clark, M. F., Savage, J. C., Disterhoft, J. F. and Kohtz, J. D. (2009). Balanced gene regulation by an embryonic brain ncRNA is critical for adult hippocampal GABA circuitry. *Nat. Neurosci.* **12**, 1020-1027. doi:10.1038/nn.2371

Brockdorff, N., Ashworth, A., Kay, G. F., McCabe, V. M., Norris, D. P., Cooper, P. J., Swift, S. and Rastan, S. (1992). The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. *Cell* **71**, 515-526. doi:10.1016/0092-8674(92)90519-I

Brown, C. J., Hendrich, B. D., Rupert, J. L., Lafreniere, R. G., Xing, Y., Lawrence, J. and Willard, H. F. (1992). The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. *Cell* **71**, 527-542. doi:10.1016/0092-8674(92)90520-M

Cajigas, I., Leib, D. E., Cochrane, J., Luo, H., Swyter, K. R., Chen, S., Clark, B. S., Thompson, J., Yates, J. R., Kingston, R. E. et al. (2015). Evf2 lncRNA/BRG1/DLX1 interactions reveal RNA-dependent inhibition of chromatin remodeling. *Development* **142**, 2641-2652. doi:10.1242/dev.126318

Cajigas, I., Chakraborty, A., Swyter, K. R., Luo, H., Bastidas, M., Nigro, M., Morris, E. R., Chen, S., VanGompel, M. J. W., Leib, D. et al. (2018). The Evf2 ultraconserved enhancer lncRNA functionally and spatially organizes megabase distant genes in the developing forebrain. *Mol. Cell* **71**, 956-972.e959. doi:10.1016/j.molcel.2018.07.024

Calin, G. A., Liu, C. G., Ferracin, M., Hyslop, T., Spizzo, R., Sevignani, C., Fabbri, M., Cimmino, A., Lee, E. J., Wojcik, S. E. et al. (2007). Ultraconserved regions encoding ncRNAs are altered in human leukemias and carcinomas. *Cancer Cell* **12**, 215-229. doi:10.1016/j.ccr.2007.07.027

Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., Jones, R. S. and Zhang, Y. (2002). Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* **298**, 1039-1043. doi:10.1126/science.1076997

Chen, C. K., Blanco, M., Jackson, C., Aznauryan, E., Ollikainen, N., Surka, C., Chow, A., Cerase, A., McDonel, P. and Guttman, M. (2016). Xist recruits the X chromosome to the nuclear lamina to enable chromosome-wide silencing. *Science* **354**, 468-472. doi:10.1126/science.aae0047

Chu, C., Zhang, Q. C., da Rocha, S. T., Flynn, R. A., Bharadwaj, M., Calabrese, J. M., Magnuson, T., Heard, E. and Chang, H. Y. (2015). Systematic discovery of Xist RNA binding proteins. *Cell* **161**, 404-416. doi:10.1016/j.cell.2015.03.025

Dickel, D. E., Ypsilanti, A. R., Pla, R., Zhu, Y., Barozzi, I., Mannion, B. J., Khin, Y. S., Fukuda-Yuzawa, Y., Plajzer-Frick, I., Pickle, C. S. et al. (2018). Ultraconserved enhancers are required for normal development. *Cell* **172**, 491-499.e415. doi:10.1016/j.cell.2017.12.017

Dodonova, S. O., Zhu, F., Dienemann, C., Taipale, J. and Cramer, P. (2020). Nucleosome-bound SOX2 and SOX11 structures elucidate pioneer factor function. *Nature* **580**, 669-672. doi:10.1038/s41586-020-2195-y

Dossin, F., Pinheiro, I., Zylicz, J. J., Roensch, J., Collombet, S., Le Saux, A., Chelmicki, T., Attia, M., Kapoor, V., Zhan, Y. et al. (2020). SPEN integrates transcriptional and epigenetic control of X-inactivation. *Nature* **578**, 455-460. doi:10.1038/s41586-020-1974-9

Fazel Darbandi, S., Poitras, L., Monis, S., Lindtner, S., Yu, M., Hatch, G., Rubenstein, J. L. and Ekker, M. (2016). Functional consequences of I56ii Dlx enhancer deletion in the developing mouse forebrain. *Dev. Biol.* **S0012-1606**, 30263-30269. doi:10.1016/j.ydbio.2016.10.015

Feng, J., Bi, C., Clark, B. S., Mady, R., Shah, P. and Kohtz, J. D. (2006). The Evf-2 noncoding RNA is transcribed from the Dlx-5/6 ultraconserved region and functions as a Dlx-2 transcriptional coactivator. *Genes Dev.* **20**, 1470-1484. doi:10.1101/gad.1416106

Flandin, P., Zhao, Y., Vogt, D., Jeong, J., Long, J., Potter, G., Westphal, H. and Rubenstein, J. L. (2011). Lhx6 and Lhx8 coordinately induce neuronal expression of Shh that controls the generation of interneuron progenitors. *Neuron* **70**, 939-950. doi:10.1016/j.neuron.2011.04.020

Fornes, O., Castro-Mondragon, J. A., Khan, A., van der Lee, R., Zhang, X., Richmond, P. A., Modi, B. P., Correard, S., Gheorghe, M., Baranasic, D. et al. (2020). JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* **48**, D87-D92. doi:10.1093/nar/gkaa516

Furlong, E. E. M. and Levine, M. (2018). Developmental enhancers and chromosome topology. *Science* **361**, 1341-1345. doi:10.1126/science.aau0320

Giorgetti, L., Lajoie, B. R., Carter, A. C., Attia, M., Zhan, Y., Xu, J., Chen, C. J., Kaplan, N., Chang, H. Y., Heard, E. et al. (2016). Structural organization of the inactive X chromosome in the mouse. *Nature* **535**, 575-579. doi:10.1038/nature18589

Guo, X., Wang, Z., Lu, C., Hong, W., Wang, G., Xu, Y., Liu, Z. and Kang, J. (2018). LincRNA-1614 coordinates Sox2/PRC2-mediated repression of developmental genes in pluripotency maintenance. *J. Mol. Cell Biol.* **10**, 118-129. doi:10.1093/jmcb/mjx041

Holmes, Z. E., Hamilton, D. J., Hwang, T., Parsonnet, N. V., Rinn, J. L., Wuttke, D. S. and Batey, R. T. (2020). The Sox2 transcription factor binds RNA. *Nat. Commun.* **11**, 1805. doi:10.1038/s41467-020-15571-8

