Transcription factor Sp2 potentiates binding of the TALE homeoproteins Pbx1:Prep1 and the histone-fold domain protein Nf-y to composite genomic sites

Received for publication, August 13, 2018, and in revised form, October 17, 2018 Published, Papers in Press, October 18, 2018, DOI 10.1074/jbc.RA118.005341

Sara Völkel[‡], Bastian Stielow[‡], Florian Finkernagel[‡], Dana Berger[‡], Thorsten Stiewe[§], Andrea Nist[§], and Guntram Suske[‡1]

From the [‡]Institute of Molecular Biology and Tumor Research (IMT) and the [§]Genomics Core Facility, Center for Tumor Biology and Immunology (ZTI), Philipps-University of Marburg, 35043 Marburg, Germany

Edited by Joel Gottesfeld

Different transcription factors operate together at promoters and enhancers to regulate gene expression. Transcription factors either bind directly to their target DNA or are tethered to it by other proteins. The transcription factor Sp2 serves as a paradigm for indirect genomic binding. It does not require its DNA-binding domain for genomic DNA binding and occupies target promoters independently of whether they contain a cognate DNA-binding motif. Hence, Sp2 is strikingly different from its closely related paralogs Sp1 and Sp3, but how Sp2 recognizes its targets is unknown. Here, we sought to gain more detailed insights into the genomic targeting mechanism of Sp2. ChIP-exo sequencing in mouse embryonic fibroblasts revealed genomic binding of Sp2 to a composite motif where a recognition sequence for TALE homeoproteins and a recognition sequence for the trimeric histone-fold domain protein nuclear transcription factor Y (Nf-y) are separated by 11 bp. We identified a complex consisting of the TALE homeobox protein Prep1, its partner PBX homeobox 1 (Pbx1), and Nf-y as the major partners in Sp2–promoter interactions. We found that the Pbx1:Prep1 complex together with Nf-y recruits Sp2 to co-occupied regulatory elements. In turn, Sp2 potentiates binding of Pbx1:Prep1 and Nf-y. We also found that the Sp-box, a short sequence motif close to the Sp2 N terminus, is crucial for Sp2's cofactor function. Our findings reveal a mechanism by which the DNA binding–independent activity of Sp2 potentiates genomic loading of Pbx1:Prep1 and Nf-y to composite motifs present in many promoters of highly expressed genes.

Transcription factors typically contain a DNA-binding domain, which mediates binding to specific DNA sequences in regulatory regions of genes *in vivo*. However, only a small fraction of DNA-binding motifs in the genome are occupied by the corresponding transcription factor raising the question of what determines binding. Chromatin accessibility and cooperative protein–protein–DNA interactions are generally believed to mediate context-specific binding (1, 2).

With the expansion of genome-wide transcription factor binding data, there is an increase in evidence that transcription factors occupy promoters and enhancers that lack the corresponding DNA-binding motifs. Indeed, dozens of transcription factors bind to so-called HOT (high-occupancy target) regions that do not harbor a cognate binding sequence (3–5). Therefore a transcription factor with a functional DNA-binding domain could be recruited to promoters as part of a complex with other proteins independent of whether the occupied promoter region contains a corresponding DNA-binding sequence. The transcription factor specificity protein 2 (Sp2)[2] provides a paradigm for such a recruitment mechanism. Sp2 mutants lacking the DNA-binding domain bind to target promoters as efficiently as the full-length protein even if the bound region contains a recognition motif (6).

Sp2 is closely related to transcription factors Sp1, Sp3, and Sp4 (7–10), which all have three consecutive C2H2-type zinc fingers near their C terminus, glutamine-rich domains, and two highly conserved short sequence motifs, the so-called Sp-box close to the N terminus and the button head box (Btd-box) preceding the first zinc finger (7) (Fig. 1*A*). All four proteins bind to GC-rich motifs (GC boxes) *in vitro* (11–14). Sp1, Sp2, and Sp3 are ubiquitously expressed, whereas Sp4 is largely restricted to neuronal cells (7, 8). Gene disruption studies of Sp1, Sp2, and Sp3 revealed that all three proteins are essential for normal mouse development (15–17).

Recently, we compared the genomic binding profiles of Sp1, Sp2, and Sp3 in mouse embryonic fibroblasts and in a human cell line (6, 11). We found that Sp1 and Sp3 bind to a common set of sites distinct from Sp2-bound regions (6, 18). Significantly, the most enriched motif in the Sp2-binding regions is not the GC box but the CCAAT motif, which is the recognition sequence for the trimeric histone-fold domain protein Nf-y. Our knockdown studies revealed that Nf-y is

This work was supported by Deutsche Forschungsgemeinschaft Grant SU102/8-2 (to G. S.). The authors declare that they have no conflicts of interest with the contents of this article.

This article contains Table S1 and Figs. S1–S6.

ChIP-seq data were deposited in ArrayExpress under accession number E-MTAB-7125.

[1] To whom correspondence should be addressed. Tel.: +49-6421-2866697; E-mail: Suske@imt.uni-marburg.de.

[2] The abbreviations used are: Sp2, specificity protein 2; MEF, mouse embryonic fibroblast; TALE, 3-amino acid loop extension; Pbx1, pre-B-cell leukemia homeobox 1; Prep1, Pbx-regulating protein 1; Nf-y, nuclear transcription factor y; GO, Gene Ontology; FRT, Flp recombination target; DAPA, DNA affinity precipitation assay; qPCR, quantitative PCR; GST, glutathione *S*-transferase; PMSF, phenylmethylsulfonyl fluoride; Btd-box, button head box.

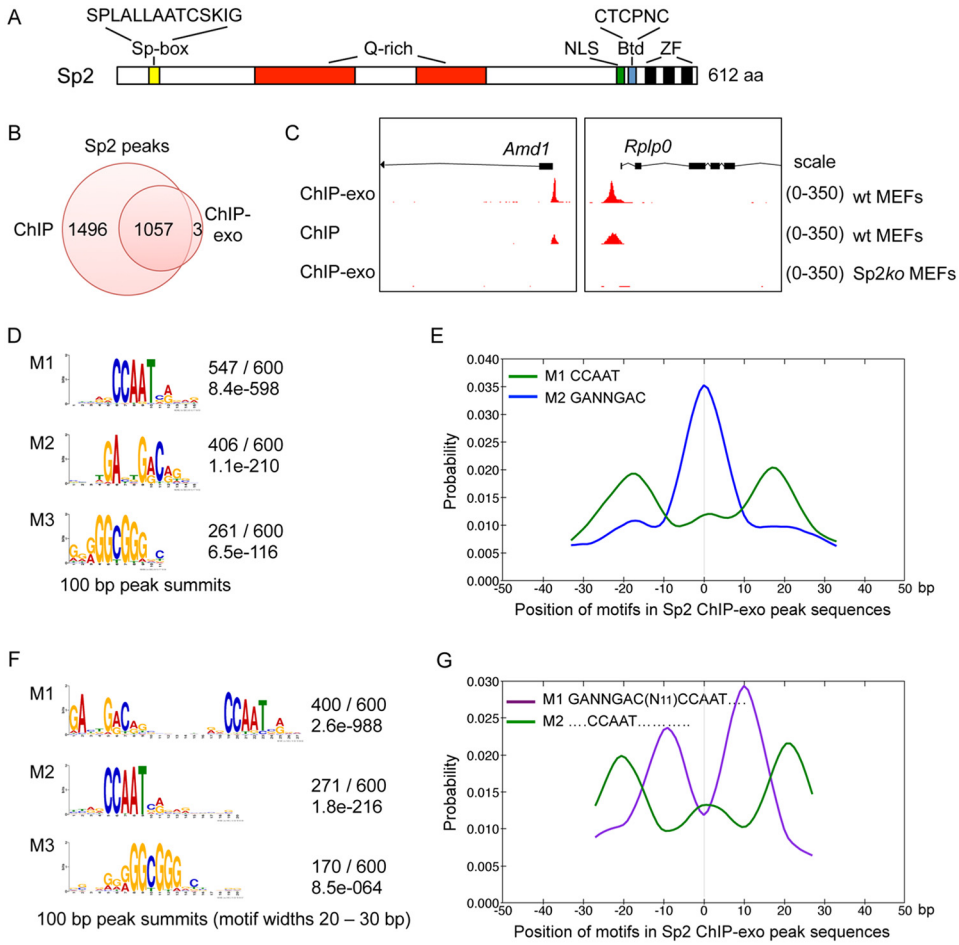

Figure 1. Sp2 localizes to composite TALE factor–Nf-y recognition motifs. Genomic binding sites of Sp2 in MEFs were determined by ChIP-exo sequencing. *A*, schematic representation of Sp2. The Sp-box (*yellow*), the glutamine-rich domains (*Q-rich*, *red*), the nuclear localization signal (*NLS*, *green*), the Btd-box (*blue*), and the zinc fingers (*ZF*, *black bars*) are indicated. *B*, Venn diagram showing the overlap of Sp2 ChIP-exo peaks with previously published Sp2 ChIP-seq peaks (11). *C*, representative genome browser screenshots showing Sp2 ChIP-exo peaks and corresponding Sp2 ChIP-seq peaks at the *Amd1* and *Rplp0* promoters. *D*, sequence motifs enriched in Sp2-binding regions. Logos were obtained by running MEME-ChIP with 100-bp sequences of the top 600 Sp2 ChIP-exo peaks using default parameters. The *numbers* next to the logos indicate the occurrence of the motifs and the statistical significance (*E*-value). *E*, local motif enrichment analysis (using CentriMo 4.12.0) of the M1 and M2 motifs shown in *D*. Of note, the GC box motif (*M3*) was not locally enriched. *F*, sequence motifs obtained by adjusting the MEME search to long motifs (20 to 30 bp widths). *G*, local motif enrichment analysis (using CentriMo 4.12.0) of the M1 and M2 motifs shown in *F*. Of note, the GC box motif (*M3*) was not locally enriched.

critical for recruitment of Sp2. We also provided a mechanistic insight into binding site selection of Sp2 *in vivo*; the glutamine-rich, positively charged N-terminal region of Sp2 is sufficient for recruitment of Sp2 to its target promoters, whereas the zinc finger DNA-binding domain is entirely dispensable (6).

Here, we report DNA binding–independent genomic loading of Sp2 to a composite motif (DECA^{ext} motif), which is bound by the dimeric TALE transcription factor Pbx1:Prep1 and Nf-y. Binding of Pbx1:Prep1 is strongly reduced in Sp2*ko* cells indicating that Sp2 potentiates Pbx1:Prep1 binding at shared sites. Expression of Sp2 mutants in Sp2*ko* cells revealed that the zinc finger DNA-binding domain is dispensable, and the Sp-box is required for potentiating genomic binding of Pbx1:Prep1. Mutational analysis of selected Sp2 target promoters identified sequence constraints required for the formation of the Sp2–Pbx1:Prep1–Nf-y complex on DECA^{ext} motifs. Both, the Pbx1:Prep1 and the Nf-y binding motifs as well as their spacing of 11 nucleotides is required for loading of Sp2. Finally, biochemical analysis revealed that Sp2 interacts directly with DNA-bound Pbx1:Prep1–Nf-y through its most N-terminal domain. Together our results provide further mechanistic insight into the role of Sp2 as an important cofactor and clearly illustrate how the interplay of different transcription factors determines their genomic binding.

