Sequentially acting Sox transcription factors in neural lineage development

Maria Bergsland,[1] Daniel Ramsköld,[1] Cécile Zaouter, Susanne Klum, Rickard Sandberg,[2] and Jonas Muhr[2,3]

Ludwig Institute for Cancer Research, Department of Cell and Molecular Biology, Karolinska Institutet, SE-171 77 Stockholm, Sweden

Pluripotent embryonic stem (ES) cells can generate all cell types, but how cell lineages are initially specified and maintained during development remains largely unknown. Different classes of Sox transcription factors are expressed during neurogenesis and have been assigned important roles from early lineage specification to neuronal differentiation. Here we characterize the genome-wide binding for Sox2, Sox3, and Sox11, which have vital functions in ES cells, neural precursor cells (NPCs), and maturing neurons, respectively. The data demonstrate that Sox factor binding depends on developmental stage-specific constraints and reveal a remarkable sequential binding of Sox proteins to a common set of neural genes. Interestingly, in ES cells, Sox2 preselects for neural lineage-specific genes destined to be bound and activated by Sox3 in NPCs. In NPCs, Sox3 binds genes that are later bound and activated by Sox11 in differentiating neurons. Genes prebound by Sox proteins are associated with a bivalent chromatin signature, which is resolved into a permissive monovalent state upon binding of activating Sox factors. These data indicate that a single key transcription factor family acts sequentially to coordinate neural gene expression from the early lineage specification in pluripotent cells to later stages of neuronal development.

During development of the CNS, neurons and glia are generated from self-renewing neural progenitor cells (NPCs) that are directed to leave the cell cycle, down-regulate progenitor identities, and activate neuronal or glial gene expression in a spatially and temporally defined manner. The mechanisms regulating gene expression in NPCs and their differentiated progeny have been extensively characterized, but it is largely unknown how neural lineage-specific gene expression programs are initially specified and activated during the course of differentiation.

An important property of pluripotent stem cells is their capacity to induce gene programs characteristic of all cell lineages. Previous studies in embryonic stem (ES) cells have demonstrated that many genes destined to become activated at later stages of development are already bound by ES cell regulatory transcription factors, including Sox2, Oct4, Nanog, and FoxD3 (Boyer et al. 2005; Lee et al. 2006; Xu et al. 2009). Moreover, genes poised for activation are often associated with bivalent histone domains consisting of repressive histone modifications combined with

modifications associated with transcriptional activation (H3K27me3 and H3K4me3) (Boyer et al. 2005; Bernstein et al. 2006; Lee et al. 2006). Bivalent histone marks are subsequently resolved as genes become activated or terminally repressed during development (Bernstein et al. 2006; Mikkelsen et al. 2007; Mohn et al. 2008). Together, these findings indicate that many silent genes in ES cells are prebound by transcription factors and epigenetically prepared for activation, but they do not demonstrate how lineage-specific gene expression programs are initially selected and later activated. Insights into these questions come from studies of the liver-specific enhancer *Alb1* (Xu et al. 2007; Zaret et al. 2008). In ES cells, the *Alb1* enhancer is prebound by FoxD3, which ensures the assembly of permissive chromatin (Xu et al. 2009). Interestingly, upon endodermal differentiation, FoxD3 binding is replaced by FoxA1, which helps to induce Alb1 expression in a liver-specific manner (Xu et al. 2009). Studies of the B-cell-specific *λ5-VpreB1* locus constitute an additional example of how a transcription factor prepares the enhancer for later activation by an alternative member of the same transcription factor family. This locus contains an intergenic enhancer to which Sox2 binds and adds an epigenetic active mark in ES cells (Liber et al. 2010). In pro-B cells, Sox2 binding is replaced by Sox4, which leads to the activation of λ5 expression (Liber et al. 2010). Although

these studies indicate the importance of pioneering functions of transcription factors, experiments are focused on specific enhancers in individual genes, and it remains unclear whether the sequential regulatory strategy is a more general requirement for activation of larger sets of gene batteries in differentiating cell lineages.

Apart from the above-mentioned gene regulatory functions in ES cells and roles in early B-cell development, transcription factors of the Sox gene family have important sequential roles in regulating the maintenance and differentiation of progenitor cells from early pluripotent stages to late steps of neurogenesis (Guth and Wegner 2008). Sox2 is necessary for the establishment and maintenance of ES cells (Avilion et al. 2003). All three SoxB1 proteins (Sox1, Sox2, and Sox3) are expressed in most neural precursors in both the developing and adult CNS, and studies conducted in chick and mouse embryos demonstrate that they act redundantly to maintain neural cells in a progenitor state and counteract neuronal differentiation (Bylund et al. 2003; Graham et al. 2003; Favaro et al. 2009). The SoxC proteins (Sox4, Sox11, and Sox12) are expressed complementary to Sox1–3 in the developing CNS and can mostly be detected in post-mitotic differentiating neurons (Fig. 1A; Bergsland et al. 2006; Hoser et al. 2008). Misexpression experiments in chicks demonstrate that SoxC proteins have the opposite function compared with Sox1–3 and can induce the expression of neuronal proteins (Bergsland et al. 2006; Hoser et al. 2008), whereas deletion of the SoxC proteins in the embryonic mouse spinal cord leads to a significant decrease in differentiated neurons and an associated increased cell death (Bhattaram et al. 2010; Thein et al. 2010).

Despite the importance of Sox factors during the course of neural development, there is very limited information concerning the control of appropriate gene expression programs that are activated in CNS progenitors and their differentiated progeny. This is partly due to the limited number of identified Sox target genes. In this study, we analyzed Sox transcription factors during neural lineage development by generating and comparing genome-wide binding data for Sox2, Sox3, and Sox11 from early lineage specification stages with the onset of neuronal gene expression. The data indicate that sequentially acting Sox transcription factors control neural lineage-specific gene expression by predisposing gene programs to become activated in NPCs and during neuronal and glial differentiation.

Results

SoxB1 and SoxC proteins share a high number of target genes

In the developing mouse CNS, SoxB1 proteins are expressed in the majority of all NPCs (Fig. 1A), whereas SoxC proteins are generally confined to post-mitotic differentiating neurons (Fig. 1A). The finding that SoxB1 maintain neural progenitors, whereas SoxC proteins promote the expression of differentiated neuronal proteins, raises the question of whether their opposite activities are mediated via the regulation of distinct or common sets of target genes. To examine this issue, we explored stage- and factor-specific genome-wide binding of Sox proteins during neurogenesis by employing chromatin immunoprecipitation
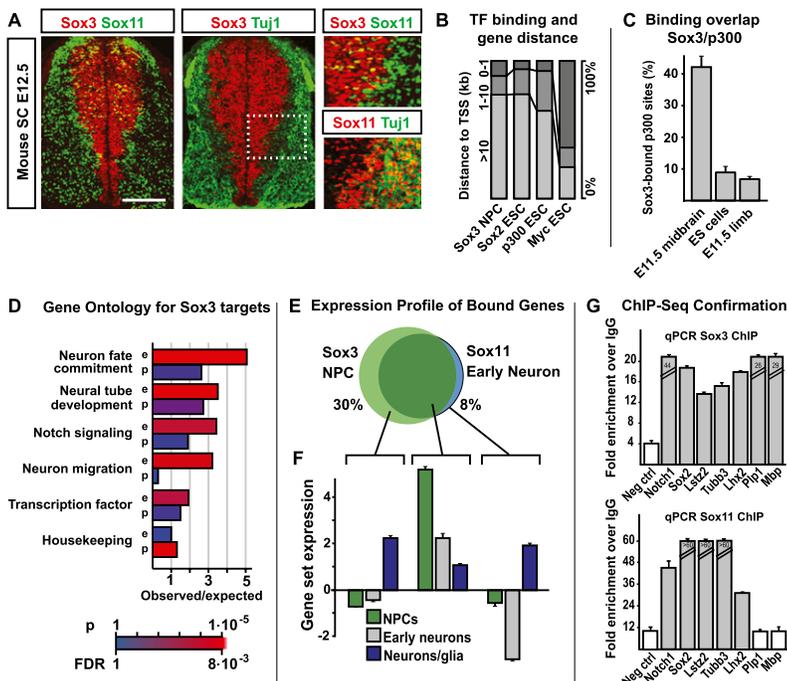
Figure 1. Genome-wide binding maps of Sox factors in neural development. (A) Expression of Sox3, Sox11, and the neuronal protein Tuj1 in developing mouse spinal cord. Bars: A, 100 μm; 40 μm. (B) Localization of genome-wide Sox3 binding relative to annotated TSSs. Percentages of sites located within 1 kb, 1–10 kb, and >10 kb from a TSS are shown for Sox3 in NPCs, as well as Sox2, p300, and Myc in ES cells. (C) Percentage overlap between Sox3 NPC peaks and p300 peaks in mouse embryonic (embryonic day 11.5 [E11.5]) brain, limb, and ES cells. Bound regions were considered overlapping if the distance between peak centers was <300 base pairs (bp). (D) Genes with Sox3 binding within enhancer regions (e) were enriched for developmentally significant functions, whereas genes with Sox3 binding in promoter regions (p) were enriched for housekeeping functions. (E) Venn diagram showing the overlap in target sites between Sox3 in NPCs and Sox11 in early-formed neurons. (F) Gene set expression for Sox3-specific genes, Sox11-specific genes, and genes bound by both Sox3 and Sox11 in E using microarray data of NPCs, PSA-NCAM1[+] early neurons, and adult neurons and glia. (G) Confirmation of Sox3 and Sox11 ChIP-seq peaks using ChIP-qPCR analysis. White bars indicate nonbound regions from ChIP-seq experiments. Error bars correspond to standard deviation of three qPCR replicates. Bars >60 indicate fold enrichment over an undetectable IgG signal after 50 cycles of PCR.