Isoda, T., Moore, A. J., He, Z., Chandra, V., Aida, M., Denholtz, M., Piet van Hamburg, J., Fisch, K. M., Chang, A. N., Fahl, S. P. et al. (2017). Non-coding transcription instructs chromatin folding and compartmentalization to dictate enhancer-promoter communication and T cell fate. *Cell* **171**, 103-119.e118. doi:10.1016/j.cell.2017.09.001

Janssens, D. H., Wu, S. J., Sarthy, J. F., Meers, M. P., Myers, C. H., Olson, J. M., Ahmad, K. and Henikoff, S. (2018). Automated in situ chromatin profiling efficiently resolves cell types and gene regulatory programs. *Epigenetics Chromatin* **11**, 74. doi:10.1186/s13072-018-0243-8

Jégu, T., Blum, R., Cochrane, J. C., Yang, L., Wang, C.-Y., Gilles, M. E., Colognori, D., Szanto, A., Marr, S. K., Kingston, R. E. et al. (2019). Xist RNA antagonizes the SWI/SNF chromatin remodeler BRG1 on the inactive X chromosome. *Nat. Struct. Mol. Biol.* **26**, 96-109. doi:10.1038/s41594-018-0176-8

Klein, F. A., Pakozdi, T., Anders, S., Ghavi-Helm, Y., Furlong, E. E. and Huber, W. (2015). FourCSeq: analysis of 4C sequencing data. *Bioinformatics* **31**, 3085-3091. doi:10.1093/bioinformatics/btv335

Langmead, B. and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357-359. doi:10.1038/nmeth.1923

Love, M. I., Huber, W. and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550. doi:10.1186/s13059-014-0550-8

Meers, M. P., Bryson, T. D., Henikoff, J. G. and Henikoff, S. (2019a). Improved CUT&RUN chromatin profiling tools. *eLife* **8**. doi:10.7554/eLife.46314

Meers, M. P., Janssens, D. H. and Henikoff, S. (2019b). Pioneer factor-nucleosome binding events during differentiation are motif encoded. *Mol. Cell* **75**, 562-575.e565. doi:10.1016/j.molcel.2019.05.025

Meers, M. P., Tenenbaum, D. and Henikoff, S. (2019c). Peak calling by Sparse Enrichment Analysis for CUT&RUN chromatin profiling. *Epigenetics Chromatin* **12**, 42. doi:10.1186/s13072-019-0287-4

Minajigi, A., Froberg, J., Wei, C., Sunwoo, H., Kesner, B., Colognori, D., Lessing, D., Payer, B., Boukhali, M., Haas, W. et al. (2015). Chromosomes. A comprehensive Xist interactome reveals cohesin repulsion and an RNA-directed chromosome conformation. *Science* **349**, 282-296. doi:10.1126/science.aab2276

Monory, K., Massa, F., Egertova, M., Eder, M., Blaudzun, H., Westenbroek, R., Kelsch, W., Jacob, W., Marsch, R., Ekker, M. et al. (2006). The endocannabinoid system controls key epileptogenic circuits in the hippocampus. *Neuron* **51**, 455-466. doi:10.1016/j.neuron.2006.07.006

Ng, S. Y., Johnson, R. and Stanton, L. W. (2012). Human long non-coding RNAs promote pluripotency and neuronal differentiation by association with chromatin modifiers and transcription factors. *EMBO J.* **31**, 522-533. doi:10.1038/emboj.2011.459

Ng, S. Y., Bogu, G. K., Soh, B. S. and Stanton, L. W. (2013). The long noncoding RNA RMST interacts with SOX2 to regulate neurogenesis. *Mol. Cell* **51**, 349-359. doi:10.1016/j.molcel.2013.07.017

Nolte, M. J., Wang, Y., Deng, J. M., Swinton, P. G., Wei, C., Guindani, M., Schwartz, R. J. and Behringer, R. R. (2014). Functional analysis of limb transcriptional enhancers in the mouse. *Evol. Dev.* **16**, 207-223. doi:10.1111/ede.12084

Nora, E. P., Lajoie, B. R., Schulz, E. G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N. L., Meisig, J., Sedat, J. et al. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381-385. doi:10.1038/nature11049

Ørom, U. A. and Shiekhattar, R. (2011). Noncoding RNAs and enhancers: complications of a long-distance relationship. *Trends Genet.* **27**, 433-439. doi:10.1016/j.tig.2011.06.009

Ørom, U. A., Derrien, T., Beringer, M., Gumireddy, K., Gardini, A., Bussotti, G., Lai, F., Zytnicki, M., Notredame, C., Huang, Q. et al. (2010). Long noncoding RNAs with enhancer-like function in human cells. *Cell* **143**, 46-58. doi:10.1016/j.cell.2010.09.001

Osterwalder, M., Barozzi, I., Tissieres, V., Fukuda-Yuzawa, Y., Mannion, B. J., Afzal, S. Y., Lee, E. A., Zhu, Y., Plajzer-Frick, I., Pickle, C. S. et al. (2018). Enhancer redundancy provides phenotypic robustness in mammalian development. *Nature* **554**, 239-243. doi:10.1038/nature25461

Pandey, R. R., Mondal, T., Mohammad, F., Enroth, S., Redrup, L., Komorowski, J., Nagano, T., Mancini-Dinardo, D. and Kanduri, C. (2008). Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. *Mol. Cell* **32**, 232-246. doi:10.1016/j.molcel.2008.08.022

Quinlan, A. R. and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842. doi:10.1093/bioinformatics/btq033

Rao, S. S. P., Huang, S. C., Glenn St Hilaire, B., Engreitz, J. M., Perez, E. M., Kieffer-Kwon, K. R., Sanborn, A. L., Johnstone, S. E., Bascom, G. D., Bochkov, I. D. et al. (2017). Cohesin loss eliminates all loop domains. *Cell* **171**, 305-320.e324. doi:10.1016/j.cell.2017.09.026

Redrup, L., Branco, M. R., Perdeaux, E. R., Krueger, C., Lewis, A., Santos, F., Nagano, T., Cobb, B. S., Fraser, P. and Reik, W. (2009). The long noncoding RNA Kcnq1ot1 organises a lineage-specific nuclear domain for epigenetic gene silencing. *Development* **136**, 525-530. doi:10.1242/dev.031328

Remenyi, A., Lins, K., Nissen, L. J., Reinbold, R., Scholer, H. R. and Wilmanns, M. (2003). Crystal structure of a POU/HMG/DNA ternary complex suggests differential assembly of Oct4 and Sox2 on two enhancers. *Genes Dev.* **17**, 2048-2059. doi:10.1101/gad.269303