Results

Transcription factor Sp2 localizes to composite TALE factor–Nf-y recognition motifs

To increase the resolution of Sp2 ChIP-seq peaks we mapped the genomic Sp2-binding sites in MEFs using the ChIP-exonuclease (ChIP-exo) technology (19). Stringent filtering of uniquely mapped reads (≥30 tags and ≥3-fold enrichment over the Sp2 knockout control) yielded >1000 high-confidence binding sites. The vast majority of the Sp2 ChIP-exo peaks (99%) overlapped with our previously published Sp2-binding sites (11) (Fig. 1*B*). Compared with the classical ChIP protocol, ChIP-exo resulted in sharper peak

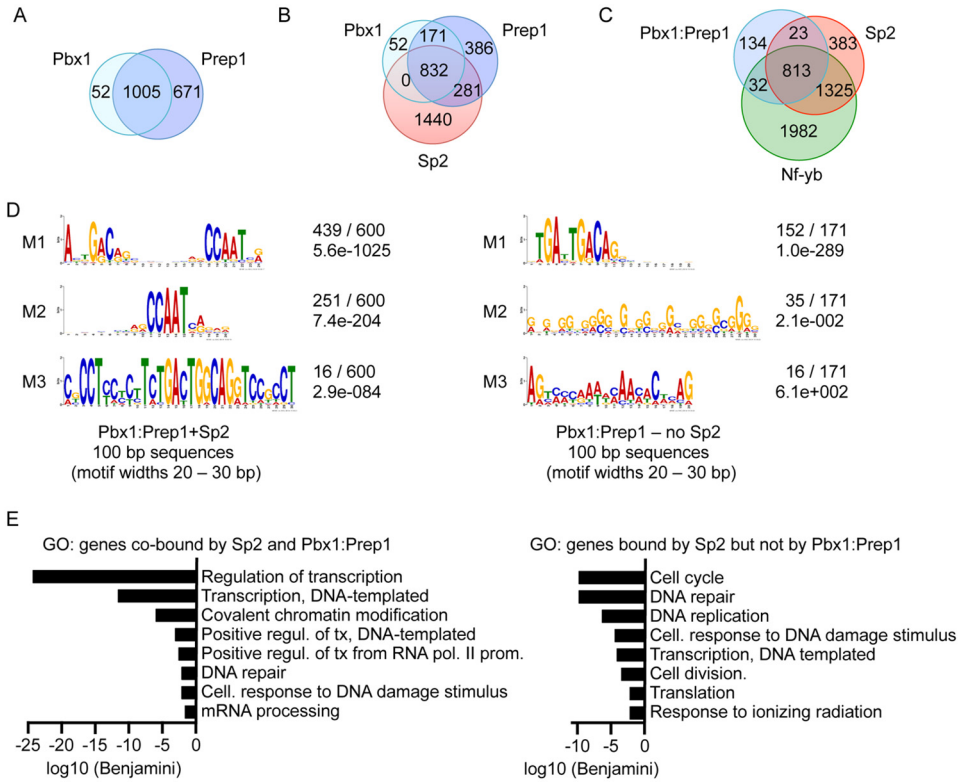

Figure 2. Overlap of Pbx1-, Prep1-, Sp2-, and Nf-y-binding sites. Genomic binding sites of Pbx1 and Prep1 in MEFs were determined by ChIP-seq. *A*, Venn diagram representing the overlap of high-confidence Pbx1 and Prep1 ChIP-seq peaks (≥30 tags, ≥3-fold enrichment over IgG control) in WT MEFs. *B*, Venn diagram showing the overlap of Pbx1- and Prep1-binding sites with Sp2-binding sites in WT MEFs. *C*, Venn diagram showing the overlap of Pbx1:Prep1-binding sites with Sp2- and Nf-yb-binding sites in WT MEFs. *D*, *left panel*, sequence motifs enriched in regions co-bound by Pbx1, Prep1, and Sp2. *Right panel*, sequence motifs in regions bound by Pbx1 and Prep1 but not by Sp2. The *numbers* next to the logos indicate the occurrence of the motifs and the statistical significance (*E* value). *E*, *left panel*, GO analyses of biological functions of genes co-bound by Pbx1, Prep1, and Sp2. *Right panel*, GO analyses of biological functions of genes bound by Sp2 but not by Pbx1 and Prep1. Enriched GO terms were retrieved using DAVID 6.8 (GOTERM_BP_DIRECT, Functional Annotation Chart). Benjamini values are plotted in $\log_{10}$ scale. Of note, genomic sites that were bound by Prep1 but not by Sp2 (see Fig. 2*B*) were mostly located remote from transcriptional start sites and could not conclusively be allocated to specific genes.

summits (Fig. 1*C*), which allowed us to map more precisely the Sp2-binding sites. A *de novo* sequence motif analysis (using the MEME Suite (20)) with 100-bp sequences extracted from the top 600 Sp2 ChIP-exo peak regions revealed three major motifs: the CCAAT motif (Nf-y-binding site) (21), a recognition sequence (GANNGAC) for the heterodimeric TALE (three amino acid loop extension) factors Pbx1:Prep1 (pre-B-cell leukemia homeobox 1:Pbx-regulating protein 1) (22), and the GC box (Sp1/Sp3-binding site) (8) (Fig. 1*D*). The GC box was not enriched at a particular position (not shown); and the CCAAT motif exhibited a symmetrical bimodal distribution (Fig. 1*E*), which is in line with our previous analysis of the Sp2-binding regions (6). Significantly, the Pbx1:Prep1 recognition sequence was enriched exactly at the center of the Sp2 peaks (Fig. 1*E*). By adjusting the motif search to motif widths between 20 and 30 bp we found that the Pbx1:Prep1 and Nf-y recognition sequences were tightly linked to each other (Fig. 1*F*). In two-thirds (400 out 600) of the top Sp2-binding sites, the Pbx1:Prep1 and Nf-y recognition sequences were found to be separated by exactly 11 nucleotides (GANNGAC(N)$_{11}$CCAAT; Fig. 1, *F* and *G*). Hence, a large fraction of the top Sp2-binding sites are characterized by a previously unrecognized composite motif containing recognition sequences for TALE factors and Nf-y.

The TALE homeoproteins Pbx1 and Prep1 colocalize with Sp2 and Nf-y

That a large fraction of the Sp2-binding sites is not only characterized by the presence of Nf-y-binding sites (CCAAT motifs) but also by a recognition motif of Pbx1:Prep1, prompted us to test whether Pbx1 and Prep1 were bound at these Sp2 target sites, and, if so, whether genomic binding of Pbx1, Prep1, Sp2, and Nf-y impinge on each other. We mapped genomic Pbx1 and Prep1-binding sites in WT MEFs by ChIP-seq. The vast majority of the Pbx1 peaks overlapped with Prep1 peaks (Fig. 2*A*), which is consistent with the association of Pbx1 and Prep1 in a stable dimeric complex (23, 24). There is also a markedly high fraction of 671 Prep1-specific peaks suggesting the existence of Pbx1-independent Prep1-binding sites. Alternatively, the absence of Pbx1 peaks at these sites could be due to a less efficient Pbx1 ChIP. Comparison of the Pbx1:Prep1-binding sites with high-confidence Sp2- and Nf-y-binding sites (6, 11) revealed that four-fifths of the Pbx1:Prep1 sites overlapped with Sp2- and Nf-y-binding sites (Fig. 2, *B* and *C*).

A *de novo* sequence analysis of the shared Pbx1:Prep1–Sp2 binding regions essentially yielded the same sequence motifs as the Sp2 peaks with a high occurrence of the composite Pbx1:Prep1–Nf-y motif (Fig. 2*D*, *left panel*). Remarkably, a very similar motif, named DECA^{ext}, was reported in a previous ChIP-

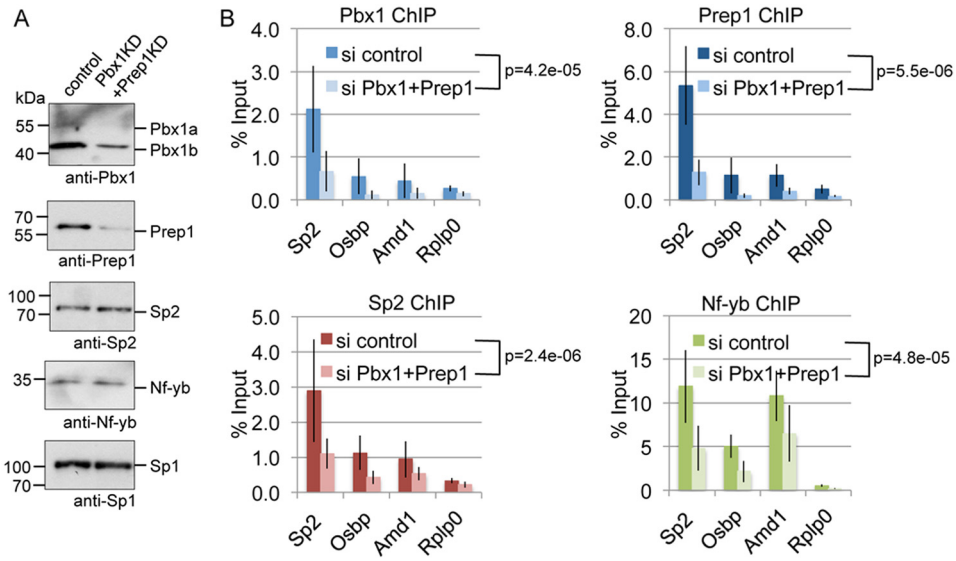

Figure 3. Pbx1:Prep1 is necessary for binding of Sp2 to shared sites. *A*, Western blot analysis of Pbx1, Prep1, Sp2, and Nf-yb expression in MEFs transfected with a nonspecific siRNA (control) or siRNAs targeting *Pbx1* and *Prep1* (*Pbx1KD*+*Prep1KD*). Re-probing for Sp1 controlled loading of extracts. *B*, ChIP-qPCR data showing reduced binding of Pbx1, Prep1, Sp2, and Nf-yb to the *Sp2*, *Osbp*, *Amd1*, and *Rplp0* promoters after simultaneous knockdown of Pbx1 and Prep1. The percent of input values are mean ± S.D. derived from three independent ChIP experiments measured in duplicate. For each factor (subpanel), we applied a 1-sample *t* test *versus* 0 on the $\log_2$ ratio (control/Pbx1KD+Prep1KD, paired by promoter and experiment).

seq study of Pbx1- and Prep1-binding sites in E11.5 mouse embryos (25). Therefore, we will refer to the composite Pbx1:Prep1–Nf-y motif from here on as the DECA^{ext} motif. Of note, sites that were bound by Pbx1:Prep1 but not by Sp2, contained also the GANNGAC motif but lacked the CCAAT motif (Fig. 2*D*, *right panel*).

