Title

Tissue-specific SMARCA4 binding at active and repressed regulatory elements during embryogenesis.

Permalink

https://escholarship.org/uc/item/4gk8x2v0

Journal

Genome research, 24(6)

ISSN

1088-9051

Authors

Attanasio, Catia
Nord, Alex S
Zhu, Yiwen
et al.

Publication Date

2014-06-01

DOI

10.1101/gr.168930.113

Peer reviewed

Tissue-specific SMARCA4 binding at active and repressed regulatory elements during embryogenesis

Catia Attanasio,[1,6] Alex S. Nord,[1,6] Yiwen Zhu,[1] Matthew J. Blow,[2] Simon C. Biddie,[2,3]
Eric M. Mendenhall,[4] Jesse Dixon,[5] Crystal Wright,[2] Roya Hosseini,[1] Jennifer A. Akiyama,[1]
Amy Holt,[1] Ingrid Plajzer-Frick,[1] Malak Shoukry,[1] Veena Afzal,[1] Bing Ren,[5]
Bradley E. Bernstein,[4] Edward M. Rubin,[1,2] Axel Visel,[1,2] and Len A. Pennacchio[1,2,7]

[1]Genomics Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA; [2]US Department of Energy Joint Genome Institute, Walnut Creek, California 94598, USA; [3]Addenbrooke's Hospital, Cambridge University NHS Trust, Cambridge CB2 0QQ, United Kingdom; [4]HHMI and Department of Pathology, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts 02114, USA; [5]Ludwig Institute for Cancer Research, UCSD School of Medicine, La Jolla, California 92093, USA

The SMARCA4 (also known as BRG1 in humans) chromatin remodeling factor is critical for establishing lineage-specific chromatin states during early mammalian development. However, the role of SMARCA4 in tissue-specific gene regulation during embryogenesis remains poorly defined. To investigate the genome-wide binding landscape of SMARCA4 in differentiating tissues, we engineered a Smarca4[FLAG] knock-in mouse line. Using ChIP-seq, we identified ~51,000 SMARCA4-associated regions across six embryonic mouse tissues (forebrain, hindbrain, neural tube, heart, limb, and face) at mid-gestation (E11.5). The majority of these regions was distal from promoters and showed dynamic occupancy, with most distal SMARCA4 sites (73%) confined to a single or limited subset of tissues. To further characterize these regions, we profiled active and repressive histone marks in the same tissues and examined the intersection of informative chromatin states and SMARCA4 binding. This revealed distinct classes of distal SMARCA4-associated elements characterized by activating and repressive chromatin signatures that were associated with tissue-specific up- or down-regulation of gene expression and relevant active/repressed biological pathways. We further demonstrate the predicted active regulatory properties of SMARCA4-associated elements by retrospective analysis of tissue-specific enhancers and direct testing of SMARCA4-bound regions in transgenic mouse assays. Our results indicate a dual active/repressive function of SMARCA4 at distal regulatory sequences in vivo and support its role in tissue-specific gene regulation during embryonic development.

[Supplemental material is available for this article.]

SMARCA4, also known as BRG1 in humans, is a major catalytic subunit of the SWI/SNF chromatin remodeling complex (Saha et al. 2006), which is involved in stem cell renewal and pluripotency (Ho et al. 2009; Kidder et al. 2009) and induction of lineage-specific gene programs (de la Serna et al. 2006; Ho and Crabtree 2010). An active role of SMARCA4 in cell fate determination has been reported in a broad range of differentiation programs such as neurogenesis (Bultman et al. 2000; Seo et al. 2005; Zhan et al. 2011; Yu et al. 2013), erythropoiesis (Armstrong et al. 1998; Griffin et al. 2008; Kim et al. 2009), myogenesis (Simone et al. 2004; Zhang et al. 2011), T cell (Zhao et al. 1998), and heart development (Stankunas et al. 2008; Hang et al. 2010). Its essential role during development has been further confirmed by targeted disruption of *Smarca4* in mice resulting in embryonic lethality (Bultman et al. 2000). In addition, germline mutations affecting normal SMARCA4 activity in vivo are associated with a number of developmental phenotypes (Bilodeau et al. 2006; Hang et al. 2010; Takeuchi et al. 2011), and its role as a tumor suppressor is well documented in cancer (Decristofaro et al. 2001; Schneppenheim et al. 2010; Wilson and Roberts 2011; Robinson et al. 2012). SMARCA4 control of gene expression occurs via both activation and repression of transcription (Battaglioli et al.