(ChIP) combined with massively parallel sequencing (ChIP-seq). Using specific antibodies for Sox2, Sox3, and Sox11 (Supplemental Fig. 1), and mouse ES cell-derived NPCs or neurons as cellular sources, we generated nine different ChIP-seq characterizations (three biological replicates per Sox factor), resulting in thousands of significant binding regions (peaks) per factor (Supplemental Tables 1–4). Individual ChIP-seq replicates showed high concordance, and overall Sox binding followed developmental stage (Supplemental Fig. 2). In all Sox ChIP-seq experiments, a Sox motif (van Beest et al. 2000; Maruyama et al. 2005) was found most highly enriched (Supplemental Table 5). Interestingly, we observed that the core motif deviated slightly when extracted from Sox2 targeted regions compared with regions targeted by Sox3 and Sox11 (Supplemental Fig. 3A). In a previous study (Engelen et al. 2011), the binding of Sox2 in ES cell-derived NPCs was reported to be enriched at promoter regions. In contrast, we identified the majority of Sox-binding regions >10 kb from the closest transcription start site (TSS) (Fig. 1B; data not shown), which is similar to the binding pattern of Sox2 in ES cells (Fig. 1B; Supplemental Fig. 4; Chen et al. 2008; Marson et al. 2008). These distal regions were significantly enriched for conserved Sox motifs, whereas peaks in promoter regions (<1 kb to the TSS) had a lower frequency of associated Sox motifs (Supplemental Fig. 3B–D) and were enriched for housekeeping genes (Fig. 1D).

To characterize Sox-bound regions for possible enhancer functions, we compared the binding of Sox3 with the binding of the transcriptional coactivator p300, which has been shown to accurately predict active enhancers in the developing embryo in a tissue-specific manner (Visel et al. 2009). Notably, >40% of the Sox3-bound regions overlapped with enhancers marked by p300 in embryonic brains, but not in ES cells or limb tissue (Fig. 1C). Similar trends were obtained by comparing the binding pattern of p300 with that of Sox11 (data not shown). Thus, many Sox3- and Sox11-bound regions appear to function as active enhancers in the developing CNS. In line with this finding, genes associated with distal Sox3 peaks (closest neighbor) (see the Materials and Methods) were enriched for those encoding transcription factors and proteins involved in CNS development (Fig. 1D; Supplemental Table 6). These genes include, for instance, many genes involved in neurogenesis (including Sox factors, components of the Notch signaling pathway, and proneural genes), genes encoding secreted molecules (such as Shh and members of the FGF, Wnt, and TGFβ families), and transcription factors implicated in cell type specification (including members of the Nkx and Pax transcription factor families). However, Sox3 binding was also observed at genes with selective expression in Sox3-negative differentiated neurons (such as *Tubb3*, *SCG10*, *Lhx2*, and *Pax2*) (Supplemental Table 7).

Since Sox3 binding could be detected at genes expressed in NPCs but also at many genes expressed in Sox3$^-$/Sox11$^+$ neurons, we next determined the degree of overlap in the binding of Sox proteins acting at different stages of neurogenesis. The binding of the SoxB1 proteins Sox2 and Sox3, which act redundantly in NPCs (Bylund et al. 2003), overlapped extensively, and Sox3 occupied the majority (96%) of the Sox2-bound sites in NPCs. Surprisingly, Sox3 binding could also be identified at most (92%) of the sites later targeted by Sox11 in newly formed neurons (Fig. 1E), and Sox3 and Sox11 were only uniquely bound to 30% and 8% of their targets, respectively (Fig. 1E). Thus, despite their differential expression and activities during neurogenesis, the binding of SoxB1 and SoxC factors overlaps extensively.

The finding that Sox3 and Sox11 share many of their bound regions raises the question how their unique and common target genes are expressed during neurogenesis. To address this issue, we separated genes bound by Sox3 from genes bound by both Sox3 and Sox11 as well as from genes bound by Sox11 alone. These gene sets were compared with publicly available expression profiles for NPCs, differentiating neurons, and late populations of neurons/glia. These comparisons revealed that genes uniquely targeted by Sox3 are most significantly expressed in late populations of neurons/glia (Fig. 1F; Supplemental Fig. 6), whereas genes first bound by Sox3 and later by Sox11 could mainly be detected in expression profiles for NPCs and/or populations of neurons/glia (Fig. 1F). Genes occupied by Sox11 only had the highest expression in late neurons/glia (Fig. 1F). The common binding of Sox3 and Sox11 was validated for a subset of these genes using ChIP and quantitative PCR (ChIP-qPCR) (Fig. 1G; Supplemental Fig. 5). Together, these analyses demonstrate that Sox3 occupies many genes with vital regulatory functions in NPCs but also many silent genes, of which several are subsequently activated and bound by Sox11 in differentiating neurons (Supplemental Table 7).

Forced expression of SoxB1 and SoxC proteins up-regulates NPC genes and neuronal genes, respectively

The finding that both Sox3 and Sox11 target genes with a restricted expression pattern in NPCs or neurons/glia raises the question about their gene regulatory role in mammalian neurogenesis. To begin to address this issue, we next characterized the activity of Sox3 and Sox11 in NPCs. Mouse ES cell-derived NPCs cultured for 12 d under differentiating conditions had generally suppressed the expression of the progenitor marker Sox1 (Fig. 2A,D) and instead up-regulated expression of the pan-neuronal marker Tuj1 (Fig. 2A,D) or the glial proteins GFAP and S100β (Fig. 2B,C). In contrast, NPCs stably overexpressing exogenous Sox3 under the control of the neural-specific Nestin enhancer (Lothian and Lendahl 1997) were maintained in a self-renewing and undifferentiated Sox1$^+$ state (Fig. 2E,H; data not shown), and only few cells had up-regulated Tuj1, GFAP, or S100β expression even after 12 d under differentiating conditions (Fig. 2E–H). siRNA-mediated knockdown of Sox2 and Sox3 expression in NPCs increased the rate of neuronal differentiation, and after 4 d, many cells were Tuj1$^+$ and devoid of Sox1 expression (Supplemental Fig. 7). Misexpression of a Myc-tagged version of Sox11 had the opposite activity compared with Sox3, and misexpression in NPCs for 24 h resulted in the induction of Tuj1 expression (Fig. 2L,M), whereas only a few Tuj1-expressing cells could be detected in the control
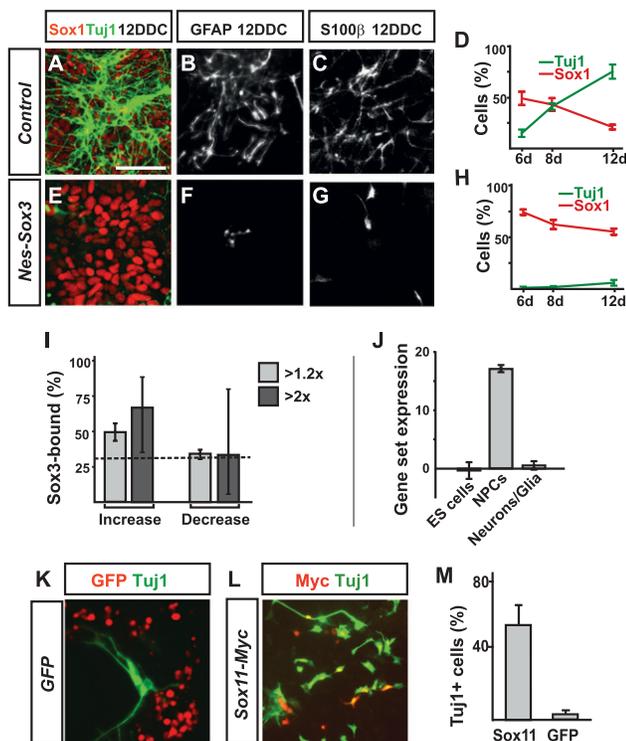
Figure 2. Function of Sox factors in neurogenesis. (*A–C*) Expression of the neural progenitor marker Sox1, the pan-neuronal marker Tuj1, and the astrocytic markers GFAP and S100β in ES cell-derived differentiating neurons and glia after 12 d in differentiation conditions (DDC). (*D*) The fraction of NPCs that expresses Sox1 or up-regulated Tuj1 after 6, 8, and 12 DDC. (*E–G*) Expression of Sox1, Tuj1, GFAP, and S100β in Sox3 overexpressing ES cell-derived NPCs (Nes-Sox3) after 12 DDC. (*H*) The fraction of Nes-Sox3 NPCs expressing Sox1 and Tuj1 after 6, 8, and 12 DDC. (*I*) Percent of up-regulated and down-regulated genes (identified by Sox3 overexpression microarray experiment at fold change levels 2 and 1.2) that are bound by Sox3 (ChIP-seq experiment). Error bars represent 95% confidence intervals. The dashed line denotes the expected fraction of Sox3 binding. (*J*) Gene set expression profile of genes that were both bound by Sox3 and up-regulated in the Sox3-overexpressing NPCs. Error bars indicate standard error of the mean. (*K–M*) Sox11-misexpressing NPCs show Tuj1 expression in 52% of the transfected cells at 20 h post-transfection (*L,M*) compared with <5% of GFP transfected NPCs (*K,M*). Bars: *A–C,E–G*, 20 μm; *K,L*, 40μm.

cells (Fig. 2K,M). The glial protein GFAP could not be detected in Sox11-expressing NPCs either 24 h or 72 h after transfection (data not shown). Together, these analyses demonstrate that despite the high number of common target genes, Sox3 and Sox11 have opposite gene regulatory activities and promote NPC- and neuronal-specific gene expression, respectively.