We determined whether there are specific functional features shared among promoters bound by Sp2 and Pbx1:Prep1. Genes co-bound by Sp2 and Pbx1:Prep1 were highly enriched in Gene Ontology (GO) terms related to transcriptional regulation. GO terms associated with genes bound by Sp2 but not by Pbx1:Prep1 were particularly associated with cell cycle, DNA repair, replication, and translation (Fig. 2*E*). This finding suggests that promoters, which are co-bound by Pbx1:Prep1 and Sp2, and promoters that are bound by Sp2 alone regulate distinct gene sets involved in different cellular processes.

Pbx1:Prep1 is necessary for Sp2 binding

Colocalization of Sp2 and Pbx1:Prep1 at DECA^{ext} motifs led us to ask whether Pbx1:Prep1 is necessary for recruitment of Sp2 to these genomic sites. We simultaneously knocked down both isoforms of Pbx1 (Pbx1a and Pbx1b) and Prep1 by RNAi (Fig. 3*A*), and analyzed the binding of Pbx1, Prep1, Sp2, and Nf-yb to a panel of target promoters (*Sp2, Osbp, Amd1*, and *Rplp0*). Knockdown of Pbx1:Prep1 resulted in reduced binding of Sp2 and Nf-yb (Fig. 3*B*). Importantly, the protein levels of Sp2 and Nf-y were not affected upon knockdown of Pbx1 and Prep1 (Fig. 3*A*). This finding suggests that Pbx1:Prep1 is necessary for recruitment of Sp2. Reduced binding of Nf-yb upon Pbx1:Prep1 knockdown suggests that Pbx1:Prep1 may promote also binding of Nf-y to the adjacent CCAAT box within the DECA^{ext} motif.

Sp2 potentiates binding of Pbx1:Prep1

In parallel to the ChIP-seq analysis in WT MEFs, we performed ChIP-seq analysis of Pbx1 and Prep1 in *Sp2ko* MEFs and in *Sp2ko* MEFs expressing FLAG-tagged Sp2 (Sp2*ko*–FL–Sp2 cells). The expression levels of Pbx1 and Prep1 were similar in WT, Sp2*ko*, and FLAG–Sp2-expressing Sp2*ko* cells (Fig. 4*A*). Moreover, co-immunoprecipitation experiments revealed that the interaction of Pbx1 with Prep1 was similar in WT and Sp2*ko* cells (Fig. S1). However, heat map and genome browser track views of the binding densities revealed that occupancy of Pbx1 and Prep1 at shared Pbx1:Prep1–Sp2–Nf-y sites was greatly reduced in Sp2*ko* cells, and was restored in Sp2*ko* MEFs expressing FLAG-tagged Sp2 (Fig. 4, *B–D*). Of note, consistent with what we observed earlier, binding of Nf-y was also reduced in Sp2*ko* MEFs at a subset of Sp2–Nf-y target regions (Fig. 4, *B* and *D*) (6). Reduced binding of Pbx1 and Prep1 in Sp2*ko* MEFs suggests that Sp2 potentiates binding of Pbx1:Prep1 at shared sites. To further substantiate this conclusion, we probed a panel of Pbx1:Prep1–Sp2 target promoters by conventional ChIP-qPCR using independent chromatin preparations. These experiments validated strongly reduced binding of Pbx1 and Prep1 in Sp2*ko* cells and almost WT levels of Pbx1 and Prep1 binding in Sp2*ko* cells expressing FLAG–Sp2 (Fig. 4*E*). Finally, we expressed tagged versions of Pbx1 and Prep1 (FLAG–Pbx1 and FLAG–Prep1) in WT MEFs and in Sp2*ko* MEFs, and subsequently analyzed binding to selected promoters (*Sp2*, *Osbp*, *Amd1*, and *Rplp0*). We observed significantly lower binding levels of FLAG–Pbx1 and FLAG–Prep1 in Sp2*ko* cells than in WT cells (Fig. S2). Together, these results show that Sp2 potentiates genomic binding of Pbx1:Prep1 at shared sites.

The zinc finger domain of Sp2 is dispensable for potentiating genomic binding of Pbx1:Prep1

The most striking feature of Sp2 is its capacity to bind its target promoters in the absence of the zinc finger DNA-binding domain as efficiently as full-length Sp2 (6). To test whether the zinc fingers are necessary to potentiate Pbx1:Prep1 binding, we examined binding of Pbx1 and Prep1 in

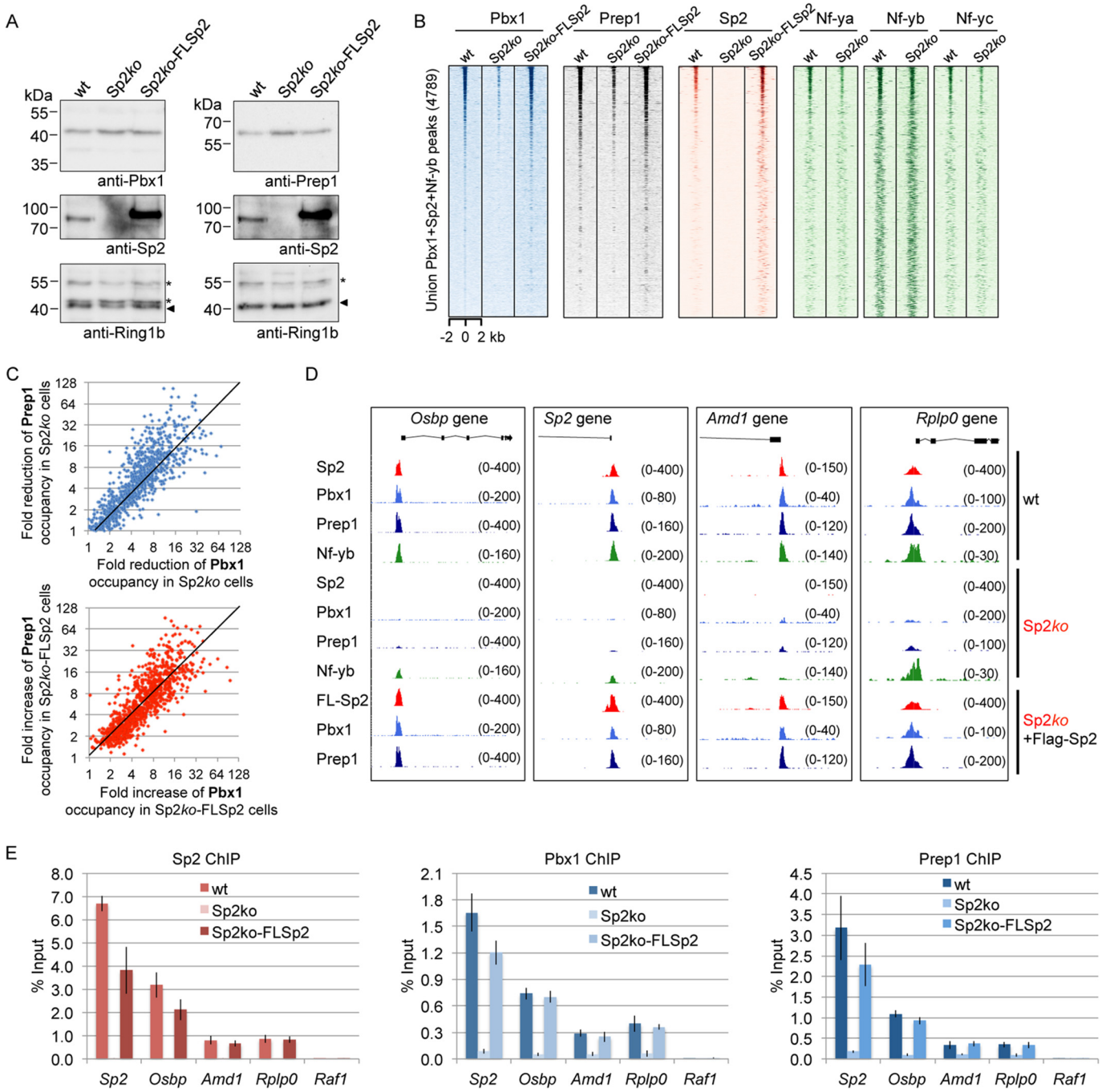

Figure 4. Sp2 potentiates binding of Pbx1 and Prep1 at shared binding sites. *A*, Western blot analysis of Pbx1, Prep1, and Sp2 expression in WT MEFs (*wt*), Sp2*ko* MEFs, and in Sp2*ko* MEFs expressing FLAG-tagged Sp2 (Sp2*ko*–FLSp2). Re-probing for Ring1b-controlled loading of extracts. The *asterisks* at the Ring1b blots denote unstripped Pbx1 antibodies and a nonspecific band. *B*, ChIP-seq heat maps of union Pbx1, Prep1, Sp2, and Nf-y (*Nf-ya, Nf-yb,* and *Nf-yc*) peaks in MEFs at ±2 kb regions centered over the Pbx1 peaks. *wt*, WT MEFs; Sp2*ko*, Sp2-deficient MEFs; *Sp2ko–FLSp2*, Sp2*ko* MEFs expressing FLAG-tagged Sp2. *C*, *upper panel*, scatter plot comparing the extent of reduction (fold-change of normalized tag counts) of Pbx1 binding with the extent of reduction of Prep1 binding in Sp2*ko* cells at individual peaks. *Lower panel*, scatter plot comparing the extent of increase of Pbx1 binding with the extent of increase of Prep1 binding after expression of FLAG-tagged Sp2 in Sp2*ko* MEFs. *D*, representative genome browser screenshots showing binding of Sp2, Pbx1, and Prep1 to the *Osbp*, *Sp2*, *Amd1*, and *Rplp0* promoters in WT MEFs, Sp2*ko* MEFs, and Sp2*ko*-FLSp2 MEFs. Binding of Nf-yb is shown in WT MEFs and Sp2*ko* MEFs. *E*, ChIP-qPCR data of Sp2, Pbx1, and Prep1 binding to selected promoters (*Sp2*, *Osbp*, *Amd1*, and *Rplp0*) in wt MEFs, Sp2*ko* MEFs, and Sp2*ko*–FLSp2 MEFs. The *Raf1* promoter, which is bound by Sp1 and Sp3 but not by Sp2 (6), served as a negative control region. Percent of input values represent the mean of at least three independent experiments ± S.D.

Sp2*ko* cells expressing zinc finger-deficient Sp2 mutants (Fig. 5, *A* and *B*). Binding of Pbx1 and Prep1 was fully restored in Sp2*ko* cells expressing the Sp2(1–525) mutant lacking the zinc finger domain, and by the Sp2(1–506) mutant lacking in addition the Btd-box (Fig. 5*C*). Thus, both the Sp2 zinc finger domain as well as the Btd-box are dispensable for potentiation of Pbx1:Prep1 binding.