2002; Chi et al. 2002; Zhang et al. 2002; Ho et al. 2009; Hang et al. 2010; Zhan et al. 2011), potentially through direct interaction between SMARCA4 and histone modification enzymes or transcription factors at gene promoters and distal regulatory elements (Battaglioli et al. 2002; Hassan et al. 2002; Seo et al. 2005; Chandrasekaran and Thompson 2007; Rada-Iglesias et al. 2011; Lee et al. 2012; Yu et al. 2013). While genetic evidence from human and mouse studies indicates a critical role for SMARCA4 in multiple biological processes, the genome-wide binding profile and association of SMARCA4 with regulatory sequences in vivo remain poorly explored. In this study, we describe genome-wide maps of SMARCA4 binding in multiple embryonic mouse tissues, which revealed tissue-specific in vivo association of SMARCA4 with both active and repressive distant-acting regulatory sequences.

Results

To facilitate the genome-wide mapping of SMARCA4 during mouse development, we used homologous recombination in embryonic stem cells (ESCs) to generate a knock-in mouse strain in which the

endogenous SMARCA4 protein carries a carboxy-terminal FLAG epitope (Fig. 1; Supplemental Fig. S1) and where the Neo selection cassette was removed to prevent any interference with proper *Smarca4* expression (Supplemental Fig. S1A,C). Correct expression of the FLAG-tagged protein was detectable via Western blot (Supplemental Fig. S1E), and ChIP-seq analysis of SMARCA4[FLAG] targets in ESCs recapitulated previously reported endogenous SMARCA4 binding (Supplemental Fig. S2; Kidder et al. 2009). Following germline transmission of the engineered *Smarca4* allele, heterozygous *Smarca4[FLAG]* mice were intercrossed. Offspring had the expected Mendelian genotype ratios and no obvious phenotypes were observed in homozygous targeted mice, suggesting that the FLAG epitope does not interfere with normal function of SMARCA4 (Bultman et al. 2000; Bultman et al. 2005).

To map SMARCA4 binding sites in vivo, we collected forebrain, hindbrain, neural tube, face, limb, and heart tissue from *Smarca4[FLAG]* mouse embryos at embryonic day 11.5 (E11.5) and performed ChIP-seq, targeting the FLAG epitope (Fig. 1; Methods). Sequencing reads were mapped to the mouse genome (mm9) and SMARCA4 peaks were identified in each tissue. Across all six tissues we identified 51,204 sites, which rises to a total of 55,967 unique SMARCA4-enriched regions identified in the mouse genome when ESC peaks are included. Enriched regions within 1 kb of an annotated RefSeq or UCSC transcription start site (TSS) were classified as promoter-proximal ($n = 14,513$) and the remaining sites ($n = 41,454$) as distal (Fig. 2A). Distal SMARCA4-bound regions were strongly enriched for sequence regions exhibiting significant evolutionary constraint (Supplemental Fig. S3) with tissue-specific differences in relative constraint levels of SMARCA4-bound regions consistent with previous studies of distal enhancers (Blow et al. 2010; Nord et al. 2013), suggesting that distal SMARCA4 binding occurs at regulatory elements.

To explore the dynamic properties of SMARCA4 in vivo, we characterized the tissue specificity of SMARCA4 binding sites across embryonic mouse tissues and ESCs. SMARCA4 enrichment at all merged SMARCA4 regions was scored for enrichment within each tissue based on coverage derived from the respective ChIP-seq

data set (see Methods). For these analyses, we considered all E11.5 tissues and ESCs. We first examined promoter-associated (proximal) SMARCA4-bound regions and observed that the majority (82%) were enriched in ≥4 samples, with only 4% specific to a single tissue or to ESCs. In comparison to proximal SMARCA4 binding, distal regions bound by SMARCA4 across embryonic mouse tissues were highly tissue specific, with 73% exhibiting enrichment in ≤3 of the samples surveyed and 28% specific to a single tissue or to ESCs (Fig. 2B). To illustrate these genome-wide patterns, we plotted SMARCA4 enrichment near TSSs (Fig. 2C) and in distal regions (Fig. 2D) across three representative tissues (forebrain, limb, and heart). These tissue-specific distal SMARCA4-bound regions are enriched near genes involved in expected tissue-specific biological processes and mouse phenotypes (Supplemental Tables S1–S3). This genome-wide analysis indicates that SMARCA4 is bound at both proximal and distal sites in the genome, but that distal SMARCA4 binding sites are more dynamic across tissues.