To get a more complete understanding of how Sox3 regulates targeted genes in NPCs, we next used global expression profiling to analyze gene expression in NPCs stably overexpressing Sox3. These analyses revealed ~350 up-regulated genes and >800 down-regulated genes. Among the activated genes, there was a strong enrichment

for genes previously identified as Sox3 targets (Fig. 2I). No such correlation could be identified among the genes down-regulated (Fig. 2I), consistent with the function of Sox3 as a transcriptional activator. Moreover, comparisons of the up-regulated genes with expression profiles for ES cells, NPCs, and neurons/glia confirmed that Sox3 activates mainly genes that are most highly expressed in NPCs (Fig. 2J). In line with these findings, siRNA-mediated knockdown of Sox2 and Sox3 expression suppressed NPC genes and lead to an up-regulation of genes expressed by neurons and glia (Supplemental Fig. 7). Thus, despite the binding of a large number of neuronal and glial genes, Sox3 activates primarily genes expressed in NPCs.

Competitive Sox3 and Sox11 activities at the transition between NPCs and post-mitotic neurons

To further explore the regulatory activities conferred by Sox3 and Sox11 on sequentially targeted genes, we next generated reporter constructs containing putative enhancers of the neuronal genes *Lhx2*, *Pax2*, and *Tubb3* (Fig. 3A–C). Mouse genomic DNA fragments, defined by our Sox3 and Sox11 ChIP-seq analyses, were isolated and cloned into reporter vectors consisting of the minimal thymidine kinase (TK) promoter and the reporter gene luciferase or the minimal β-globin promoter and the reporter gene *EGFP*. EGFP reporter expression was used to determine enhancer activity 45 h after electroporation into the neural tube of Hamburger Hamilton (HH) stages 10–12 chick embryos, whereas a ubiquitously active *CMV/TK-LacZ* vector was used as an internal expression control. Out of four isolated DNA fragments (see the Materials and Methods), three showed enhancer activity and could drive EGFP expression in the chick spinal cord, whereas no EGFP expression could be detected in chick embryos electroporated with reporters containing DNA fragments corresponding to the alternative Sox peak of the *Pax2* gene (see the Material and Methods). Notably, the activity of the enhancers was confined to Sox3$^-$/Sox11$^+$ differentiating neurons (Fig. 3D–F; data not shown) and recapitulated the expression pattern of their respective neuronal genes along the medial–lateral axis of the neural tube (Fig. 3A–C). Coelectroporation of the reporter constructs together with Sox11 expression vectors resulted in a broad EGFP activation throughout the transfected neural tubes (Fig. 3G–I). Moreover, all three luciferase reporters were activated in P19 cells or COS1 cells (data not shown) in the presence of Sox11 expression but not Sox3 (Fig. 3J–L), showing that the Sox3- and Sox11-bound genomic regions of the neuronal genes *Tubb3*, *Lhx2*, and *Pax2* can function as Sox11-activated neuronal enhancers both in vitro and in vivo. The presence of Sox3 expression efficiently suppressed Sox11-mediated reporter activation (Fig. 3J–L). This suppression could result from either competitive binding between Sox3 or Sox11 for the same DNA motif or Sox3's function as an active repressor on these enhancers. To discriminate between these two possibilities, we expressed three derivatives of Sox3: the DNA-binding HMG domain alone, the HMG domain fused to the repressor domain of *Drosophila* Engrailed protein (HMG-EnR) (Bylund et al.
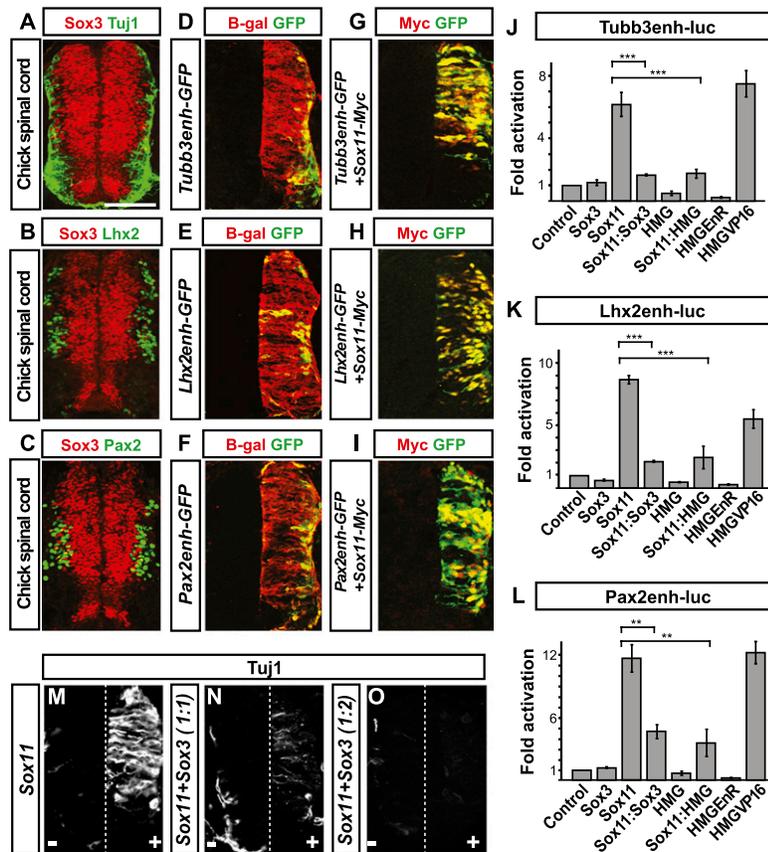
Figure 3. Competitive Sox binding at neuronal genes. (*A–C*) Expression of Sox3 and the neuronal proteins Tuj1 (*A*), Lhx2 (*B*), and Pax2 (*C*) in developing chick spinal cord. (*D–F*) Sox3- and Sox11-bound enhancers of the neuronal genes *Tubb3* (*D*), *Lhx2* (*E*), and *Pax2* (*F*) can drive the expression of a GFP reporter in post-mitotic neurons of the electroporated chick spinal cord. Chick embryos were electroporated at HH stages 9–11 and harvested after 45 h of incubation. β-Galoctosidase represents electroporation control. (*G–I*) Cotransfection of a Sox11-Myc expression vector broadly activated all GFP reporters in *D–F* throughout the electroporated neural tube. Chick embryos were electroporated at HH stages 10–12 and harvested after 24 h of incubation. (*J–L*) Transactivation assays in P19 cells with *Tubb3-luc* (*J*), *Lhx2-luc* (*K*), or *Pax2-luc* (*L*) reporter constructs in the presence of vectors expressing Sox11, Sox3, or the HMG domain of Sox3 either alone or fused to the EnR repression domain or VP16 activation domain. Results are represented as mean ± SEM from three to nine experiments. (**) $P < 0.01$; (***) $P < 0.001$. (*M–O*) Tuj1 expression in chick spinal cord 24 h after electroporation (at HH stages 9–11) with Sox11 alone (*M*) or together with Sox3 at a ratio of 1:1 (*N*) or 1:2 (*O*). A plus sign (+) denotes the electroporated side of the spinal cord and a minus sign (−) denotes the control side. Bars: *A–I*, 50 μm; *M–O*, 40 μm.

2003), and the HMG domain fused to the activator domain of the viral protein VP16 (HMG-VP16) (Bylund et al. 2003). The HMG domain alone behaved as full-length Sox3 and blocked Sox11 activity (Fig. 3J–L). The HMG-EnR variant efficiently repressed the activity of the luciferase reporters, whereas these reporters were activated by HMG-VP16 (Fig. 3J–L). Hence, the ability of Sox3 to suppress Sox11-mediated activation can be mimicked by its DNA-binding HMG domain, indicating that Sox3 blocks Sox11 through competitive DNA binding.