We went on to identify domains of Sp2 that are necessary to potentiate Pbx1:Prep1 binding, and stably expressed a panel of additional C-terminal Sp2 deletion mutants in Sp2*ko* cells (Fig.

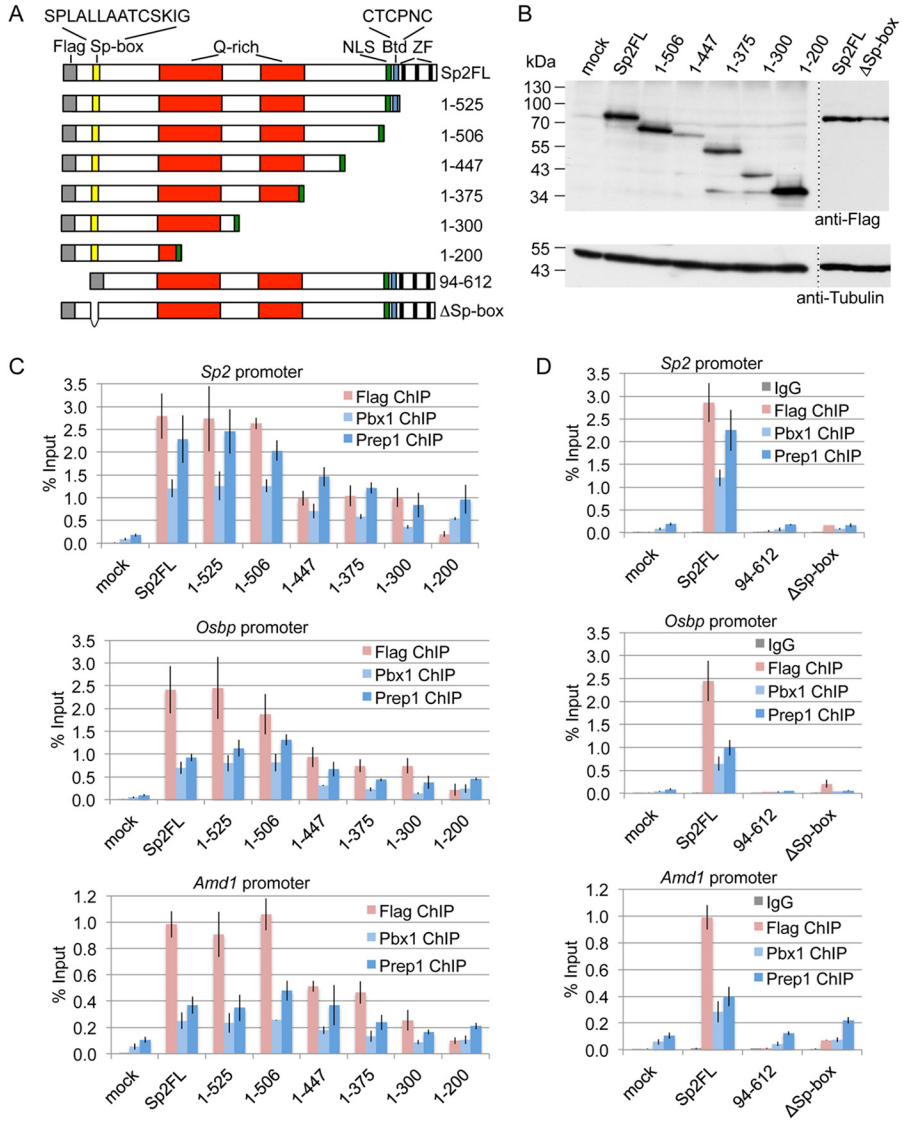

Figure 5. Identification of Sp2 domains involved in promoting Pbx1:Prep1 binding. *A*, schematic representation of Sp2 deletion mutants expressed in *Sp2ko* MEFs. The FLAG-epitope at the N termini (*gray*), the Sp-box (*yellow*), the glutamine-rich domains (*Q-rich*, *red*), the nuclear localization signal (*NLS*, *green*), the Btd-box (*blue*), and the zinc fingers (*ZF*, *black bars*) are indicated. *B*, Western blot analysis (*anti-Flag*) of Sp2 mutants expressed in Sp2*ko* MEFs. Re-probing for tubulin controlled loading of Sp2-containing extracts. *C* and *D*, ChIP-qPCR data of Pbx1 and Prep1 binding to the *Sp2*, *Osbp*, and *Amd1* promoters in Sp2*ko* cells expressing FLAG-tagged Sp2 C-terminal mutants (*C*) or FLAG-tagged Sp2 N-terminal mutants (*D*). The percent of input values are mean ± S.D. ($n = 3$).

5, *A* and *B*). Attachment of a nuclear localization signal ensured nuclear localization of all Sp2 mutants (Fig. S3). The Sp2 deletion mutants Sp2(1–447), Sp2(1–375), and Sp2(1–300) bound to the selected target promoters ~2–3-fold less efficiently. Concomitantly, WT levels of Pbx1:Prep1 binding were only partly restored (Fig. 5*C*). The Sp2(1–200) mutant still bound to the Sp2 target promoters albeit at an ~10-fold reduced level than full-length Sp2. Despite low level binding of the Sp2(1–200) fragment, binding of Pbx1 and Prep1 to the *Sp2*, *Osbp*, and *Amd1* promoters was still markedly increased compared with Sp2*ko* cells (compare "mock" with "1–200" in Fig. 5*C*).

The Sp-box of Sp2 is required for promoting genomic binding of Pbx1:Prep1

The N-terminal 200-amino acid Sp2 fragment contains the Sp-box, a hallmark of the Sp transcription factor family members, at position 33–46 (SPLALLAATCSKIG) (7). Moreover, the most 94 N-terminal amino acids of Sp2 are required for binding of Sp2 *in vivo* (Fig. 5*D*) (6). To test whether the Sp-box contributes to chromatin binding, we deleted the Sp-box in the context of full-length Sp2 (Sp2ΔSp-box) (Fig. 5, *A* and *B*). The Sp2ΔSp-box mutant bound to the selected target promoters ~10-fold less efficiently. Importantly, the binding level of Pbx1:Prep1 at the *Sp2* and *Osbp* promoters in Sp2ΔSp-box expressing cells was similar as in Sp2*ko* cells (Fig. 5*D*) suggesting an important function of the Sp-box in potentiating binding of the Pbx1:Prep1 complex. Together, we conclude that a large part of the capacity of Sp2 to potentiate genomic binding of Pbx1:Prep1 resides in the most N-terminal 200 amino acids.

Pbx1:Prep1 and the Nf-y recognition sequences are both necessary for recruitment of Sp2

We sought to determine in detail the DNA sequence constraints, which are necessary for genomic loading of Sp2. We