Remenyi, A., Scholer, H. R. and Wilmanns, M. (2004). Combinatorial control of gene expression. *Nat. Struct. Mol. Biol.* **11**, 812-815. doi:10.1038/nsmb820

Richart, L. and Margueron, R. (2020). Drugging histone methyltransferases in cancer. *Curr. Opin. Chem. Biol.* **56**, 51-62. doi:10.1016/j.cbpa.2019.11.009

Rinn, J. L. and Chang, H. Y. (2020). Long noncoding RNAs: molecular modalities to organismal functions. *Annu. Rev. Biochem.* **89**, 283-308. doi:10.1146/annurev-biochem-062917-012708

Sandelin, A., Bailey, P., Bruce, S., Engstrom, P. G., Klos, J. M., Wasserman, W. W., Ericson, J. and Lenhard, B. (2004). Arrays of ultraconserved non-coding regions span the loci of key developmental genes in vertebrate genomes. *BMC Genomics* **5**, 99. doi:10.1186/1471-2164-5-99

Shaham, O., Smith, A. N., Robinson, M. L., Taketo, M. M., Lang, R. A. and Ashery-Padan, R. (2009). Pax6 is essential for lens fiber cell differentiation. *Development* **136**, 2567-2578. doi:10.1242/dev.032888

Skene, P. J. and Henikoff, S. (2017). An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *eLife* **6**, e21856. doi:10.7554/eLife.21856

Skene, P. J., Henikoff, J. G. and Henikoff, S. (2018). Targeted in situ genome-wide profiling with high efficiency for low cell numbers. *Nat. Protoc.* **13**, 1006-1019. doi:10.1038/nprot.2018.015

Stojnic, R. and Diez, D. (2020). PWMEnrich: PWM enrichment analysis. R package version 4.26.0. doi:10.18129/B9.bioc.PWMEnrich

Takahashi, K. and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663-676. doi:10.1016/j.cell.2006.07.024

van de Werken, H. J., de Vree, P. J., Splinter, E., Holwerda, S. J., Klous, P., de Wit, E. and de Laat, W. (2012). 4C technology: protocols and data analysis. *Methods Enzymol.* **513**, 89-112. doi:10.1016/B978-0-12-391938-0.00004-5

Wang, W., Côté, J., Xue, Y., Zhou, S., Khavari, P. A., Biggar, S. R., Muchardt, C., Kalpana, G. V., Goff, S. P., Yaniv, M. et al. (1996). Purification and biochemical heterogeneity of the mammalian SWI-SNF complex. *EMBO J.* **15**, 5370-5382. doi:10.1002/j.1460-2075.1996.tb00921.x

Wang, Z., Su, Y., Zhuang, D. and Lan, T. (2020). The role of EZH2 inhibitor, GSK-126, in seizure susceptibility. *J. Mol. Neurosci.* **71**, 556-564. doi:10.1007/s12031-020-01677-7

Waymack, R., Fletcher, A., Enciso, G. and Wunderlich, Z. (2020). Shadow enhancers can suppress input transcription factor noise through distinct regulatory logic. *eLife* **9**. doi:10.7554/eLife.59351

Williams, D. C., Jr, Cai, M. and Clore, G. M. (2004). Molecular basis for synergistic transcriptional activation by Oct1 and Sox2 revealed from the solution structure of the 42-kDa Oct1.Sox2.Hoxb1-DNA ternary transcription factor complex. *J. Biol. Chem.* **279**, 1449-1457. doi:10.1074/jbc.M309790200

Woolfe, A., Goodson, M., Goode, D. K., Snell, P., McEwen, G. K., Vavouri, T., Smith, S. F., North, P., Callaway, H., Kelly, K. et al. (2005). Highly conserved non-coding sequences are associated with vertebrate development. *PLoS Biol.* **3**, e7. doi:10.1371/journal.pbio.0030007

Yamazaki, T., Souquere, S., Chujo, T., Kobelke, S., Chong, Y. S., Fox, A. H., Bond, C. S., Nakagawa, S., Pierron, G. and Hirose, T. (2018). Functional domains of NEAT1 architectural lncRNA induce paraspeckle assembly through phase separation. *Mol. Cell* **70**, 1038-1053.e1037. doi:10.1016/j.molcel.2018.05.019

Yi, W., Li, J., Zhu, X., Wang, X., Fan, L., Sun, W., Liao, L., Zhang, J., Li, X., Ye, J. et al. (2020). CRISPR-assisted detection of RNA-protein interactions in living cells. *Nat. Methods* **17**, 685-688. doi:10.1038/s41592-020-0866-0

Zerucha, T., Stuhmer, T., Hatch, G., Park, B. K., Long, Q., Yu, G., Gambarotta, A., Schultz, J. R., Rubenstein, J. L. and Ekker, M. (2000). A highly conserved enhancer in the Dlx5/Dlx6 intergenic region is the site of cross-regulatory interactions between Dlx genes in the embryonic forebrain. *J. Neurosci.* **20**, 709-721. doi:10.1523/JNEUROSCI.20-02-00709.2000

Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., Nusbaum, C., Myers, R. M., Brown, M., Li, W. et al. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137. doi:10.1186/gb-2008-9-9-r137

Table S1. 4Cseq-Dlx5/6UCE Bait counts on mouse chr6:

Evf2$^{+/+}$*: Evf2* (+) regulated

Evf2$^{TS/TS}$*: Evf2* (-) regulated

Conserved: *Evf2* (I) independent

Sox2 regulated: *Sox2*$^{fl/fl}$ (cre negative, Sox2 (+) regulated), *Sox2*$^{fl/fl}$*:Dlx5/6cre* (Sox2 (-) regulated