To further examine whether the prebinding of Sox3 to neuronal genes affects their later activation by Sox11, Sox11 was either misexpressed alone in chick neural tubes or together with increasing amounts of Sox3. Misexpression of Sox11 for 24 h resulted in a strong ectopic expression of Tuj1 on the electroporated side of the neural tube (Fig. 3M; Bergsland et al. 2006; Hoser et al. 2008). Sox3 counteracted this induction, and the capacity of Sox11 to induce ectopic Tuj1 expression was completely abolished when electroporated together with the doubled amount of Sox3 expression vectors (Fig. 3N,O). Together, these results suggest that despite the short period during neurogenesis at which Sox3 overlaps with the expression of Sox11, one role of Sox3 prebinding may be to prevent premature Sox11-mediated induction of neuronal genes until differentiating cells have down-regulated progenitor-specific gene expression.

Sox3 binding establishes epigenetic changes

In ES cells, Sox2 binds many silent genes that are induced later during development. Several of these genes are associated with activating histone modifications (H3K4me3) as well as repressive histone modifications (H3K27me3) (Boyer et al. 2005; Lee et al. 2006). The finding that Sox3 targets genes in NPCs that are first activated upon neuronal differentiation and Sox11 binding prompted us to examine histone modifications associated with these genes. Using sequential ChIP-qPCR, we found that Sox3-bound genes expressed in NPCs (Fig. 4A) were associated with H3K4me3 only (Fig. 4B). In contrast, Sox3-bound neuronal and glial genes, which are silent in NPCs (Fig. 4A), were associated both with activating and repressing histone marks (Fig. 4B). Notably, the bivalent chromatin profiles at neuronal genes were resolved into monovalent H3K4me3$^+$ domains as the genes become occupied and activated by Sox11 in early neurons (Fig. 4C,D), whereas the binding of Sox11 to NPC genes was associated with a replacement of H3K4me3 with H3K27me3 (Fig. 4D). Thus, in NPCs, Sox3-bound neuronal genes are associated with bivalent chromatin, which are resolved into a monovalent active state upon the binding of Sox11.

It is possible that Sox3 participates in defining the epigenetic status of chromatin by conferring changes in histone modifications. To address how Sox3 influences the presence of histone modifications, we expressed Sox3
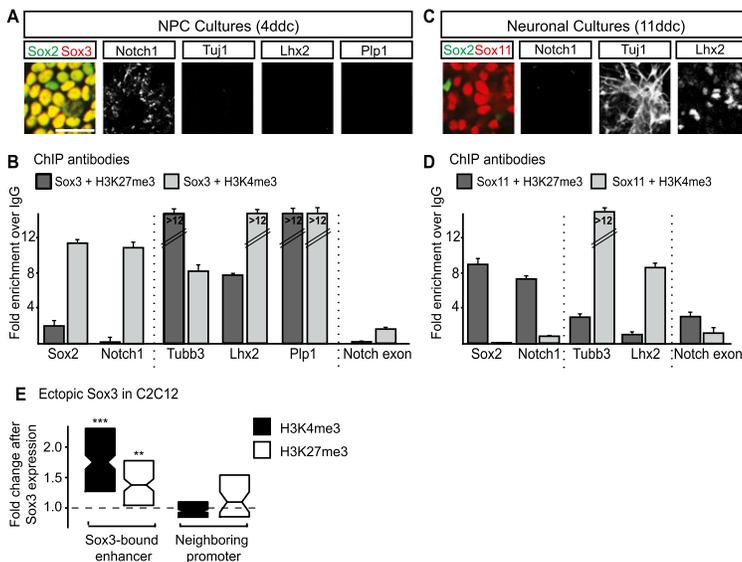
Figure 4. Active histone modifications associated with Sox3 binding. (*A*) Expression of NPC proteins (Sox2 and Notch1), neuronal proteins (Tuj1 and Lhx2), and the glial protein Plp1 in Sox3-expressing ES cell-derived NPCs (4 DDC). (*B*) Histone modifications of Sox3 targeted genes were measured by sequential ChIP experiments. Chromatin precipitation of Sox3-bound regions in NPCs (4 DDC, shown in *A*) were followed by H3K4me3- or H3K27me3-specific chromatin precipitations and qPCR analysis. (*C*) Expression of NPC proteins (Sox2 and Notch1) and neuronal proteins (Tuj1 and Lhx2) in Sox11-expressing ES cell-derived neurons (11 DDC). (*D*) Histone modifications of Sox11 targeted genes were measured by sequential ChIP experiments. Chromatin precipitation of Sox11-bound regions in neurons (11 DDC, shown in *C*) was followed by H3K4me3- or H3K27me3-specific chromatin precipitations and qPCR analysis. As the Plp1 gene is not bound by Sox11, it was excluded from this analysis. Error bars represent the standard deviation of triplicate qPCR measurements from one representative ChIP experiment out of three. Samples denoted with >12-fold change had no detectable IgG signal after 50 cycles of PCR. Bars: *A*,*C*, 15 μm. (*E*) Fold change in histone modifications in C2C12 cells, shown as box plots, at all enhancers bound by ectopic Sox3 or at their neighboring promoters. Asterisks indicate positive correlation between fold change in methylation and Sox3-binding strength ([**] $P < 0.01$; [***] $P < 0.001$), giving further support of direct effects.

ectopically in C2C12 mesodermal progenitors, which are normally devoid of SoxB1 gene expression (our unpublished observation). Twenty-four hours after transfection, cells were harvested and genome-wide ChIP-seq data for Sox3 as well as H3K4me3 and H3K27me3 were generated. Importantly, at Sox3-bound enhancers, we could identify a significant increase in H3K4me3 and, to a lesser degree, also H3K27me3 (Fig. 4E). At Sox3-bound promoters, no significant change in either H3K4me3 or H3K27me3 could be detected (data not shown). Sox3 enhancer binding also did not lead to an alteration in the levels of H3K4me3 or H3K27me3 at neighboring promoter regions (Fig. 4E). Hence, by regulating the presence of histone modifications, Sox3 appears to have the capacity to induce local epigenetic changes at targeted enhancers.

Sox2 preselects neural lineage-specific gene programs in pluripotent cells

The finding that Sox3 prebinds genes that are later targeted and activated by Sox11 during neuronal differentiation raises the question of whether Sox3-activated NPC genes, in a similar manner, are prebound by alternative Sox proteins already in pluripotent stem cells. To address this issue, we compared the binding of Sox2 in ES cells (Chen et al. 2008; Marson et al. 2008) with the binding of Sox3 in NPCs. Out of ~9000 significant peaks identified for Sox3 in NPCs, nearly 50% mapped to regions also bound by Sox2 in ES cells (Fig. 5A). The overlapping binding pattern of Sox2 in ES cells and Sox3 in NPCs could reflect either a sequential binding of Sox2 and Sox3 to genes that are expressed in both ES cells and NPCs or that Sox2 prebinds genes that are later occupied and activated by Sox3 during neural lineage development. To address these possibilities, we separated genes bound only by Sox2 in ES cells (closest

neighbor; 1373 genes) from genes bound by both Sox2 in ES cells and Sox3 in NPCs (1532 genes) and genes bound by Sox3 alone (2474 genes) and analyzed their expression in ES cells, NPCs, and neurons/glia. Genes bound only by Sox2 were most significantly expressed in ES cells (Fig. 5B), whereas genes sequentially bound by Sox2 in ES cells and Sox3 in NPCs were most highly expressed in NPCs (Fig. 5B). Genes targeted only by Sox3 were mostly expressed in differentiated neurons and glia (Fig. 5B). Thus, in analogy with the prebinding of Sox3 to neuronal and glial genes, Sox2 prebinds many silent genes in ES cells that are targeted and activated at a succeeding stage of neurogenesis by NPC-expressed Sox2 and Sox3 proteins (Supplemental Table 8).