Recent in vitro observations indicate that SMARCA4 binding occurs at distal regulatory elements, including at enhancers (De et al. 2011; Hu et al. 2011; Rada-Iglesias et al. 2011). To test the in vivo regulatory capability of distal regions exhibiting enrichment of SMARCA4, we used an established transgenic mouse enhancer assay to examine 54 randomly chosen SMARCA4-bound candidate enhancer sequences. Each candidate sequence was cloned upstream of a minimal promoter driving the expression of a *lacZ* reporter gene, which was then used to generate embryos via pronuclear injection. Only patterns present in at least three independent E11.5 embryos were considered reproducible (Pennacchio et al. 2006; Visel et al. 2009). Of the 54 sequences tested, 25 (46%) performed as enhancers in vivo, confirming that distal SMARCA4-bound regions can activate tissue-specific transcription in vivo (Fig. 2E; Supplemental Table S4). However, the correlation between SMARCA4 tissue-specific binding and tissue-specific in vivo enhancer activity observed in transgenic embryos (Supplemental Fig. S4) was much weaker than for other marks of active enhancers such as EP300 (Visel et al. 2009). This is further observed in a retrospective analysis of all tested sequences ($n = 1747$) published on the VISTA Enhancer Browser (http://enhancer.lbl.gov/; Visel et al. 2007) that are bound in any of the analyzed tissues by SMARCA4 (Supplemental Fig. S5). This indicates that while SMARCA4 binding occurs at developmental enhancers, it is not consistently associated with in vivo enhancer activity of the bound region in the matched tissue. This observation suggests that SMARCA4, rather than being universally associated with active enhancers, is associated with more complex tissue-specific regulatory chromatin states in vivo.