Click here to Download Table S1

Fig_S1

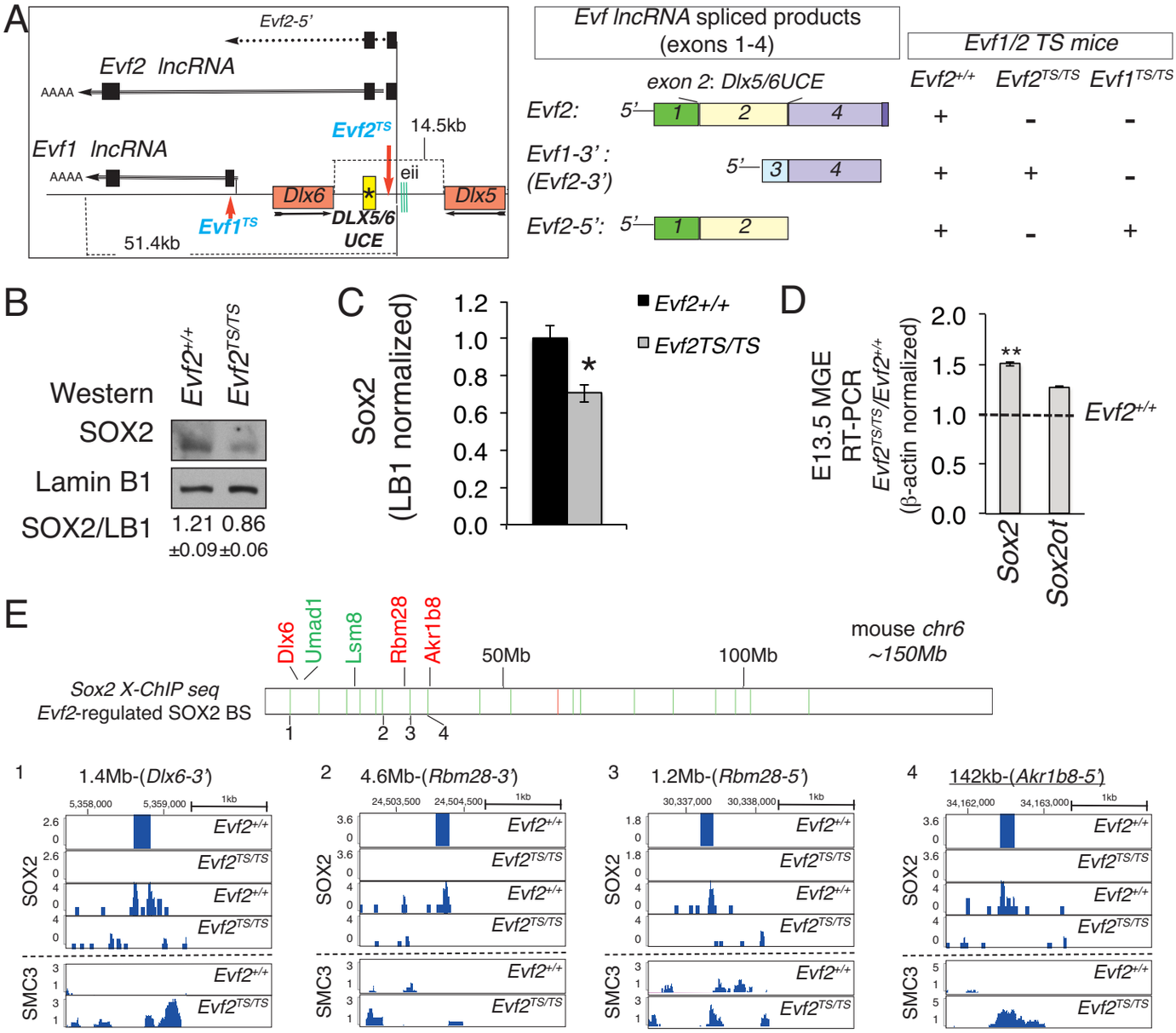


Figure S1. *Evf2* regulation of *Evf2* Sox2 RNP. A. (Left) On mouse chr6, the sites of *Evf1/2* transcription stop insertions (TS, blue) are shown with respect to repressed genes Dlx5 and Dlx6 (red boxes), and polyadenylated and spliced *Evf1, Evf2, and Evf2-5'* transcripts (exons 1-4 in black). *Dlx5/6UCE* sequences are in yellow with star. (Right) Evf contains 4 exons; depending on placement of the TS insertion, different combination of Evf spliced forms are produced. **B.C.**. Western analysis of Sox2 protein levels in *Evf2* expressing (*Evf2*$^{+/+}$) and *Evf2* lacking (*Evf2*$^{TS/TS}$) E13.5GE extracts, (normalized to Lamin B1) shows ~25% decrease in Sox2 protein levels in the absence of *Evf2*. **D.** Taqman qRT-PCR of Sox2 and Sox2ot (a ~118kb overlapping transcript) (RNA levels in *Evf2*$^{+/+}$ and *Evf2*$^{TS/TS}$ E13.5GE shows ~50% increase in Sox2 RNA, but no change in Sox2ot RNA (normalized to β-actin). A-C. n=3/genotype, Student's *t* test. **E.** *Evf2* regulation of *Evf2* RNPs Sox2 and Smc3. Sox2 ChIPseq profiles across mouse chr6 from crosslinked chromatin isolated from *Evf2* expressing (*Evf2*$^{+/+}$) and *Evf2* lacking (*Evf2*$^{TS/TS}$) E13.5GE. *Evf2* positively regulated Sox2 binding sites (BS) (green bars), and negatively regulated (red bars) are shown. *Evf2* repressed gene targets are labeled in red (Dlx, Rbm28, Akr1b8), activated gene targets in green (Umad1, Lsm8). Peaks from sites labelled 1-4 overlap with Evf2 positively regulated Sox2 binding (differential MACS2 (peak calling algorithm used for ChIPseq, Zhang et al. 2008) binding site shown in top two rows) and Smc3 binding peaks. Site 1: Sox2 (+) Smc3 (-), Site 2: Sox2 (+) Smc3 (no peak), Site 3: Sox2 (+) Smc3 binding (but not Evf2 regulated), Site 4:Sox2 (+) Smc3 (-), n=2/genotype, Sites 1-4 top two rows (differential MACS2), rows 3-6 (MACS2. FDR<0.05).