The finding that more than half of all Sox2-bound genes in ES cells are targeted by Sox3 in NPCs raises the question of whether Sox2-prebound genes are predominantly expressed during neural lineage differentiation or whether Sox2 binding in ES cells is evenly distributed among genes of all cellular lineages. To address this issue, we examined the expression pattern of Sox2 targeted genes with bivalent chromatin modifications, since these marks have been associated with genes that become activated at later stages of development (Bernstein et al. 2006; Boyer et al. 2006; Lee et al. 2006). Interestingly, among populations of ES cells and progenitor populations of endodermal, mesodermal, and neural origin, we found that bivalent genes prebound by Sox2 in ES cells are strongly expressed in NPCs, but not in cells of the other lineages (Fig. 5C). Bivalent histone marks, regardless of Sox2 binding, were not associated with genes expressed in a particular lineage (Fig. 5C). Characterization of bivalent genes bound by Oct4 or Nanog showed a similar, but not as significant, bias for NPC expression (Supplemental Fig. 8). This could possibly be explained by the fact that Sox3 targeted sites
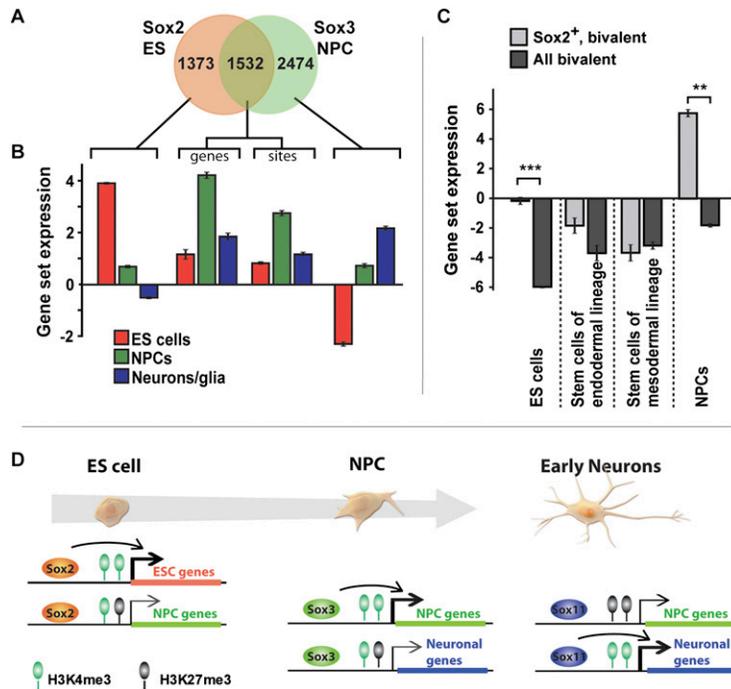
Figure 5. Bivalent NPC genes prebound by Sox2 in ES cells. (*A*) Venn diagram showing the overlap in number of target genes between Sox2 in ES cells and Sox3 in NPCs. (*B*) Expression profile for genes bound by Sox2 in ES cells and Sox3 in NPCs, as shown in *A*. Gene set expression in ES cells, NPCs, and neurons/glia is presented as percentile rank above average, with error bars showing standard error of the mean among replicates. Overlap in Sox binding at both the level of genes and further separated into those genes bound by Sox2 and Sox3 at the same site (56%–58% of genes). (*C*) Gene set expression in stem and progenitor cells of different origins for genes with Sox2 binding close (<5 kb) to bivalent domains containing both H3K4me3 and H3K27me3 marks in ES cells. All genes with bivalent marks are shown as a control. Significant differences (paired *t*-test) are indicated. (**) $P < 0.01$ or (***) $P < 0.001$. (*D*) Model depicting the sequential binding of Sox proteins to common downstream genes in stem cells differentiation along the neural lineage, highlighting the association between Sox prebinding and bivalent histone modifications.

contain accompanying Oct4-binding motifs to a lower degree than Sox2-bound sites in ES cells (Supplemental Fig. 9). Nevertheless, in parallel to its function in maintaining gene expression in pluripotent stem cells, Sox2 prebinding specifies neural lineage-specific gene programs.

Discussion

Sequentially acting Sox proteins are necessary from early pluripotent stem cell stages to the generation of differentiated neuronal progeny (Avilion et al. 2003; Bylund et al. 2003; Graham et al. 2003; Bergsland et al. 2006; Hoser et al. 2008), but their involvement in coordinating programs of neural lineage-specific gene expression has remained elusive. In this study, we compared genome-wide binding of SoxB1 and SoxC proteins during neurogenesis and demonstrated that gene sets designated to be activated in cells differentiating along the neural lineage are preselected and activated by sequentially acting Sox proteins. Thus, a single family of transcription factors uses several regulatory means to coordinate neural lineage-specific gene expression from early pluripotent stem cell stages to the onset of neuronal and glial gene expression (Fig. 5D).

Transactivation experiments and analyses in the developing chick neural tube suggest that the regulatory properties of SoxB1 proteins depend, at least in part, on their function as transcriptional activators (Kamachi et al. 2001; Bylund et al. 2003; Graham et al. 2003). This function relies on the presence of a *trans*-activation domain in their C-terminal regions and its interaction with coactivators, including p300 (Bernadt et al. 2004). Indeed, overexpression of Sox3 in NPCs up-regulated hundreds of target genes, while knockout of Sox2 in ES cells (Masui et al. 2007) leads

to a general decrease in the expression of Sox2-bound genes (data not shown). Apart from functioning as bona fide transcription factors, Sox proteins also have properties of architectural proteins that upon DNA binding can induce bending of DNA and local unwinding of chromatin (Ferrari et al. 1992). Based on these findings, it is tempting to speculate that the prebinding of Sox proteins to silent genes may protect enhancers from epigenetic repression, such as DNA methylation or heterochromatin formation, and promote the formation of permissive chromatin that facilitates gene activation as the proper cellular context of activating Sox factors and their associated partner factors has developed. These ideas are consistent with the fact that genes that are prebound by Sox proteins are in a poised state (associated with bivalent chromatin marks) and with our and others' finding (Liber et al. 2010) that SoxB1 proteins have the capacity to establish local epigenetic changes by promoting H3K4 trimethylations at bound enhancers. Moreover, Sox2-binding regions in human ES cells are depleted of DNA methylation (Lister et al. 2009). Thus, apart from Sox3's capacity to prevent premature activation of prebound neuronal genes by competing for binding sites with Sox11, it is likely that SoxB1 proteins also facilitate prebound genes to be activated at later stages of neural development. The prebinding of silent genes is not only intrinsic to neural lineage differentiation. For instance, the liver-specific enhancer of the *Alb1* gene is protected from methylation in ES cells by FoxD3 binding, which appears to be a prerequisite for its later activation by FoxA1 in liver cells (Xu et al. 2009). These findings demonstrate that prebinding transcription factors have vital regulatory functions, but to gain a deeper understanding for their role during stem cell differentiation, it will be necessary to further measure how prebinding

participates in the control of lineage-specific gene expression through the modulation of gene regulatory methylations and the accessibility of DNA.

In comparison with other transcription factors, Sox proteins bind DNA with relatively low affinities (Kamachi et al. 2000) and recognize a binding motif of only 6–8 bases (Remenyi et al. 2003). To increase the binding strength to DNA and refine their target gene selection in different cellular contexts, Sox proteins often interact with other transcriptional regulators (Kamachi et al. 2000; Bernard and Harley 2010). Indeed, we found that most peaks (>88%) harbored only one strong Sox motif (data not shown). Considering the genome-wide binding in NPCs, we were unable to distinguish the pattern of Sox2 from that of Sox3. However, a comparison with the binding in ES cells revealed that nearly half of the sites occupied by Sox2 were unique to ES cells and were not targeted by Sox2 in NPCs. Thus, the genome-wide binding of Sox2 in two distinct cell types—ES cells and NPCs—constitutes a striking example of how the target gene selection of one particular transcription factor is dependent on developmental stage-specific constraints. One possible explanation for the differences in the binding specificity between ES cells and NPCs is the limitation of partner factors. Concordantly, variations in motif occurrence could be detected in regions bound by Sox2 in ES cells and NPCs, respectively. The strongest difference was found for the binding motif of the Oct4 transcription factor, which has been demonstrated to bind in synergy with Sox2 to activate ES cell genes (Yuan et al. 1995; Nichols et al. 1998; Avilion et al. 2003; Boyer et al. 2005). This motif was very commonly associated with Sox-bound sites in ES cells (Boyer et al. 2005), but was more rarely found in Sox-bound sites identified in NPCs (Supplemental Fig. 9). Moreover, apart from controlling the binding of Sox proteins, the selective expression or affinity for heterodimerizing partner factors may also underlie the inability of preoccupying Sox proteins to activate gene expression. For instance, while many genes are targeted by Sox2 in both ES cells and NPCs, their expression is first initiated in NPCs, and although many neuronal genes are prebound by Sox2 and Sox3 in NPCs in vitro and in vivo, experiments demonstrate that they can only be activated by Sox11. Thus, the interaction with DNA-binding proteins is likely to underlie the capacity of Sox proteins to regulate specific sets of genes in distinct types of cells. It is noteworthy to mention that we failed to detect any Sox11 binding at several of the Sox3-bound genes that are activated in neurons/glia. One likely possibility is that these genes are targeted by an alternative Sox factor as NPCs commit to differentiation. One such candidate Sox factor is constituted by Sox10, which, similar to Sox11 in neurons, is necessary for the activation of genes expressed in differentiated oligodendrocytes (Stolt and Wegner 2010).

SoxB1 proteins maintain ES cells as well as NPCs in an undifferentiated state, probably by maintaining the expression of a large set of progenitor genes that can contribute to growth and self-renewal (Avilion et al. 2003; Bylund et al. 2003; Graham et al. 2003; Boyer et al. 2005). At the same time, these Sox proteins are preparing cells for differentiation by occupying and epigenetically predisposing genes to be activated at later steps of neurogenesis. Together with studies on liver cell-specific and B-cell-specific genes (Xu et al. 2009; Liber et al. 2010), our genome-wide data reinforce an emerging paradigm for coordinated lineage selection and maintenance from early pluripotency to later post-mitotic differentiation steps by sequentially acting members of different transcription factor families (Fig. 5D). Further studies will be necessary to determine whether additional transcription factor families are predisposing gene expression programs along other developmental lineages. A likely outcome of these analyses is that development of a particular lineage, such as the neural lineage, involves the sequential activation of preselected gene expression programs.