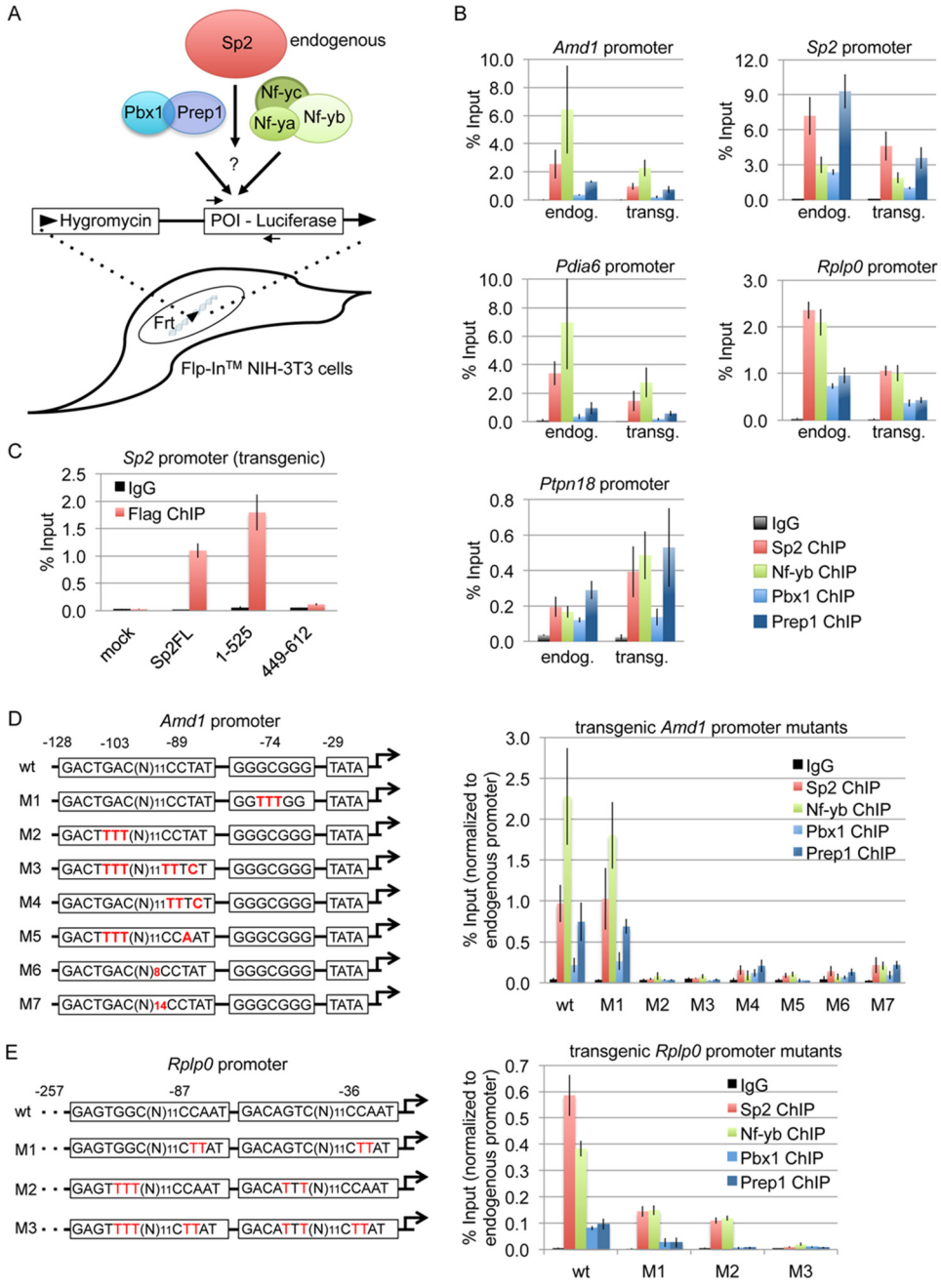

Figure 6. Pbx1:Prep1- and Nf-y-binding sites in the DECA^{ext} motif are crucial for genomic binding of Sp2 to the *Amd1* and *Rplp0* promoters. *A*, schematic representation of the experimental design. Selected Sp2 target promoters (*POI*, promoter of interest) were placed as luciferase fusions into the single FRT site of Flp-In^{TM} NIH-3T3 cells, and subsequently analyzed for binding of endogenous Sp2, Nf-yb, Pbx1, and Prep1. *B*, ChIP-qPCR data showing binding of Sp2, Nf-yb, Pbx1, and Prep1 to the endogenous and transgenic *Amd1*, *Sp2*, *Pdia6*, *Rplp0,* and *Ptpn18* promoters in Flp-In^{TM} NIH-3T3 cells. The percent of input values are mean ± S.D. (n = 3). *C*, binding of Sp2 to the transgenic *Sp2* promoter is zinc finger-independent. FLAG-tagged full-length Sp2 (*Sp2FL*), the Sp2(1–525) mutant lacking the zinc finger domain, or the Sp2(449–612) mutant encompassing the zinc finger domain were ectopically expressed in Flp-In^{TM} NIH-3T3 cells containing an integrated transgenic *Sp2* promoter. Binding of the FLAG–Sp2 mutants to the transgenic promoter was analyzed by ChIP-qPCR. Anti-FLAG antibodies were used for ChIP. The percent of input values are mean ± S.D. (n = 3). *D*, *left panel*, schematic representation of the *Amd1* promoter fragment (−128 to +134) and corresponding mutants placed into the FRT site of Flp-In^{TM} NIH-3T3 cells. Shown are the sequences of the composite Pbx1:Prep1–Nf-y binding motif, the GC box, and the TATA box. Mutations are highlighted in *red*. The full promoter sequences are shown in Fig. S4. *Right panel*, ChIP-qPCR data of Sp2, Nf-yb, Pbx1, and Prep1 binding to the transgenic WT *Amd1* promoter and the corresponding *Amd1* promoter mutants. The percent of input values are mean ± S.D. (n = 3). *E*, *left panel*, schematic representation of the *Rplp0* promoter (−257 to +37) and corresponding mutants. Shown are the sequences of the two composite Pbx1:Prep1–Nf-y binding motifs. Mutations are highlighted in *red*. The full promoter sequences are shown in Fig. S4. *Right panel*, ChIP-qPCR data of Sp2, Nf-yb, Pbx1, and Prep1 binding to the transgenic *Rplp0* promoter mutants. The percent of input values are mean ± S.D. (n = 3).

employed Flp-In^{TM} NIH-3T3 cells, which contain a single Flp recombination target (FRT) site in their genome. We introduced several Sp2 target promoters (*Amd1*, *Sp2*, *Pdia6*, *Rplp0*, and *Ptpn18*) into the FRT site of Flp-In^{TM} NIH-3T3 cells by homologous recombination (Fig. 6*A* and Fig. S4) and subsequently tested whether endogenous Sp2, Pbx1, Prep1, and Nf-y were bound to these transgenic promoters. These analyses revealed that Sp2, Pbx1, Prep1, and Nf-y were also bound to the

transgenic promoters (Fig. 6*B*) demonstrating that Sp2 is recruited to its target promoters when present at different genomic loci. Thus, only the promoter sequence and not the exact genomic position determines genomic loading of Sp2.

Next, we tested whether binding of Sp2 to the transgenic promoters is also mediated by its N-terminal region. We expressed FLAG-tagged full-length Sp2 (Sp2FL), the Sp2 mutant lacking the zinc fingers (Sp2(1–525) in Fig. 6*C*), or the Sp2 zinc finger domain on its own (Sp2(449–612)) in Flp-In™ NIH-3T3 cells carrying the transgenic *Sp2* promoter. The Sp2(1–525) mutant but not the Sp2(449–612) mutant bound to the transgenic *Sp2* promoter (Fig. 6*C*) suggesting that binding of Sp2 to the transgenic promoters is also independent of the zinc finger domain.

To determine the sequence elements, which are required for genomic loading of Sp2, we analyzed the murine *Amd1* promoter by mutagenesis. The *Amd1* promoter contains a single DECA^{ext} element and a GC box, the prototypical Sp1-binding motif (Fig. 6*D* and Fig. S4). Sp2, Nf-yb, Pbx1, and Prep1 still bound to an *Amd1* promoter mutant where the GC box is destroyed (GGGCGGG to GGTTTGG in the *Amd1* M1 mutant) unambiguously demonstrating that genomic loading of Sp2 to the *Amd1* promoter does not involve binding to the "classical" Sp-factor recognition sequence. In contrast, Sp2, Nf-yb, Pbx1, and Prep1 did not bind to *Amd1* promoters where the Pbx1:Prep1 recognition site is mutated (GACTGAC to GACTTTT in the M2 and M3 mutants). Thus, the Pbx1:Prep1 recognition motif appears to be required not only for binding of the Pbx1:Prep1 dimer but also for binding of Sp2 and Nf-y. Finally, binding of all four factors was also strongly reduced when the Pbx1:Prep1 recognition sequence was intact but the noncanonical Nf-y-binding site mutated (CCTAT to TTTCT in M4). These results suggest that both binding of Pbx1:Prep1 as well as binding of Nf-y to the composite motif DECA^{ext} motif is required for genomic loading of Sp2 on the *Amd1* promoter.

We were surprised that binding of Nf-y was lost upon mutating the Pbx1:Prep1 recognition site. Loss of Nf-y binding could be due to the absence of a canonical CCAAT sequence in the DECA^{ext} element of the *Amd1* promoter. To test this assumption, we generated an *Amd1* promoter mutant, in which the Pbx1:Prep1 recognition site was destroyed and simultaneously the CCTAT motif was replaced by a canonical Nf-y recognition motif (CCTAT to CCAAT in M5, Fig. 6*D*). Strikingly, Nf-y did not bind to this promoter mutant. This result suggests that direct DNA binding of Pbx1:Prep1 to the *Amd1* promoter is not only required for loading of Sp2 but also for binding of Nf-y.

A characteristic feature of the composite DECA^{ext} motif is the fixed distance of 11 bp between the Pbx1:Prep1 and the Nf-y recognition motifs. To examine whether the distance between these motifs is crucial for binding of Sp2, Nf-y, and Pbx1:Prep1, we generated *Amd1* mutant promoters with shortened (8 bp) or extended (14 bp) distances between the Pbx1:Prep1 and the Nf-y recognition motifs (M6 and M7 in Fig. 6*D*). Both, shortening as well as extending the distance between the Pbx1:Prep1 and the Nf-y–binding motifs greatly reduced binding of Sp2, Nf-y, and Pbx1:Prep1. This result strongly suggests that the 11-bp distance between the Pbx1:Prep1 and Nf-y–recognition site is crucial for establishing the Sp2–Pbx1:Prep1–Nf-y complex on DNA.

To further substantiate the conclusion that both Pbx1:Prep1 and Nf-y recognition sequences are necessary for loading of Sp2, we introduced mutations in another promoter (*Rplp0*), which contains two tandemly arranged DECA^{ext} motifs (Fig. 6*E*). Binding of Sp2, Nf-yb, Pbx1, and Prep1 was strongly reduced on mutating the two CCAAT motifs (*M1* in Fig. 6*E*). As expected, Pbx1 and Prep1 did not bind to an *Rplp0* promoter mutant where the Pbx1:Prep1 recognition motifs are mutated (*M2* in Fig. 6*E*). Remarkably, Sp2 and Nf-y still bound to the M1 and M2 mutants albeit at a markedly reduced level. Most importantly, binding of Pbx1, Prep1, and Nf-yb as well as Sp2 was completely abolished when both the recognition sequences of Pbx1:Prep1 and Nf-y were mutated simultaneously (*M3* in Fig. 6*E*). This result confirms the conclusion that the Pbx1:Prep1 and Nf-y recognition sequences are crucial for loading of Sp2 to promoters that contain the DECA^{ext} motif.

Finally, we tested whether the transgenic *Amd1* and *Rplp0* promoter fragments were transcriptionally active (Fig. S5). Compared with Flp-In™ NIH-3T3 cells carrying a promoterless luciferase construct, we observed an ~33- and 90-fold higher luciferase activity in extracts of cells carrying the *Amd1* or *Rplp0*-luciferase construct. Interestingly, none of the mutations in the *Amd1* promoter affected luciferase expression (Fig. S5*A*). In contrast, the M1 and M2 *Rplp0* promoter mutants were 2-fold, and the M3 *Rplp0* promoter mutant 10-fold less active than the WT promoter (Fig. S5*B*). Thus, the establishment of the Pbx1:Prep1–Nf-y–Sp2 complex is necessary for the full activity of the *Rplp0* but not for the *Amd1* promoter fragment. This observation suggests that binding of the Pbx1:Prep1–Nf-y–Sp2 complex to DECA^{ext} sites is not sufficient to drive gene expression, but may require other transcription factors. This finding is in line with a recent report that revealed that DECA and Nf-y sites-containing genomic elements are not sufficient to act as enhancers (26).

Pbx1:Prep1 physically interacts with Sp2 and recruits it to the DECA^{ext} motif in vitro

Considering that Pbx1:Prep1 together with Nf-y recruits Sp2 to the DECA^{ext} motifs *in vivo*, which in turn may strengthen DNA binding of Pbx1:Prep1 and Nf-y, we sought to reconstitute binding of Pbx1:Prep1, Nf-y, and Sp2 to the DECA^{ext} motif *in vitro* using recombinant proteins. The dimeric Pbx1:Prep1 complex was obtained by co-expressing GST–Pbx1 and Prep1 in *Escherichia coli*, and the trimeric Nf-y complex (Nf-ya:Nf-yb:Nf-yc) as well as DNA binding-deficient Sp2 mutants (Sp2(1–525) and Sp2(94–525)) were obtained by baculovirus-mediated expression in Sf9 cells (Fig. S6). We performed DNA affinity precipitation assays (DAPA) using a biotinylated oligonucleotide that contains the DECA^{ext} motif of the *Amd1* promoter as a probe (Fig. 7*A*). Pbx1:Prep1 and Nf-ya:b:c but not Sp2(1–525) bound to the WT DECA^{ext} motif oligonucleotide (Fig. 7*B*, *left upper panel*). However, Sp2(1–525) bound weakly to the DECA^{ext} motif oligonucleotide in the presence Nf-y. Strong binding was observed in the presence of either Pbx1:Prep1 alone or Pbx1:Prep1 and Nf-y (Fig. 7*B*, *right upper panel*). In contrast, the Sp2(94–525) mutant lacking the 93 N-terminal