To functionally stratify the nature of the SMARCA4-bound regions and gain insights into their role in gene regulation in vivo, we generated ChIP-seq data for both activating (H3K4me3, H3K4me1, H3K27ac) and repressive histone marks (H3K27me3) in forebrain and limb tissues. We also used E11.5 forebrain and limb transcriptomes (RNA-seq data sets,
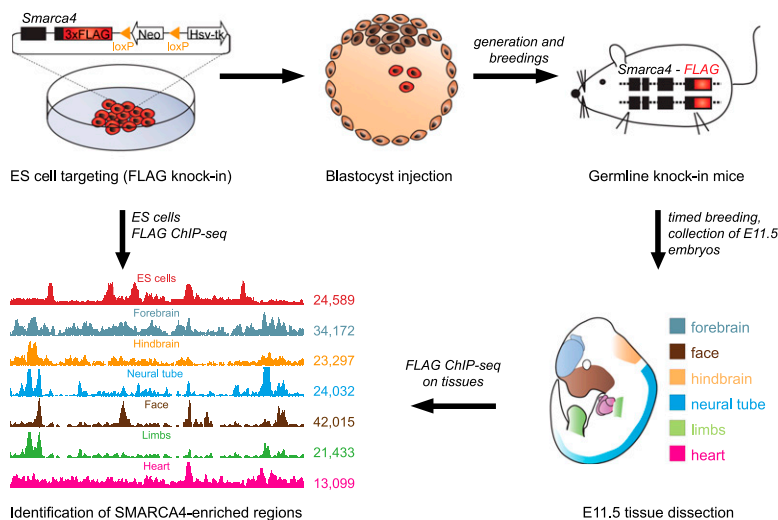
Figure 1. Overview of *Smarca4[FLAG]* mouse generation and SMARCA4 binding site mapping in vivo. A knock-in mouse strain carrying a FLAG epitope at the carboxy terminus of SMARCA4 was engineered and used to perform ChIP-seq from a panel of embryonic tissues. For each ChIP-seq data set, the total number of identified enriched SMARCA4 regions is indicated near the corresponding coverage map (schematics).
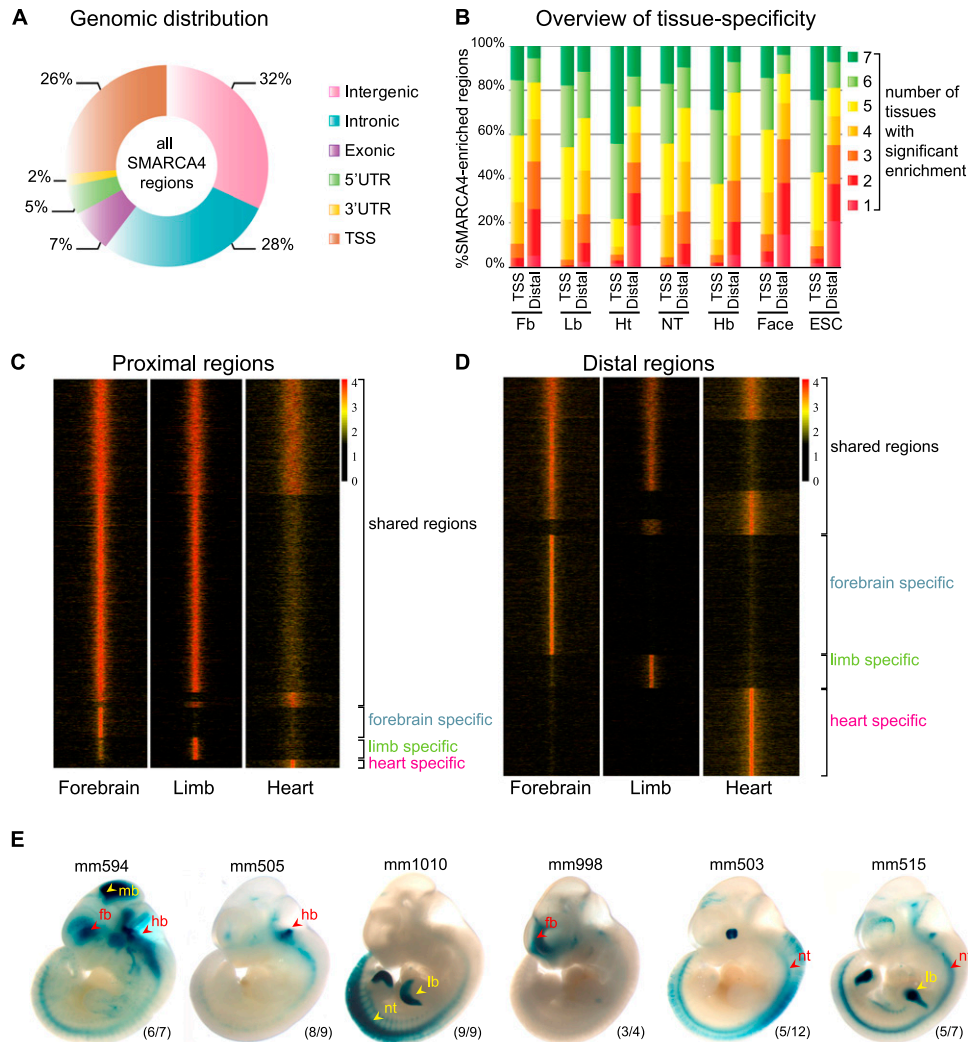
Figure 2. ChIP-seq identifies SMARCA4-enriched regions genome-wide. (*A*) Genomic distribution of SMARCA4-enriched regions. (*B*) Tissue-specific SMARCA4 enrichment in TSS versus distal regions in each data set. The bar graph illustrates the lower percentage of tissue-specific SMARCA4-bound regions among the proximal sites (TSS) versus distal sites by indicating the number of tissues in which a specific SMARCA4 region is significantly enriched (red, present in one tissue; dark green, present in seven tissues). (*C*,*D*) Heatmaps displaying SMARCA4 coverage across three tissues (forebrain, limb, and heart). Each row of heatmap represents one enhancer, with coverage plotted across the 10 kb surrounding enrichment peak. (*C*) SMARCA4 enrichment proximal to TSSs. (*D*) SMARCA4 enrichment at distal regions. (*E*) Selection of six reproducible SMARCA4-bound enhancers at E11.5, illustrating SMARCA4 activity across different tissues. Reproducible predicted patterns are indicated by red arrows while unpredicted reproducible tissue-specific activity is indicated by yellow arrows. (Fb) Forebrain; (Mb) midbrain; (Hb) hindbrain; (Nt) neural-tube; (Lb) limbs; (Ht) heart. The reproducibility of the indicated patterns is indicated for each embryo (number of observed/total number of embryos).

see Methods) to correlate the chromatin state at SMARCA4 genomic loci with neighboring gene expression.