Materials and methods

ES cell culturing and generation of Nes-Sox3 cells

Mouse ES cells (E14.1) were cultured as described (Andersson et al. 2006). For in vitro differentiation, cells were grown in N2B27 (Ying and Smith 2003) supplemented with 20 ng/mL bFGF (Invitrogen), 8 nM SHH (R&D Systems), and 0.5 μM retinoic acid (all-trans; Sigma) for 0–12 d. The Nes-Sox3 stable line was generated by nucleofecting (Amaxa) mouse ES cells with NesE vector (Andersson et al. 2006) expressing a myc-tagged version of Sox3. After selection, individual clones were expanded and tested for transgene expression.

ChIP

ChIPs were performed using Millipore ChIP assay kit, ChIP-IT express (Active Motif), or Re-ChIP-it (Active Motif) according to the recommendations of the manufacturers' instructions. Chromatin was sheered by sonication (Bioruptur, Diagenode), 30 sec on/30 sec off, for 9–12 min. Antibodies against full-length mSox2 (Millipore), mSox3 (T. Edlund, Umea University), mSox11 (M. Wegner, University of Erlangen), H3K4me3 (Abcam), and H3K27me3 (Millipore) were used. Detection of ChIP signal was done by qPCR (Rotor Gene RG-3000A, Corbett) (primer sequences available on request) using SYBR Green (Biotools). ChIP signals were considered positive when Ct_{sample} (negative region) $-$ Ct_{sample} (positive region) was >2 after normalization of Ct_{IgG} for the corresponding qPCR run. Sequence libraries were generated using Illumina ChIP-seq kit and sequencing was done on an Illumina Genome Analyzer IIx (Fasteris SA) and Illumina HiSeq. Sox2 and Sox3 ChIPs were performed on cells after 4 d of differentiation conditions (DDC), and Sox11 ChIP was performed after 11 DDC. We could detect an overlap between Sox3 and Sox11 expression in ~1%–5% of the cells cultured for 4 and 11 d of differentiation.

RNA sequencing of Sox2/3 knockdown NPCs and early neurons

ES cell-derived NPCs were transfected with siRNA targeting Sox3 (5′-GCGGAAAUGGGACUUGCUA-3′, 5′-UAGCAAGUCCCA UUUCCGC-3′) (Sigma) and Sox2 (Liber et al. 2010) and universal siRNA control using Lipofectamine 2000 and harvested 24 h post-transfection. Magnetic sorting of mouse ES cell-derived early neurons was performed according to protocol (Miltenyi Biotec) using PSA-NCAM antibodies (Chemicon). RNA was extracted with Qiagen RNA extraction kit and prepared using Illumina

mRNA-seq kit, and samples were sequenced with an Illumina Genome Analyzer IIx. RNA-seq data were mapped using Bowtie (Langmead et al. 2009), and expression levels were calculated for RefSeq genes (Ramskold et al. 2009) that were probed on Affymetrix Mouse 430-2 GeneChip to enable comparisons with microarray data.

Immunohistochemistry

Antibody stainings were performed as previously described (Tsuchida et al. 1994). Antibodies that were used but were not described earlier were as follows: Plp1 (AbCam), GFAP (DACO), Sox1 (provided by S. Wilson, Umeå University), Sox2 (provided by T. Edlund, Umea University), Tuj1 (Covance), Notch1 (Santa Cruz Biotechnology), c-Myc (Santa Cruz Biotechnology), S100β (Abcam), and Lhx2 (Hybridoma Bank).

Immunoprecipitation of ^{35}S-methionine-labeled proteins

In vitro transcribed and translated mouse Sox1, Sox2, Sox3, and Sox11 were labeled with ^{35}S-methionine and allowed to react individually with 5–10 µg of Sox1, Sox2 (Millipore), Sox3, or Sox11 antibodies; precipitated with protein A agarose beads (Invitrogen); and analyzed on polyacrylamide gel.

Subcloning of enhancers and P19 cell assays

Mouse genomic regions (mm9) for Tubb3 (chr8: 125935175–125935403), Lhx2 (chr2: 38203511–38203819), and Pax2 (chr19: 44770650–44770952 and chr19: 44852024–44852490; nonfunctional) were selected based on our Sox3 and Sox11 ChIP-seq experiments (Supplemental Tables 3, 4) and conservations. Sox-bound regions were amplified by PCR and subcloned into multiple cloning sites of the βglobin-GFP-MCSIII and pTK-luc vectors. To analyze the activity of the pTK-luc vector, P19 cells (expressing endogenous levels of Sox1–3 but not Sox11) or COS1 cells (devoid of Sox1–3 or Sox11 expression) were transfected with 125 ng of DNA and 100,000 cells. Plasmids (pTK-luciferase reporter, CMV-lacZ and pCAGG-CMV constructs) (Bergsland et al. 2006; Bylund et al. 2003) were cotransfected, and luciferase activity was measured 24 h post-transfection. For 293 cell transfections, 50 ng of DNA and 100,000 cells of pCAGG-mSox1, pCAGG-mSox2, pCAGG-mSox3, or pCAGG-mSox11 were used, and cells were fixed and processed for immunohistochemistry 24 h post-transfection. Cell transfections and activity assay methods have been described elsewhere (Wang et al. 2003).

Peak calling and gene mapping for ChIP-seq data

Bowtie (Langmead et al. 2009) was used to align reads to the mouse genome (mm9). Erange (Johnson et al. 2007) was used for peak calling for the Sox3 ChIP-seq data using the following criteria: minimum of five reads, maximum of 75-base-pair (bp) gaps, options set to strict and listpeak. Peaks that were independently found in both biological replicate 1 and pooled 2/3 were used for further analyses. For the Sox2 (in NPCs) and Sox11 ChIP-seq data, peaks were called with Site Identification from Short Sequence Reads (SISSRS) (Jothi et al. 2008) to allow less-stringent criteria (5% false discovery rate [FDR]). We removed peaks with two or more reads within 300 bp in the negative control, with less than two reads in any of the replicates, or with <13 reads in total using all biological replicates. The conclusions derived from analyses of these Sox-bound regions were robust to differences in peak calling (Supplemental Fig. 5). For the comparison between promoter and enhancer peaks (Supplemental Fig. 3B–D), Erange

was used as above, but we pooled all data from Sox3 replicates. In all ChIP-seq analyses, sites were mapped to the closest RefSeq TSSs and we limited the analyses to distal enhancers within 1 kb to 1 Mb of a TSS, but the results were not sensitive to the upper cutoff. DAVID 2008 (Dennis et al. 2003) was used for gene ontology analysis. For the gene set of Sox2 targeted bivalent domains in ES cells, we used peaks for H3K27me3 in ES cells and kept those that overlapped a peak for H3K4me3 (Mikkelsen et al. 2007). In addition, their midpoint was required to be within 10 kb of a RefSeq TSS and within 5 kb of a Sox2 peak's midpoint. Sox2 peaks within 1 kb of a TSS were not included.

Gene set expression analysis

We downloaded all raw data from Gene Expression Omnibus (GEO) for the Affymetrix Mouse 430-2 platform. Each experiment was processed independently using Affymetrix Power Tools with RMA and custom probe definitions (CustomCDF version 11) (Dai et al. 2005) mapping to RefSeq transcripts. Genes in each sample were first ranked by expression. We then calculated the mean of ranks for a gene set in a sample and subtracted the mean of ranks for the same gene set across all other samples on the platform in GEO. The difference in mean rank was then divided by the number of genes on the array and multiplied by 100 to get a percentile rank above average. We note that a percentile rank difference of 4 corresponds to a 20%–30% increase in expression level. The calculated percentile rank above background was used as "Gene Set Expression" values throughout the study and was found to be robust and general enough to be used on both microarray and RNA sequencing data (Supplemental Fig. 6).

Reanalyses of previously published microarray and ChIP-seq data

Microarray data were obtained from GEO, as follows: ES cells (wild type) and NPCs were from GSE12982; neurons and glia were from GSE13379, excluding cultured samples and unbound fractions; stem and progenitor cells of mesodermal lineage were from GSE5011, GSR6933, GSE6503, GSE9198, GSE10627, GSE7012, GSE11415, and GSE12993; stem and progenitor cells of endodermal lineage were from GSE3216 and GSE8818; and C2C12 cells were from GSE7863 and GSE13347.

ChIP-seq data on Sox2, Myc, E2f1, and p300 in ES cells were taken from GSE11431, and p300 in embryonic brains was obtained from GSE13845 and converted to mm9 assembly using liftOver. RNA-seq data for mouse tissues comes from Mortazavi et al. (2008).