Fig_S2

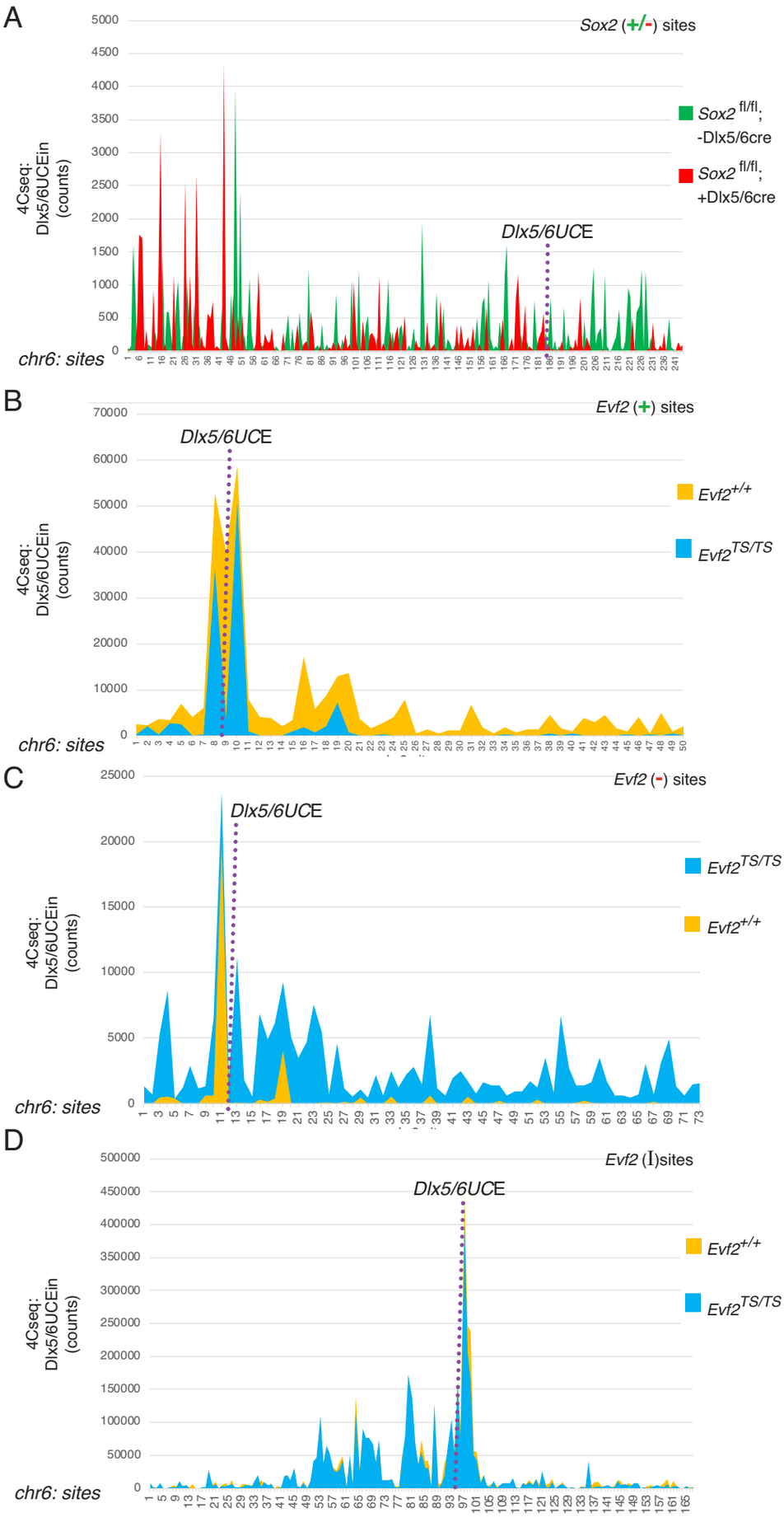


Figure S2. 4Cseq counts identify Sox2 and *Evf2* regulated *Dlx5/6UCEins* across mouse *chr6* in E13.5GEs **A**. *Sox2 (+)* positively regulated *Dlx5/6UCE*ins from Sox2fl/fl;-Dlx5/6cre (presence of Sox2), and Sox2 (-) negatively regulated Dlx5/6UCEins, Sox2fl/fl; +Dlx5/6cre (absence of Sox2). n=4/each genotype 4Cseq intersection of two computational methods (FourCseq and DEseq). Scaled counts for FDR <0.01 are plotted.

B. *Evf2 (+)* positively-regulated *Dlx5/6UCEin* sites (orange) $Evf2^{+/+}$ (presence of *Evf2*), and Evf2 (-) negatively regulated Dlx5/6UCEins (blue) $Evf2^{TS/TS}$ (absence of *Evf2*).

C. *Evf2* (-) negatively regulated sites (blue) $Evf2^{+/+}$ (presence of *Evf2*), and $Evf2^{TS/TS}$ (absence of *Evf2*).

D. Dlx5/6UCEins independent of Evf2 (Evf2 (I) sites). In B-D data is summarized from Cajigas et al 2018.

note:

-*Y-axis:* count numbers from 4Cseq using Dlx5/6UCE as bait

-*X-axis:* numbers correspond to chr6 location sites listed for each genotype (Excel).