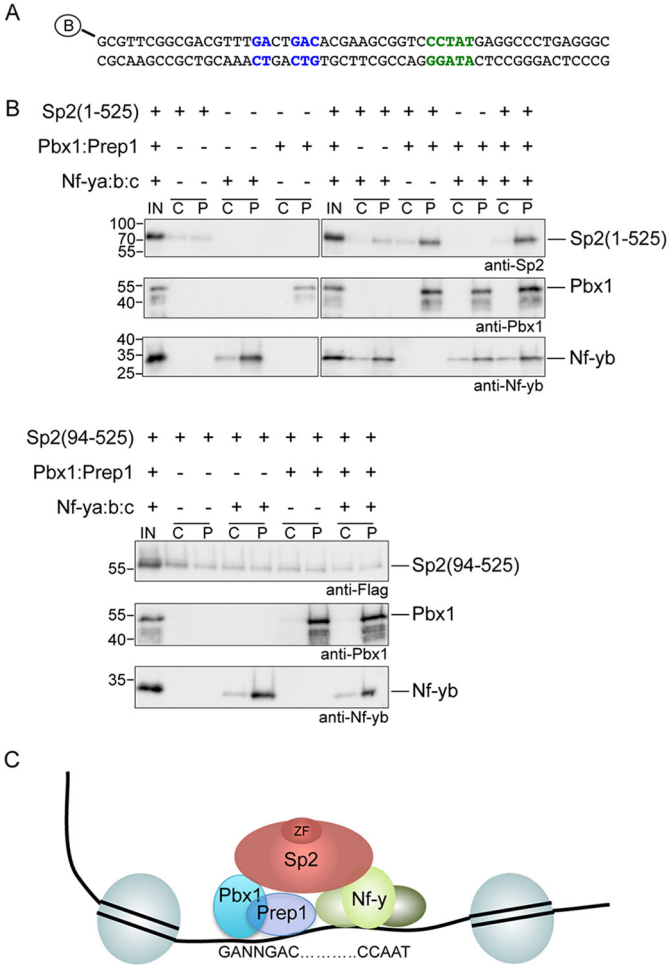

Figure 7. Sp2 binds directly to DNA-bound Pbx1:Prep1 via its N-terminal domain. *A*, sequence of the biotinylated double-stranded oligonucleotide used in DAPA. *B*, recombinant Sp2(1–525), Sp2(94–525), Pbx1:Prep1, and Nf-ya:b:c were incubated with the biotinylated double-stranded oligonucleotide as indicated. Bound proteins were detected by Western blotting. *IN* indicates input 10%; *C* indicates nonspecific binding control (magnetic beads); and *P* indicates probe (biotinylated oligonucleotide bound to magnetic beads). *C*, model diagram depicting the recruitment of Sp2 to its target promoters. Pbx1:Prep1 and Nf-ya:b:c bind to composite motifs where the Pbx1:Prep1 and Nf-y recognition sequences are separated by 11 bp. Sp2 interacts directly with Pbx1:Prep1 via its Sp-box and bridges Pbx1:Prep1 with Nf-y.

amino acids did not bind to the DECA^{ext} motif oligonucleotide, neither in the presence of Pbx1:Prep1 nor in the presence of Pbx1:Prep1 and Nf-y (Fig. 7*B*). These results show that Sp2 can physically interact with DNA-bound Pbx1:Prep1 *in vitro*. Collectively, our data suggest that Sp2 is tethered to the DECA^{ext} motifs *in vivo* by directly interacting with DNA-bound Pbx1:Prep1.

Discussion

The zinc finger domain of Sp2 binds to GC-rich sequences *in vitro* (11, 12). However, several lines of evidence revealed that the GC box is not the target sequence of Sp2 *in vivo*. (i) Sp2 mutants lacking the zinc finger DNA-binding domain bind to GC box-containing promoters as efficiently as WT Sp2. (ii) The GC box is not locally enriched in Sp2 ChIP peak summits. (iii) Mutation of the GC box in the *Amd1* promoter does not affect binding of endogenous Sp2. Our finding that the *in vitro* DNA binding motif of Sp2 is not the *in vivo* binding site suggests caution is required in the interpretation of ChIP-seq data. It appears to be risky to assume that the mere presence of a well-characterized DNA motif in a set of high-quality ChIP-seq peaks indicates direct binding of the corresponding transcription factor *in vivo*.

Our data suggest that Sp2 functions as a cofactor rather than as a DNA-binding factor. We show that Sp2 potentiates genomic binding of Pbx1:Prep1 and Nf-y to a composite motif, where the Pbx1:Prep1 and Nf-y recognition motifs are located at a fixed distance of 11 bp from each other. Importantly, Sp2's function to potentiate Pbx1:Prep1 and Nf-y binding does not require the zinc finger DNA-binding domain. Our results led us to propose a model whereby Pbx1:Prep1 and Nf-y bind directly to the DECA^{ext} motif. Sp2 is tethered by Pbx1:Prep1 and bridges Pbx1:Prep1 with Nf-y (Fig. 7*C*). Considering that Nf-y induces a strong bend in the DNA (27), Nf-y may impose spatial constraints necessary to establish the entire Sp2–Pbx1:Prep1–Nf-y complex.

An Sp2 mutant protein with a deletion of the Sp-box displayed strongly reduced genomic binding and failed to promote binding of Pbx1:Prep1, highlighting a critical role for this motif in establishing the Sp2–Pbx1:Prep1–Nf-y complex. The Sp-box is also found in other Sp1-related transcription factor family members (8) as well as in members of the NET (Noc, Nlz, Elbow, and Tlp-1) transcription factors (28, 29) but its molecular function is largely unknown. The Sp-box in these factors may also function as a protein–protein interaction domain. A physiological role of the Sp-box motif has been shown for the *Drosophila* Elb (Elbow) protein. The Sp-box in Elb is crucial for Elb's function in specification of polarization-sensitive photoreceptors in the dorsal rim area of the fly retina (30).

Pbx1 also dimerizes with Meis1 (Myeloid ecotropic viral integration site-1) (31, 32), another member of the TALE family of homeoproteins (24). Meis1 is expressed in mouse embryonic fibroblasts, where it can compete with Prep1 for Pbx1 binding (33). However, genomic Pbx1:Prep1- and Pbx1:Meis1-binding sites display different core motifs and largely do not overlap. Pbx1:Prep1 preferentially bind to promoter regions, whereas Pbx1:Meis1 predominantly bind to enhancers (24, 25). Visual inspection of genome browser tracks of a Meis1 ChIP-seq data set in MEFs (GEO58802) (33) revealed that the shared Sp2–Pbx1:Prep1–Nf-y binding sites were not bound by Meis1. Conversely, regions bound by Meis1 were not bound by Sp2. This observation suggests that the recruitment mechanism of Sp2 described here is specific for the Pbx1:Prep1 dimer.

Why is Sp2, which by itself does not bind to DNA, necessary for efficient binding of Pbx1:Prep1 and Nf-y to adjacent sites? Given that Sp2 is present at promoters of highly expressed ubiquitous genes, we hypothesize that the Sp2–Pbx1:Prep1–Nf-y complex may be required to establish or to sustain an active open chromatin state. In zebrafish embryos maternally expressed Prep1 occupies Pbx:Prep1 DECA motifs with nearby Nf-y sites already at blastula stage when many genomic loci are still occupied by nucleosomes suggesting a pioneer role of Pbx:Prep1 and Nf-y at DECA sites (26). Interestingly, Sp2 is conserved in zebrafish and, like Pbx, Prep1, and Nf-y, maternally expressed (34). Thus, it is tempting to speculate that Sp2 may

act also together with TALE factors and Nf-y during zebrafish embryogenesis.

Experimental procedures

Antibodies

We used affinity purified homemade antibodies for Western blotting and/or ChIP of Sp1 and Sp2 (6, 17). The following commercially available antibodies were used for detection of Pbx1, Prep1, and Nf-y: anti-Pbx1 (Cell Signaling Technology, catalog number 4342, 3 μg for ChIP, 1/1,000 dilution for Western blot), anti-Prep1 (Santa Cruz, catalog number sc-25282, 3 μg for ChIP, 1/500 dilution for Western blot), anti-Nf-ya (Santa Cruz, catalog number sc-10779, 3 μg for ChIP, 1/1,000 dilution for Western blot), anti-Nf-yb (Genespin, catalog number PAb001, 3 μg for ChIP, 1/2,000 dilution for Western blot), anti-Ring1b (Abcam, catalog number ab101273, 1/1,000 dilution for Western blot), anti-FLAG M2 (Sigma, catalog number F3165, 3 μg for ChIP, 1/2,000 dilution for Western blot), anti-tubulin (Merck Millipore, catalog number MAB3408, 1/15,000 dilution for Western blot).

Cell lines and cell growth conditions

The generation of WT and Sp2*ko* MEFs were described in Refs. 6 and 17. MEFs were cultured in a 1:1 mixture of Dulbecco's modified Eagle's medium/high glucose (4.5 g/liter) (Gibco Thermo Fisher Scientific) and Ham's F-10 (Gibco Thermo Fisher Scientific) supplemented with 10% (v/v) fetal calf serum (Sigma) and 1% penicillin-streptomycin. Sp2*ko* MEFs expressing Sp2 mutants were selected and propagated in the presence of 2 μg/ml of puromycin. Immunohistochemical detection of FLAG–Sp2 mutants shown in Fig. S3 was performed essentially as described previously (6, 35). The Flp-In^{TM} NIH-3T3 cell line was purchased from Thermo Fisher Scientific and cultured in Dulbecco's modified Eagle's medium/high glucose (4.5 g/liter) (Gibco Thermo Fisher Scientific) supplemented with 10% (v/v) bovine donor serum (Gibco Thermo Fisher Scientific) and 1% penicillin-streptomycin.

Knockdown of Pbx1 and Prep1

For RNAi-mediated depletion of mouse Pbx1 and Prep1, pools of five (Pbx1) and three (Prep1) siRNAs were used (Santa Cruz, sc-38797 and sc-38759). The siGenome nontargeting siRNA (Dharmacon number 001210-01) was used as a nonspecific control siRNA. WT MEFs on 15-cm plates were transfected with 40 nM siRNA using Oligofectamine (Invitrogen). Three days post-transfection, 2×10^6 cells were replated, and transfected a second time. An additional 3 days later, cells were collected and cross-linked chromatin was prepared. Knockdown efficiency was monitored by Western blotting.

Co-immunoprecipitation

For co-immunprecipitation of Pbx1 and Prep1, 500 μg of precleared nuclear extracts from WT and Sp2*ko* MEFs expressing 3×FLAG–Prep1 were incubated with 20 μl of equilibrated anti-FLAG M2-agarose (Sigma) for 2 h at 4 °C in a buffer containing 20 mM HEPES/KOH, pH 7.9, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 25% glycerol, 150 mM NaCl, 0.5 mM PMSF, 1× protease inhibitor mixture. Anti-FLAG-agarose–antigen complexes were washed four times in 25 mM HEPES/KOH, pH 7.9, 12.5 mM $MgCl_2$, 0.1 mM EDTA, 10% glycerol, 150 mM NaCl, 0,08% Nonidet P-40, 0.5 mM PMSF, 0.5× protease inhibitor mixture. Immune complexes were eluted with SDS sample buffer and subjected to Western blot analysis.