FACS sorting of NPCs and microarray experiment

Sox1-GFP and Sox1-GFP/Nes-Sox3 ES cells were differentiated to a Sox1-GFP-expressing state, trypsinized, and FACS-sorted. Obtained Sox1-GFP-positive cells were lysed, and RNA was extracted using RNA extraction kit (Qiagen). Gene expression analyses were performed using Affymetrix Mouse Genome 430 2.0 GeneChip arrays at the Karolinska Institute core facility (http://www.bea.ki.se). Raw data from three samples of each were analyzed as for GEO data above. We filtered for genes found present in at least two of three biological control replicates. Fold change between mean expression of biological replicates was in comparison with Sox-bound sites, and 95% confidence intervals were computed using the adjusted Wald method. The Sox3-up-regulated gene set used for gene set expression analysis required a fold change of 1.5 or more and did not use filtering.

Estimation of overlap between two ChIP-seq data sets

Two factors were considered binding to the same site if the middles of the two peaks (i.e., their summits) were within 300 bp. Overlaps between Sox2 (ES cells) and Sox3 (NPCs) presented in Figure 5A and the fraction of p300 sites with Sox binding (Fig. 1C) used all peaks for best sensitivity. Since fold enrichment in ChIP-seq experiments affect the numbers of detected peaks, we corrected for ChIP-seq strength in the comparison between Sox3 (NPC) and Sox11 (early neurons) through the development of another more sensitive method (used for Fig. 1E). First, peaks were defined by ChIP-seq reads for the first factor, then the read coverage for the second factor was scored as binding the peaks or not. We selected an expected threshold that would balance the numbers of false positive and false negative sites, with false negatives estimated from shifting the read positions 10 kb upstream or downstream. The algorithm was defined as o, b = fraction of sites with a number of reads from real ChIP-seq data or background (shifted sites), respectively.

$$cum(x, i) = \sum_{j=i} x_j \text{ (fraction of } x \text{ with at least } i \text{ reads).}$$

$$FDR_i = \frac{1 - cum(b, i)}{1 - cum(o, i)} \times \frac{cum(o, i)}{cum(b, i)} \text{ (where FDR is the estimated false discovery rate).}$$

$$t_i = o_i \times \left[1 - \frac{b_i}{o_i} \times \frac{cum(o, i)}{cum(b, i)}\right] \text{(where } t \text{ is the fraction not explained by background).}$$

$$FNR_i = 1 - \frac{cum(t, i)}{\sum t} \text{ (where FNR is the estimated false negative rate).}$$

We then found the threshold (the smallest i) where $FNR_i \geq FDR_i$. From this, we defined Sox3-specific (i.e., no Sox11-binding) sites as those with less than three Sox11 ChIP-seq reads, and Sox3-specific genes as those closest to one of these sites but not the closest to a site with three or more Sox11 ChIP-seq reads. For Sox11 sites and genes in the same analyses, however, we used those that came directly from the Sox11 ChIP-seq.

Motif discovery and enrichment

MEME (Bailey and Elkan 1994) and CisFinder (Sharov and Ko 2009) were used for de novo motif discovery with default settings and a random subset of sequences (200 or 400 bp long) centered at peak midpoints. As a background set, random sites located within 10 kb at either side of Sox3 sites were chosen and 200-bp were sequences taken. TAMO (Gordon et al. 2005) was used to score position weight matrix motif against genomic sites. Estimation of the number of motif occurrences above/below background level was calculated for sites following the background estimation model, defined for site overlap estimation above, with percentage = $100 \times sum(t)$. Ninety percent confidence intervals were inferred by bootstrapping for sample sizes >20 peaks. WebLogo (Crooks et al. 2004) was used to visualize enriched motifs. The 30-vertebrate species phastCons track downloaded from the University of California at Santa Cruz Genome Browser was used for conservation values.

In Supplemental Figure 9, to compare motifs at Sox2 (in ES cells) and Sox3 sites, we used 200-bp-long peak summit-centered sequences from the 4302 highest peaks of each and a database of motifs collected from Jaspar core vertebrates and the Bulyk PBM Database (http://the_brain.bwh.harvard.edu/pbms/webworks2; Badis et al. 2009), but removed peaks that occurred in both data sets (within 300 bp). Because Sox2- and Sox3-bound sites have different G+C content levels, we used different background sequence sets for Sox2 and Sox3, consisting of 100 dinucleotide shuffles (Altschul and Erickson 1985) of each sequence in the set. For Sox3-only against Sox3–Sox11 shared sites, we used Sox3 sites with less than three Sox11 reads and Sox11 sites with more than two Sox3 reads, without a cutoff on the number of reads to increase statistical power (however, this makes motifs higher in the shared set, which has fewer sites, less trustworthy). The comparison of the Sox motif (Supplemental Fig. 3A) used the highest 1000 peaks (by number of reads), since Sox motif degeneracy correlates with binding strength. The sequence CCTTTGTT (the most common bases in the identified Sox3 motif) was used, and we iteratively scanned the Sox-bound regions for matching CCTTTGTT sequence variants where one position was allowed to vary at a time. We repeated the scan with regions shifted 200 bp to estimate background. These frequencies were subtracted from the sequence frequencies in the Sox-bound regions for visualizing motifs (Supplemental Fig. 3A) and were added to the expected frequencies for χ^2 tests for each position, where P-values were then Benjamini-Hochberg-adjusted.

C2C12 ChIP-seq analyses

C2C12 cells were transfected with 5 μg of pCAGG-mSox3 or pCAGG-GFP. Cells were harvested 24 h after transfection and processed for ChIP-seq experiments. Sox3 peaks were called by SISSRS (5% FDR), filtering out peaks with two or more IgG reads within 300 bp of a peak summit. For analyses of changes in H3K4me3 and H3K27me3 levels at enhancers, we counted reads within 250 bp of the top 1000 Sox3 sites (located >1 kb from a TSS) in ChIP-seq before and after Sox3 introduction. We obtained a threshold of detection for respective histone methylation by taking the 99th percentile of read coverage at random genomic locations (0.82 reads per kilobase per million [RPKM] for H3K4me3 and 0.97 RPKM for H3K27me3) and subsequently used these thresholds to filter out Sox3-bound regions with an average RPKM (before and after transfection) below threshold. Changes in histone methylation at promoters were based on 1000 promoters (±250 bp) neighboring the top 1000 Sox3 sites located within 20 kb of annotated TSSs. Fold changes in methylation levels were computed as read count ratios (after/before Sox3 transfection). For clarity of presentation, we computed a normalized fold change through the division of a constant (0.89 for H3K4me3 and 0.70 for H3K27me3). The constants were computed so that methylation levels at promoters, regardless of Sox3 binding, had a median fold change of 1. Finally, to provide further evidence for direct effects, we computed the correlation between Sox3 read counts and the histone methylation fold change using all Sox3 sites.

Accession numbers

The reported sequence read data have been deposited to the Sequence Read Archive at NCBI (SRP009040 and SRP009041), and microarray data have been deposited to the GEO (GSE33024).

References

Altschul SF, Erickson BW. 1985. Significance of nucleotide sequence alignments: A method for random sequence permutation that preserves dinucleotide and codon usage. *Mol Biol Evol* **2:** 526–538.

Andersson E, Tryggvason U, Deng Q, Friling S, Alekseenko Z, Robert B, Perlmann T, Ericson J. 2006. Identification of intrinsic determinants of midbrain dopamine neurons. *Cell* **124:** 393–405.

Avilion AA, Nicolis SK, Pevny LH, Perez L, Vivian N, Lovell-Badge R. 2003. Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes Dev* **17:** 126–140.

Badis G, Berger MF, Philippakis AA, Talukder S, Gehrke AR, Jaeger SA, Chan ET, Metzler G, Vedenko A, Chen X, et al. 2009. Diversity and complexity in DNA recognition by transcription factors. *Science* **324:** 1720–1723.

Bailey TL, Elkan C. 1994. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol* **2:** 28–36.

Bergsland M, Werme M, Malewicz M, Perlmann T, Muhr J. 2006. The establishment of neuronal properties is controlled by Sox4 and Sox11. *Genes Dev* **20:** 3475–3486.

Bernadt CT, Nowling T, Rizzino A. 2004. Transcription factor Sox-2 inhibits co-activator stimulated transcription. *Mol Reprod Dev* **69:** 260–267.

Bernard P, Harley VR. 2010. Acquisition of SOX transcription factor specificity through protein–protein interaction, modulation of Wnt signalling and post-translational modification. *Int J Biochem Cell Biol* **42:** 400–410.

Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, Fry B, Meissner A, Wernig M, Plath K, et al. 2006. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125:** 315–326.

Bhattaram P, Penzo-Mendez A, Sock E, Colmenares C, Kaneko KJ, Vassilev A, Depamphilis ML, Wegner M, Lefebvre V. 2010. Organogenesis relies on SoxC transcription factors for the survival of neural and mesenchymal progenitors. *Nat Commun* **1:** 9. doi: 10.1038/ncomms1008.

Boyer LA, Lee TI, Cole MF, Johnstone SE, Levine SS, Zucker JP, Guenther MG, Kumar RM, Murray HL, Jenner RG, et al. 2005. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122:** 947–956.