-*Dlx5/6UCE* bait location indicated by purple dotted lines

Fig_S3

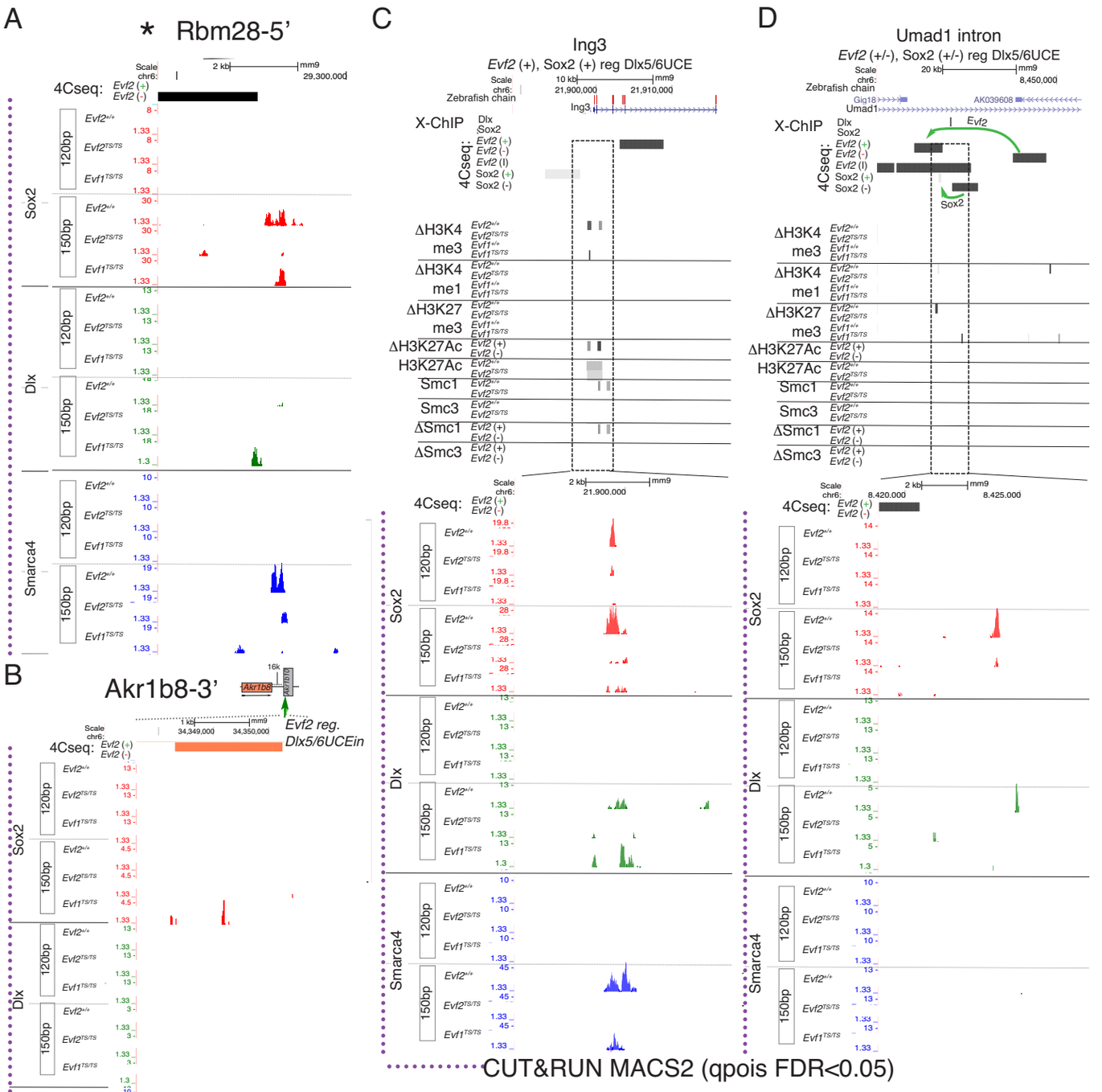
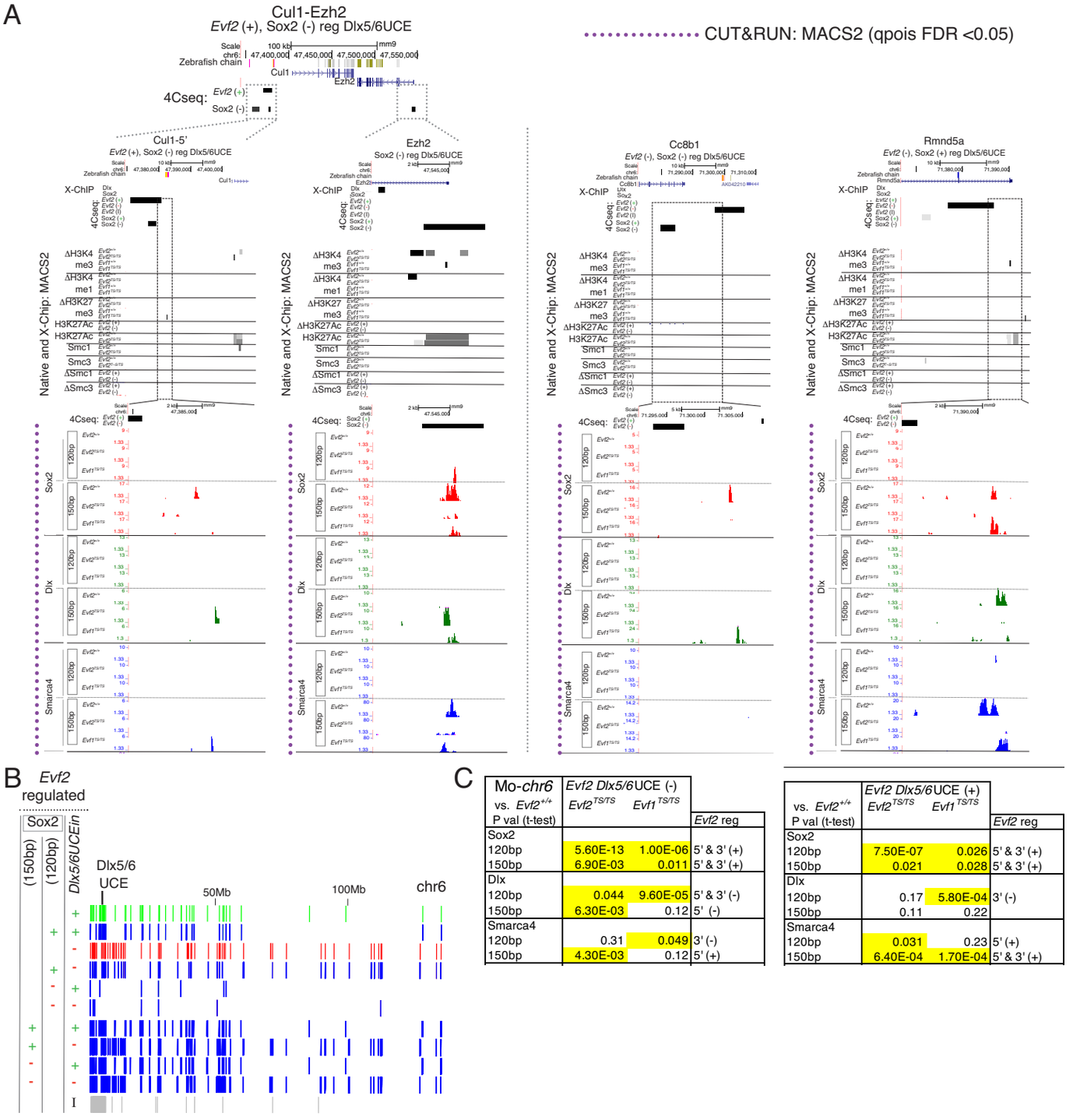