Construction of retroviral vectors and retroviral transduction

Retroviral expression plasmids for 3×FLAG–Sp2 mutants were generated by restriction cloning of PCR fragments into EcoRI–SalI restricted retroviral pBABE3×FLAG-puro. Primer sets used for PCR are available in Table S1. cDNA fragments encoding murine Pbx1a (UniProtKB P41778) and Prep1 (pKnox1, UniProtKB O70477) were obtained as gBlock DNA fragments from IDT (Integrated DNA Technologies) and cloned into BstBI-SalI–restricted pBABE3×FLAG-puro. The production of virus stocks, infection of MEFs and Flp-In^{TM} NIH-3T3 cells, and the selection of transduced cells were as described (17).

Cloning of Sp2 target promoters and targeted integration in Flp-In^{TM} NIH-3T3 cells

The cytomegalovirus promoter of the pcDNA^{TM}5/FRT targeting plasmid (Thermo Fisher Scientific) was removed by BglII–XhoI restriction and replaced by a BglII–SalI fragment containing the mouse *Sp2* promoter (−341 to +77) fused to the luciferase gene. The resulting pcDNA5_FRT_mSp2prom-Luc plasmid was used as a starting plasmid for cloning of other Sp2 target promoters. The *Ptpn18* promoter (−577 to + 41) was obtained by PCR amplification of genomic mouse DNA. The *Pdia6* (−237 to +7), *Amd1* (−128 to +134), and *Rplp0* promoters (−257 to +37) as well as corresponding mutants (Fig. 6 and Fig. S4) were obtained as gBlock DNA fragments from IDT. All promoters were cloned by replacing the *Sp2* promoter in the pcDNA5_FRT_mSp2prom-Luc plasmid through restriction with BglII and Asp718I.

The pcDNA5_FRT_promoter-Luc plasmids were transfected along with the Flp recombinase plasmid pOG44 into Flp-In^{TM} NIH-3T3 cells according to the manufacturer's instructions using the Lipofectamine^{TM} 3000 transfection reagent (Thermo Fisher Scientific). Single clones were obtained by Hygromycin B selection (150 μg/ml). Integration into the FRT site was verified by testing for their Zeocin sensitivity (600 μg/ml).

Expression and purification of recombinant proteins

The vector for co-expression of GST-Pbx1 and Prep1 in *E. coli* was described in Ref. 36. Expression of GST–Pbx1:Prep1 in BL21 cells was induced with 0.1 mM isopropyl 1-thio-D-galactopyranoside for 20 h at 16 °C. Bacterial lysates were collected by centrifugation, washed with PBS, and resuspended in lysis buffer containing 25 mM HEPES/KOH, pH 7.6, 0.1 mM EDTA, 12.5 mM $MgCl_2$, 10% glycerol, 500 mM NaCl, 0.1% Nonidet P-40, 0.5 mg/ml of lysozyme, 0.5 mM DTT, 0.5 mM PMSF, 1× protease inhibitor mixture (Roche Applied Science). Cells were sonicated, and cleared lysates were incubated with GSH-Sepharose 4B beads (GE Healthcare). Pbx1:Prep1 was eluted by cleavage with 2 units/μl of PreScission^{TM} protease (GE Health-

care) in 50 mM Tris/Cl, pH 7.5, 150 mM NaCl, 1 mM EDTA, 1 mM DTT, 0.01% Nonidet P-40 for 16 h at 4 °C.

Transfer vectors for generation of baculovirus expressing His-tagged Sp2(1–525), 3×FLAG-tagged Sp2(94–525), and His-tagged Nf-ya along with untagged Nf-yb and Nf-yc (His–Nf-ya:b:c) were constructed by PCR cloning of the respective murine open reading frames into pFBDM (37). Transformation of the pFBDM plasmids into DH10 MultiBac^{Cre} *E. coli* cells, preparation of bacmid DNA, and infection of Sf9 cells was carried out as described in Ref. 38. Baculovirus were amplified three times and then used to infect Sf9 cells for protein production. Cells were harvested 3–4 days after infection. His–Nf-ya:b:c and His–Sp2(1–525) were purified using nickel-nitrilotriacetic acid-agarose (Qiagen); and 3×FLAG–Sp2(94–525) was purified using anti-FLAG M2 affinity gel (Sigma). The eluted proteins were dialyzed twice against storage buffer (50 mM Tris/Cl, pH 8.0, 150 mM NaCl, 0.5 mM EDTA, 10% glycerol, 1 mM DTT, 0.5 mM PMSF) and frozen at −80 °C.

ChIP-exo and ChIP-seq

For Sp2 ChIP-exo analysis we used the ChIP-exo kit from Active Motif in accordance with the manufacturer's instructions. Cross-linked chromatin was generated from 2×10^7 WT and Sp2*ko* MEFs. Three individual ChIP reactions with 500 μg of sheared chromatin were performed for each cell type. Six micrograms of affinity-purified homemade Sp2 antibody #1 (6) were used per ChIP experiment. Enzymatic reactions and DNA purification steps were performed as described in the ChIP-exo kit manual. After PCR amplification, reactions were pooled and purified with AMPure magnetic beads (Beckman Coulter). The DNA library was size selected by agarose gel electrophoresis (size range 180–350 bp) and extracted DNA fragments were purified with QIAquick columns (Qiagen).

Conventional ChIPs of Pbx1, Prep1, and FLAG–Sp2 were performed as described in Ref. 11 using the One Day ChIP kit (Diagenode). Three micrograms of antibody were used per ChIP experiment, and 5 ng of precipitated DNA were used for indexed sequencing library preparation using the Microplex library preparation kit v2 (Diagenode). DNA libraries were sequenced on an Illumina HiSeq1500 platform (Illumina Inc.), rapid-run mode, single-read 50 bp (HiSeq SR Rapid Cluster Kit v2, HiSeq Rapid SBS Kit v2, 50 cycles) according to the manufacturer's instructions.

ChIP-seq data analysis

Raw Sp2 ChIP-exo reads were aligned using Subread version 1.4.3-p1 (39) against the *Mus musculus* genome assembly from Ensembl revision 74. Peak discovery was performed using MACE 1.1 with default parameters except for "-fold = 1.5" (40). Peaks were filtered to have at least 100 reads in WT MEFs, less than 50 reads in Sp2*ko* MEFs, and at least a 3-fold enrichment in WT MEFs over the Sp2*ko* control.

Raw Pbx1, Prep1, and FLAG–Sp2 ChIP-seq reads were aligned using Subread (39) version 1.4.3-p1. Reads matching multiple locations were discarded during alignment. Peaks were called with MACS (41) version 1.4.0rc2 against an IgG control ChIP (6). Filtered peaks were required to have at least 30 tags and a sequencing depth-corrected ratio over control of 3-fold. For motif search and heat maps, peaks were centered at their summits and fixed sized regions were extracted. Summits were defined as the point of highest read overlap after extending the reads to 200 bp. Heat maps show the number of reads extended to 200 bp, normalized for sequencing depth. The signal distribution was truncated at the 99th percentile in each sample to increase contrast. Regions for heat maps were ordered by the sum of signal in the first sample depicted.

Motif analysis

De novo motif search including Tomtom and CentriMo was performed online with MEME-ChIP version 4.12.0 (http://meme-suite.org/tools/meme-chip)[3] (42) within the MEME Suite (http://meme-suite.org)[3] (20) using 100 bp (±50 bp) sequences surrounding peak summits.

Gene ontology analysis

The DAVID 6.8 web-based tool (https://david.ncifcrf.gov) was used for GO analyses (43, 44).

ChIP-qPCR

ChIP-qPCRs with gene-specific primers were performed using the ImmoMix PCR reagent (Bioline) in the presence of 0.1× SYBR Green (Molecular Probes, Thermo Fisher). Enrichment was calculated relative to input. Primer sets used for ChIP-qPCR are available in Table S1.

Luciferase reporter assays

For luciferase reporter assays 3×10^4 Flp-In^{TM} NIH-3T3 cells carrying transgenic promoter-luciferase constructs were seeded on a 24-well plate and grown for 4 days. Cells were lysed in 100 μl of passive lysis buffer and luciferase activity was quantified using the Luciferase® Reporter Assay System (Promega) in an AutoLumat Plus LB 953 reader (Berthold Technologies). Firefly luciferase levels were normalized to total protein levels. For each cell line two independent experiments were performed in quadruplicate.

DAPA

One μg of biotinylated double-stranded oligonucleotides were incubated with 50 μl of equilibrated MagneSphere paramagnetic particles (Promega) for 1 h at room temperature. Unbound oligonucleotides were removed by washing the beads three times with 1× DWP-150 buffer (20 mM HEPES/KOH, pH 7.9, 0.2 mM EDTA, 150 mM NaCl, 1.5 mM $MgCl_2$, 10% glycerol, 10 μM $ZnSO_4$, 0.3% Triton X-100, 1 mM DTT, 1 mg/ml of BSA, 0.5 mM PMSF, 0.5× protease inhibitor mixture (Roche Applied Science)). Coupled oligonucleotides were incubated with 500 ng of purified His–Sp2(1–525), FLAG–Sp2(94–525), His–Nf-ya:b:c and Pbx1:Prep1 proteins in 150 μl of 1× DWP-150 buffer for 2 h at 4 °C. Unbound proteins were removed by washing the beads three times with 250 μl of 1× DWP-150 buffer. Bound proteins were eluted with 2× Laemmli SDS-PAGE sample buffer for 10 min at 100 °C and subsequently analyzed by Western blotting.

[3] Please note that the JBC is not responsible for the long-term archiving and maintenance of this site or any other third party hosted site.

Databases and data availability

Our ChIP-seq data were deposited at ArrayExpress under accession number E-MTAB-7125. For assessing the overlap of Pbx1 and Prep1 with Sp2 and Nf-y, we used our previously published ChIP-seq data sets for Sp2 (E-MTAB-994) and Nf-y (E-MTAB-2970).

Author contributions—S. V., B. S., and G. S. data curation; S. V., B. S., F. F., D. B., and A. N. investigation; S. V. and B. S. methodology; S. V., B. S., and G. S. writing-original draft; F. F. formal analysis; F. F. and G. S. visualization; T. S. resources; G. S. conceptualization; G. S. supervision; G. S. funding acquisition; G. S. writing-review and editing.

Acknowledgments—We thank Anna Mary Steitz, Irene Klassen, and Xenia Schmidt for their contribution to the acquisition and collection of data, and Iris Rohner for excellent technical assistance. Chiara Bruckmann and Franceso Blasi (IFOM, Milan, Italy) generously provided the bacterial expression vector for co-expression of GST-Pbx1 and Prep1. We thank Alexander Brehm, Jaqueline Mermoud, and Colin Dingwall for discussion and critical reading of the manuscript.