Boyer LA, Plath K, Zeitlinger J, Brambrink T, Medeiros LA, Lee TI, Levine SS, Wernig M, Tajonar A, Ray MK, et al. 2006. Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* **441:** 349–353.

Bylund M, Andersson E, Novitch BG, Muhr J. 2003. Vertebrate neurogenesis is counteracted by Sox1–3 activity. *Nat Neurosci* **6:** 1162–1168.

Chen X, Xu H, Yuan P, Fang F, Huss M, Vega VB, Wong E, Orlov YL, Zhang W, Jiang J, et al. 2008. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* **133:** 1106–1117.

Crooks GE, Hon G, Chandonia JM, Brenner SE. 2004. WebLogo: A sequence logo generator. *Genome Res* **14:** 1188–1190.

Dai M, Wang P, Boyd AD, Kostov G, Athey B, Jones EG, Bunney WE, Myers RM, Speed TP, Akil H, et al. 2005. Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data. *Nucleic Acids Res* **33:** e175. doi: 10.1093/nar/gni179.

Dennis G Jr, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA. 2003. DAVID: Database for annotation, visualization, and integrated discovery. *Genome Biol* **4:** 3. doi: 10.1186/gb-2003-4-5-p3.

Engelen E, Akinci U, Bryne JC, Hou J, Gontan C, Moen M, Szumska D, Kockx C, van Ijcken W, Dekkers DH, et al. 2011. Sox2 cooperates with Chd7 to regulate genes that are mutated in human syndromes. *Nat Genet* **43:** 607–611.

Favaro R, Valotta M, Ferri AL, Latorre E, Mariani J, Giachino C, Lancini C, Tosetti V, Ottolenghi S, Taylor V, et al. 2009. Hippocampal development and neural stem cell maintenance require Sox2-dependent regulation of Shh. *Nat Neurosci* **12:** 1248–1256.

Ferrari S, Harley VR, Pontiggia A, Goodfellow PN, Lovell-Badge R, Bianchi ME. 1992. SRY, like HMG1, recognizes sharp angles in DNA. *EMBO J* **11:** 4497–4506.

Gordon DB, Nekludova L, McCallum S, Fraenkel E. 2005. TAMO: A flexible, object-oriented framework for analyzing transcriptional regulation using DNA-sequence motifs. *Bioinformatics* **21:** 3164–3165.

Graham V, Khudyakov J, Ellis P, Pevny L. 2003. SOX2 functions to maintain neural progenitor identity. *Neuron* **39:** 749–765.

Guth SI, Wegner M. 2008. Having it both ways: Sox protein function between conservation and innovation. *Cell Mol Life Sci* **65:** 3000–3018.

Hoser M, Potzner MR, Koch JM, Bosl MR, Wegner M, Sock E. 2008. Sox12 deletion in the mouse reveals nonreciprocal redundancy with the related Sox4 and Sox11 transcription factors. *Mol Cell Biol* **28:** 4675–4687.

Johnson DS, Mortazavi A, Myers RM, Wold B. 2007. Genome-wide mapping of in vivo protein–DNA interactions. *Science* **316:** 1497–1502.

Jothi R, Cuddapah S, Barski A, Cui K, Zhao K. 2008. Genome-wide identification of in vivo protein–DNA binding sites from ChIP-seq data. *Nucleic Acids Res* **36:** 5221–5231.

Kamachi Y, Uchikawa M, Kondoh H. 2000. Pairing SOX off: With partners in the regulation of embryonic development. *Trends Genet* **16:** 182–187.

Kamachi Y, Uchikawa M, Tanouchi A, Sekido R, Kondoh H. 2001. Pax6 and SOX2 form a co-DNA-binding partner complex that regulates initiation of lens development. *Genes Dev* **15:** 1272–1286.

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10:** R25. doi: 10.1186/gb-2009-10-3-r25.

Lee TI, Jenner RG, Boyer LA, Guenther MG, Levine SS, Kumar RM, Chevalier B, Johnstone SE, Cole MF, Isono K, et al. 2006. Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell* **125:** 301–313.

Liber D, Domaschenz R, Holmqvist PH, Mazzarella L, Georgiou A, Leleu M, Fisher AG, Labosky PA, Dillon N. 2010. Epigenetic priming of a pre-B cell-specific enhancer through binding of Sox2 and Foxd3 at the ESC stage. *Cell Stem Cell* **7:** 114–126.

Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Ngo QM, et al. 2009. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462:** 315–322.

Lothian C, Lendahl U. 1997. An evolutionarily conserved region in the second intron of the human nestin gene directs gene expression to CNS progenitor cells and to early neural crest cells. *Eur J Neurosci* **9:** 452–462.

Marson A, Levine SS, Cole MF, Frampton GM, Brambrink T, Johnstone S, Guenther MG, Johnston WK, Wernig M, Newman J, et al. 2008. Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. *Cell* **134:** 521–533.

Maruyama M, Ichisaka T, Nakagawa M, Yamanaka S. 2005. Differential roles for Sox15 and Sox2 in transcriptional

control in mouse embryonic stem cells. *J Biol Chem* **280:** 24371–24379.

Masui S, Nakatake Y, Toyooka Y, Shimosato D, Yagi R, Takahashi K, Okochi H, Okuda A, Matoba R, Sharov AA, et al. 2007. Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. *Nat Cell Biol* **9:** 625–635.

Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, Alvarez P, Brockman W, Kim TK, Koche RP, et al. 2007. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448:** 553–560.

Mohn F, Weber M, Rebhan M, Roloff TC, Richter J, Stadler MB, Bibel M, Schübler D. 2008. Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. *Mol Cell* **30:** 755–766.

Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-seq. *Nat Methods* **5:** 621–628.

Nichols J, Zevnik B, Anastassiadis K, Niwa H, Klewe-Nebenius D, Chambers I, Scholer H, Smith A. 1998. Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell* **95:** 379–391.

Ramskold D, Wang ET, Burge CB, Sandberg R. 2009. An abundance of ubiquitously expressed genes revealed by tissue transcriptome sequence data. *PLoS Comput Biol* **5:** e1000598. doi: 10.1371/journal.pcbi.1000598.

Remenyi A, Lins K, Nissen LJ, Reinbold R, Scholer HR, 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.

Sharov AA, Ko MS. 2009. Exhaustive search for over-represented DNA sequence motifs with CisFinder. *DNA Res* **16:** 261–273.

Stolt CC, Wegner M. 2010. SoxE function in vertebrate nervous system development. *Int J Biochem Cell Biol* **42:** 437–440.

Thein DC, Thalhammer JM, Hartwig AC, Crenshaw EB 3rd, Lefebvre V, Wegner M, Sock E. 2010. The closely related transcription factors Sox4 and Sox11 function as survival factors during spinal cord development. *J Neurochem* **115:** 131–141.

Tsuchida T, Ensini M, Morton SB, Baldassare M, Edlund T, Jessell TM, Pfaff SL. 1994. Topographic organization of embryonic motor neurons defined by expression of LIM homeobox genes. *Cell* **79:** 957–970.

van Beest M, Dooijes D, van De Wetering M, Kjaerulff S, Bonvin A, Nielsen O, Clevers H. 2000. Sequence-specific high mobility group box factors recognize 10–12-base pair minor groove motifs. *J Biol Chem* **275:** 27266–27273.

Visel A, Blow MJ, Li Z, Zhang T, Akiyama JA, Holt A, Plajzer-Frick I, Shoukry M, Wright C, Chen F, et al. 2009. ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* **457:** 854–858.

Wang Z, Benoit G, Liu J, Prasad S, Aarnisalo P, Liu X, Xu H, Walker NP, Perlmann T. 2003. Structure and function of Nurr1 identifies a class of ligand-independent nuclear receptors. *Nature* **423:** 555–560.

Xu J, Pope SD, Jazirehi AR, Attema JL, Papathanasiou P, Watts JA, Zaret KS, Weissman IL, Smale ST. 2007. Pioneer factor interactions and unmethylated CpG dinucleotides mark silent tissue-specific enhancers in embryonic stem cells. *Proc Natl Acad Sci* **104:** 12377–12382.

Xu J, Watts JA, Pope SD, Gadue P, Kamps M, Plath K, Zaret KS, Smale ST. 2009. Transcriptional competence and the active marking of tissue-specific enhancers by defined transcription factors in embryonic and induced pluripotent stem cells. *Genes Dev* **23:** 2824–2838.

Ying QL, Smith AG. 2003. Defined conditions for neural commitment and differentiation. *Methods Enzymol* **365:** 327–341.

Yuan H, Corbi N, Basilico C, Dailey L. 1995. Developmental-specific activity of the FGF-4 enhancer requires the synergistic action of Sox2 and Oct-3. *Genes Dev* **9:** 2635–2645.

Zaret KS, Watts J, Xu J, Wandzioch E, Smale ST, Sekiya T. 2008. Pioneer factors, genetic competence, and inductive signaling: Programming liver and pancreas progenitors from the endoderm. *Cold Spring Harb Symp Quant Biol* **73:** 119–126.