Figure S3. Relationships between *Evf2*-regulated *Dlx5/6UCEin*, RNP binding, and histone modifications at specific sites on mouse *chr6*. CUT&RUN 120bp and 150bp profiles of *Evf2*-regulated binding *Evf2* RNPs Sox2, Dlx, Smarca4, histone modifications (H3K27ac/me3, H3K4me3, H3K4me1), and Smc1a and Smc3 at select *Evf2*-regulated *Dlx5/6UCEin*s from *Evf2*$^{+/+}$, *Evf2*$^{TS/TS}$, and *Evf1*$^{TS/TS}$ E13.5GE. **A.** *Evf2* regulates Sox2 and Smarca4 binding near the *Evf2* negatively regulated *Rbm28-5'-Dlx5/6UCEin*. **B.** *Evf2* regulates Smarca4, but not Sox2 at the *Evf2* positively regulated *Dlx5/6UCEin* located 16kb 3' of *Evf2* repressed target gene Akr1b8. **C-D.** ChIPseq profiles of histone modifications (native chromatin:H3K27me3, H3K4me3, H3K4me1 cross linked: H3K27ac), Smc1a/Smc3 (crosslinked chromatin), Sox2, Dlx, Smarca4 (CUT&RUN) overlapping with Evf2 regulated Dlx5/6UCEins generated from 4Cseq analysis. ChIPseq MACS2 peak (FDR<0.05), n=2-4/genotype, histone modification ChIPseq and Smc1/3 ChiPseq, and 4Cseq FourCseq intersection with DEseq data was based on previously reported in Cajigas et al. 2018.

Fig_S4



Mo-*chr6* vs. *Evf2*$^{+/+}$ P val (t-test)	*Evf2 Dlx5/6*UCE (-) *Evf2*$^{TS/TS}$	*Evf1*$^{TS/TS}$	*Evf2* reg
Sox2			
120bp	5.60E-13	1.00E-06	5' & 3' (+)
150bp	6.90E-03	0.011	5' & 3' (+)
Dlx			
120bp	0.044	9.60E-05	5' & 3' (-)
150bp	6.30E-03	0.12	5' (-)
Smarca4			
120bp	0.31	0.049	3' (-)
150bp	4.30E-03	0.12	5' (+)

vs. *Evf2*$^{+/+}$ P val (t-test)	*Evf2 Dlx5/6*UCE (+) *Evf2*$^{TS/TS}$	*Evf1*$^{TS/TS}$	*Evf2* reg
Sox2			
120bp	7.50E-07	0.026	5' & 3' (+)
150bp	0.021	0.028	5' & 3' (+)
Dlx			
120bp	0.17	5.80E-04	3' (-)
150bp	0.11	0.22	
Smarca4			
120bp	0.031	0.23	5' (+)
150bp	6.40E-04	1.70E-04	5' & 3' (+)

Figure S4. *Evf2*-regulated *Dlx5/6UCEin*, RNP binding, and histone modifications on mouse *chr6*. CUT &RUN profiles of *Evf2*-regulated *Evf2* RNPs Sox2, Dlx, Smarca4, histone modifications (H3K27ac/me3, H3K4me3, H3K4me1) and *Evf2*-regulated *Dlx5/6UCEin*s from *Evf2*$^{+/+}$, *Evf2*$^{TS/TS}$, and *Evf1*$^{TS/TS}$ E13.5GE. **A.** Relationships between ChIPseq profiles of histone modifications (crosslinked and native chromatin), Smc1a/Smc3 (crosslinked chromatin), and Sox2, Dlx, Smarca4(CUT&RUN, native chromatin) overlapping with *Evf2* regulated *Dlx5/6UCEins* generated from 4Cseq analysis. **B.** *Evf2* regulated Sox2 binding sites overlapping with *Evf2* positively regulated (+) and negatively regulated (-) and independent *Dlx5/6UCEins* across *chr6*. **C**. Statistically significant differences (yellow highlights) in RNP binding (CUT&RUN ChIPseq) at *Evf2* positively and negatively regulated *Dlx5/6UCEins* (4Cseq) from *Evf2*$^{+/+}$, *Evf2*$^{TS/TS}$, and *Evf1*$^{TS/TS}$ E13.5GE.