References

1. Slattery, M., Zhou, T., Yang, L., Dantas Machado, A. C., Gordân, R., and Rohs, R. (2014) Absence of a simple code: how transcription factors read the genome. *Trends Biochem. Sci.* **39,** 381–399 CrossRef Medline
2. Lelli, K. M., Slattery, M., and Mann, R. S. (2012) Disentangling the many layers of eukaryotic transcriptional regulation. *Annu. Rev. Genet.* **46,** 43–68 CrossRef Medline
3. Yip, K. Y., Cheng, C., Bhardwaj, N., Brown, J. B., Leng, J., Kundaje, A., Rozowsky, J., Birney, E., Bickel, P., Snyder, M., and Gerstein, M. (2012) Classification of human genomic regions based on experimentally determined binding sites of more than 100 transcription-related factors. *Genome Biol.* **13,** R48 CrossRef Medline
4. Xie, D., Boyle, A. P., Wu, L., Zhai, J., Kawli, T., and Snyder, M. (2013) Dynamic trans-acting factor colocalization in human cells. *Cell* **155,** 713–724 CrossRef Medline
5. Foley, J. W., and Sidow, A. (2013) Transcription-factor occupancy at HOT regions quantitatively predicts RNA polymerase recruitment in five human cell lines. *BMC Genomics* **14,** 720 CrossRef Medline
6. Völkel, S., Stielow, B., Finkernagel, F., Stiewe, T., Nist, A., and Suske, G. (2015) Zinc finger independent genome-wide binding of Sp2 potentiates recruitment of histone-fold protein Nf-y distinguishing It from Sp1 and Sp3. *PLOS Genet.* **11,** e1005102 CrossRef Medline
7. Suske, G., Bruford, E., and Philipsen, S. (2005) Mammalian SP/KLF transcription factors: bring in the family. *Genomics* **85,** 551–556 CrossRef Medline
8. Suske, G. (1999) The Sp-family of transcription factors. *Gene* **238,** 291–300 CrossRef Medline
9. Schaeper, N. D., Prpic, N. M., and Wimmer, E. A. (2010) A clustered set of three Sp-family genes is ancestral in the Metazoa: evidence from sequence analysis, protein domain structure, developmental expression patterns and chromosomal location. *BMC Evol. Biol.* **10,** 88 CrossRef Medline
10. Philipsen, S., and Suske, G. (1999) A tale of three fingers: the family of mammalian Sp/XKLF transcription factors. *Nucleic Acids Res.* **27,** 2991–3000 CrossRef Medline
11. Terrados, G., Finkernagel, F., Stielow, B., Sadic, D., Neubert, J., Herdt, O., Krause, M., Scharfe, M., Jarek, M., and Suske, G. (2012) Genome-wide localization and expression profiling establish Sp2 as a sequence-specific transcription factor regulating vitally important genes. *Nucleic Acids Res.* **40,** 7844–7857 CrossRef Medline
12. Moorefield, K. S., Fry, S. J., and Horowitz, J. M. (2004) Sp2 DNA binding activity and trans-activation are negatively regulated in mammalian cells. *J. Biol. Chem.* **279,** 13911–13924 CrossRef Medline
13. Hagen, G., Müller, S., Beato, M., and Suske, G. (1992) Cloning by recognition site screening of two novel GT box binding proteins: a family of Sp1 related genes. *Nucleic Acids Res.* **20,** 5519–5525 CrossRef Medline
14. Hagen, G., Dennig, J., Preiss, A., Beato, M., and Suske, G. (1995) Functional analyses of the transcription factor Sp4 reveal properties distinct from Sp1 and Sp3. *J. Biol. Chem.* **270,** 24989–24994 CrossRef Medline
15. Marin, M., Karis, A., Visser, P., Grosveld, F., and Philipsen, S. (1997) Transcription factor Sp1 is essential for early development but dispensable for cell growth and differentiation. *Cell* **89,** 619–628 CrossRef Medline
16. Bouwman, P., Göllner, H., Elsässer, H. P., Eckhoff, G., Karis, A., Grosveld, F., Philipsen, S., and Suske, G. (2000) Transcription factor Sp3 is essential for post-natal survival and late tooth development. *EMBO J.* **19,** 655–661 CrossRef Medline
17. Baur, F., Nau, K., Sadic, D., Allweiss, L., Elsässer, H. P., Gillemans, N., de Wit, T., Krüger, I., Vollmer, M., Philipsen, S., and Suske, G. (2010) Specificity protein 2 (Sp2) is essential for mouse development and autonomous proliferation of mouse embryonic fibroblasts. *PloS One* **5,** e9587 CrossRef Medline
18. Suske, G. (2017) NF-Y and SP transcription factors: new insights in a long-standing liaison. *Biochim. Biophys. Acta* **1860,** 590–597 CrossRef
19. Rhee, H. S., and Pugh, B. F. (2011) Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. *Cell* **147,** 1408–1419 CrossRef Medline
20. Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., Ren, J., Li, W. W., and Noble, W. S. (2009) MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* **37,** W202–W208 CrossRef Medline
21. Dolfini, D., Zambelli, F., Pavesi, G., and Mantovani, R. (2009) A perspective of promoter architecture from the CCAAT box. *Cell Cycle* **8,** 4127–4137 CrossRef Medline
22. Blasi, F., Bruckmann, C., Penkov, D., and Dardaei, L. (2017) A tale of TALE, PREP1, PBX1, and MEIS1: interconnections and competition in cancer. *BioEssays* **39,** 10.1002/bies.201600245 CrossRef
23. Berthelsen, J., Zappavigna, V., Mavilio, F., and Blasi, F. (1998) Prep1, a novel functional partner of Pbx proteins. *EMBO J.* **17,** 1423–1433 CrossRef Medline
24. Longobardi, E., Penkov, D., Mateos, D., De Florian, G., Torres, M., and Blasi, F. (2014) Biochemistry of the Tale transcription factors PREP, MEIS, and PBX in vertebrates. *Dev. Dyn.* **243,** 59–75 CrossRef Medline
25. Penkov, D., Mateos San Martin, D., Fernandez-Díaz, L. C., Rosselló, C. A., Torroja, C., Sánchez-Cabo, F., Warnatz, H. J., Sultan, M., Yaspo, M. L., Gabrieli, A., Tkachuk, V., Brendolan, A., Blasi, F., and Torres, M. (2013) Analysis of the DNA-binding profile and function of TALE homeoproteins reveals their specialization and specific interactions with Hox genes/proteins. *Cell Rep.* **3,** 1321–1333 CrossRef Medline
26. Ladam, F., Stanney, W., Donaldson, I. J., Yildiz, O., Bobola, N., and Sagerström, C. G. (2018) TALE factors use two distinct functional modes to control an essential zebrafish gene expression program. *Elife* **7,** e36144 CrossRef Medline
27. Nardini, M., Gnesutta, N., Donati, G., Gatta, R., Forni, C., Fossati, A., Vonrhein, C., Moras, D., Romier, C., Bolognesi, M., and Mantovani, R. (2013) Sequence-specific transcription factor NF-Y displays histone-like DNA binding and H2B-like ubiquitination. *Cell* **152,** 132–143 CrossRef Medline
28. Pereira, F., Duarte-Pereira, S., Silva, R. M., da Costa, L. T., and Pereira-Castro, I. (2016) Evolution of the NET (NocA, Nlz, Elbow, TLP-1) protein family in metazoans: insights from expression data and phylogenetic analysis. *Sci. Rep.* **6,** 38383 CrossRef Medline
29. Nakamura, M., Runko, A. P., and Sagerström, C. G. (2004) A novel subfamily of zinc finger genes involved in embryonic development. *J. Cell. Biochem.* **93,** 887–895 CrossRef Medline
30. Wernet, M. F., Meier, K. M., Baumann-Klausener, F., Dorfman, R., Weihe, U., Labhart, T., and Desplan, C. (2014) Genetic dissection of photoreceptor subtype specification by the *Drosophila melanogaster* zinc finger proteins Elbow and No ocelli. *PLOS Genet.* **10,** e1004210 CrossRef Medline
31. Knoepfler, P. S., Calvo, K. R., Chen, H., Antonarakis, S. E., and Kamps, M. P. (1997) Meis1 and pKnox1 bind DNA cooperatively with Pbx1 uti-

lizing an interaction surface disrupted in oncoprotein E2a-Pbx1. *Proc. Natl. Acad. Sci. U.S.A.* **94,** 14553–14558 CrossRef Medline

32. Berthelsen, J., Kilstrup-Nielsen, C., Blasi, F., Mavilio, F., and Zappavigna, V. (1999) The subcellular localization of PBX1 and EXD proteins depends on nuclear import and export signals and is modulated by association with PREP1 and HTH. *Genes Dev.* **13,** 946–953 CrossRef Medline
33. Dardaei, L., Longobardi, E., and Blasi, F. (2014) Prep1 and Meis1 competition for Pbx1 binding regulates protein stability and tumorigenesis. *Proc. Natl. Acad. Sci. U.S.A.* **111,** E896–E905 CrossRef Medline
34. Xie, J., Yin, H., Nichols, T. D., Yoder, J. A., and Horowitz, J. M. (2010) Sp2 is a maternally inherited transcription factor required for embryonic development. *J. Biol. Chem.* **285,** 4153–4164 CrossRef Medline
35. Sapetschnig, A., Koch, F., Rischitor, G., Mennenga, T., and Suske, G. (2004) Complexity of translationally controlled transcription factor Sp3 isoform expression. *J. Biol. Chem.* **279,** 42095–42105 CrossRef Medline
36. Mathiasen, L., Bruckmann, C., Pasqualato, S., and Blasi, F. (2015) Purification and characterization of a DNA-binding recombinant PREP1:PBX1 complex. *PloS One* **10,** e0125789 CrossRef Medline
37. Berger, I., Fitzgerald, D. J., and Richmond, T. J. (2004) Baculovirus expression system for heterologous multiprotein complexes. *Nat. Biotechnol.* **22,** 1583–1587 CrossRef Medline
38. Fitzgerald, D. J., Berger, P., Schaffitzel, C., Yamada, K., Richmond, T. J., and Berger, I. (2006) Protein complex expression by using multigene baculoviral vectors. *Nat. Methods* **3,** 1021–1032 CrossRef Medline
39. Liao, Y., Smyth, G. K., and Shi, W. (2013) The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res.* **41,** e108 CrossRef Medline
40. Wang, L., Chen, J., Wang, C., Uusküla-Reimand, L., Chen, K., Medina-Rivera, A., Young, E. J., Zimmermann, M. T., Yan, H., Sun, Z., Zhang, Y., Wu, S. T., Huang, H., Wilson, M. D., Kocher, J. P., and Li, W. (2014) MACE: model based analysis of ChIP-exo. *Nucleic Acids Res.* **42,** e156 CrossRef Medline
41. Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., Nusbaum, C., Myers, R. M., Brown, M., Li, W., and Liu, X. S. (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9,** R137 CrossRef Medline
42. Machanick, P., and Bailey, T. L. (2011) MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics* **27,** 1696–1697 CrossRef Medline
43. Huang da, W., Sherman, B. T., and Lempicki, R. A. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4,** 44–57 CrossRef Medline
44. Huang da, W., Sherman, B. T., and Lempicki, R. A. (2009) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* **37,** 1–13 CrossRef Medline