Fig_ S5

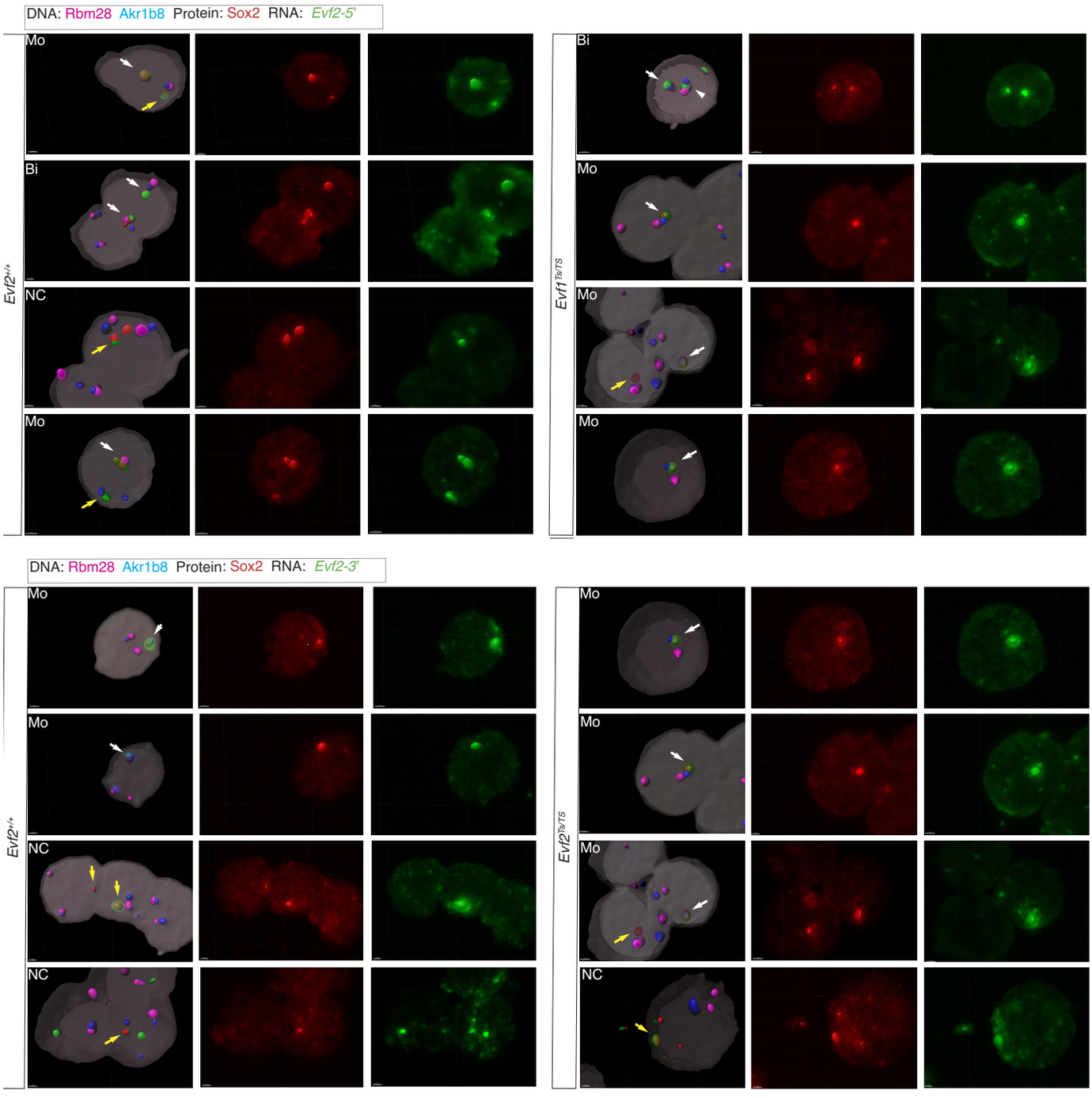


Figure S5. Additional examples of Evf2 RNA cloud and Sox2 protein pool association with re-pressed target genes Akr1b8 and Rmb28 in Evf2 mutants. Imaris 3D reconstructions based on confocal analysis in *Evf2*$^{+/+}$, *Evf2*$^{TS/TS}$, *Evf1*$^{TS/TS}$ E13.5 GE nuclei: FISH (fluorescent in situ hybridization of DNA probes) Rbm28 DNA (pink), Akr1b8 DNA (blue), Evf2-5' RNA (green), Evf2-3' RNA (green) and detection of Sox2 protein using anti-Sox2 antibody: Sox2 protein pools (red). Mo indicates monoallelic, Bi indicates biallelic, white arrows indicate colocalization of RNA or protein with a DNA target, yellow arrows indicate RNA or protein not colocalized (NC) with a DNA target.

Fig_ S6

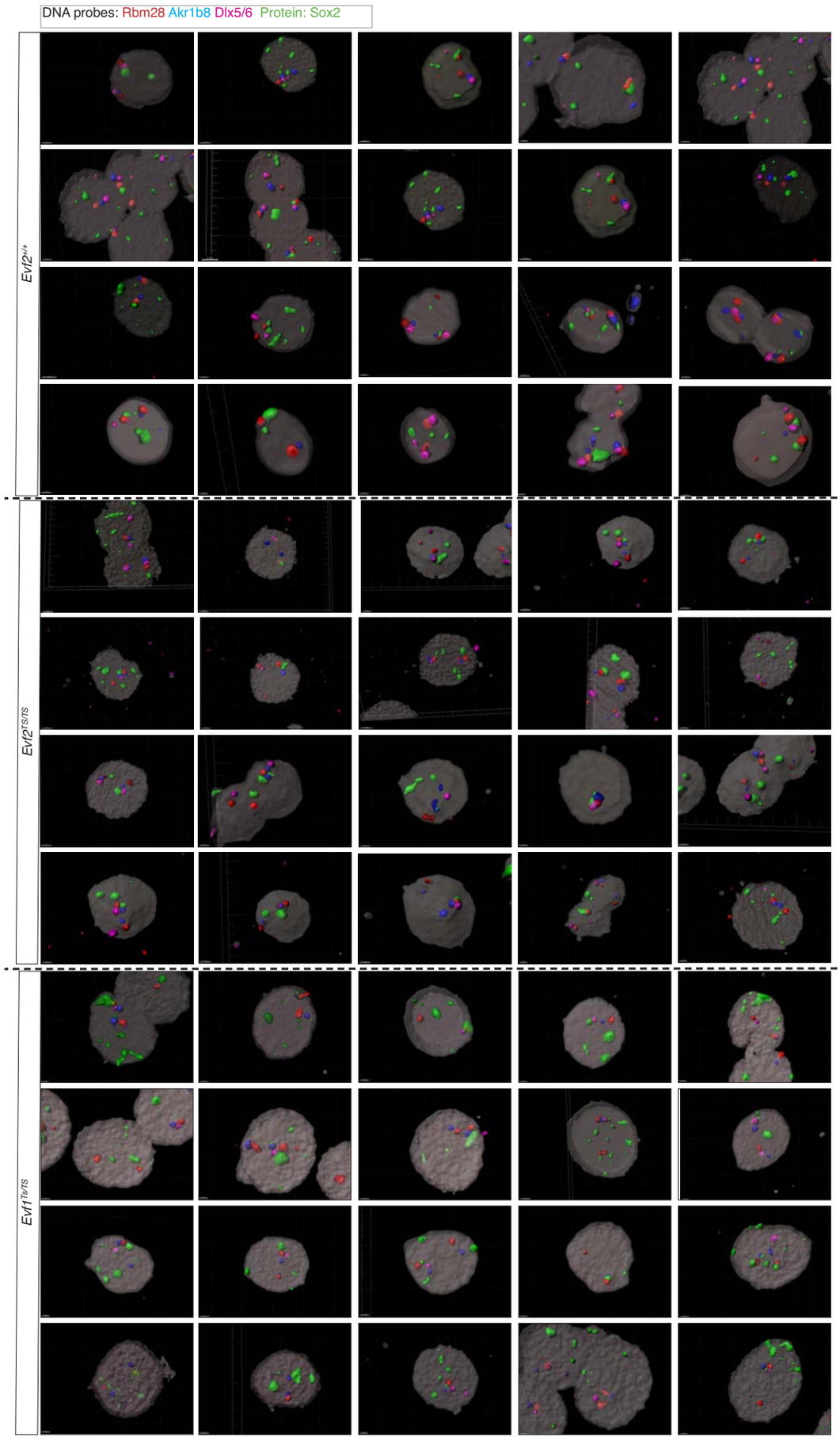


Figure S6. Additional examples of Sox2 protein pools associated with repressed gene targets Rbm28 and Akr1b8 and Dlx5/6 in Evf2 mutants. Imaris 3D reconstructions based on Z-stacks from confocal analysis in *Evf2*$^{+/+}$, *Evf2*$^{TS/TS}$, *Evf1*$^{TS/TS}$ E13.5 GE nuclei: FISH (fluorescent in situ hybridization of DNA probes) Rbm28 DNA (red), Akr1b8 DNA (blue), Dlx5/6 DNA (pink), and detection of Sox2 protein using anti-Sox2 antibody: Sox2 protein pools (green).