For proximal SMARCA4 regions (forebrain n = 10,737; limb n = 10,193; combined n = 11,234), we observed that almost all (94% forebrain, 98% limb, 95% overall) regions are associated with H3K4me3, a hallmark of active promoters (Hon et al. 2009). Two states at active proximal regions are further differentiated by the presence or absence of the H3K27me3 histone modification as (1) active promoters (no H3K27me3) and (2) bivalent promoters (with H3K27me3) (Fig. 3A,B; Supplemental Table S5; Bernstein et al. 2006; Mikkelsen et al. 2007). Analysis of the level of expression and known function of the genes associated with proximal SMARCA4 regions revealed that predicted active promoters are associated with high levels of relative expression and strong enrichment for general/housekeeping functional terms. In comparison, predicted bivalent promoters show low levels of expression and are associated with tissue-specific developmental terms, consistent with previous studies (Fig. 3C,D; Supplemental Fig. S6; Bernstein et al. 2006; Mikkelsen et al. 2007; Pan et al. 2007; Zhao et al. 2007). Overall, SMARCA4 enrichment is present at 89% of all forebrain and 79% of all limb promoters marked by H3K4me3 or H3K27me3. Interestingly, SMARCA4-enriched promoters show significantly increased levels of expression compared with promoters that are not bound by SMARCA4 for both active and bivalent classes (Supplemental Fig. S7).

For distal regions (forebrain n = 12,269; limb n = 6,918; combined n = 13,570), SMARCA4 enrichment is largely accompanied by H3K4me1 across both tissues (78% forebrain, 97% limb,
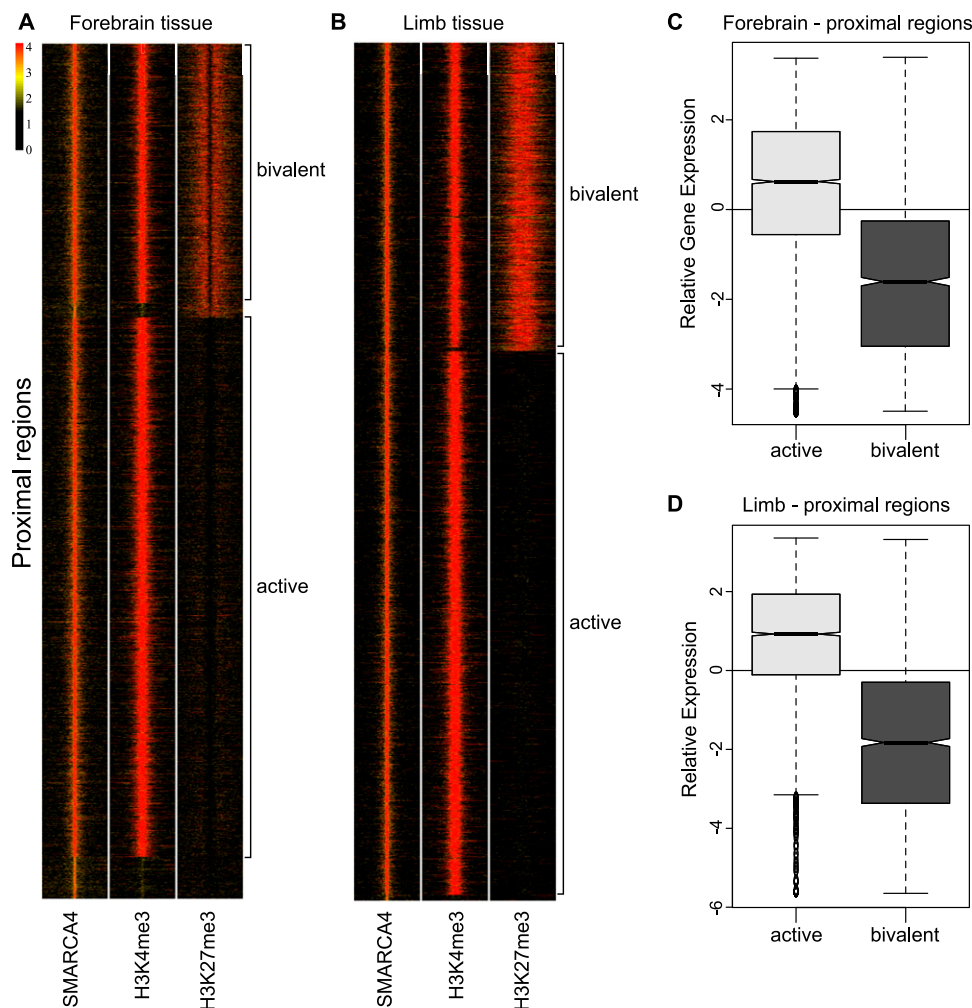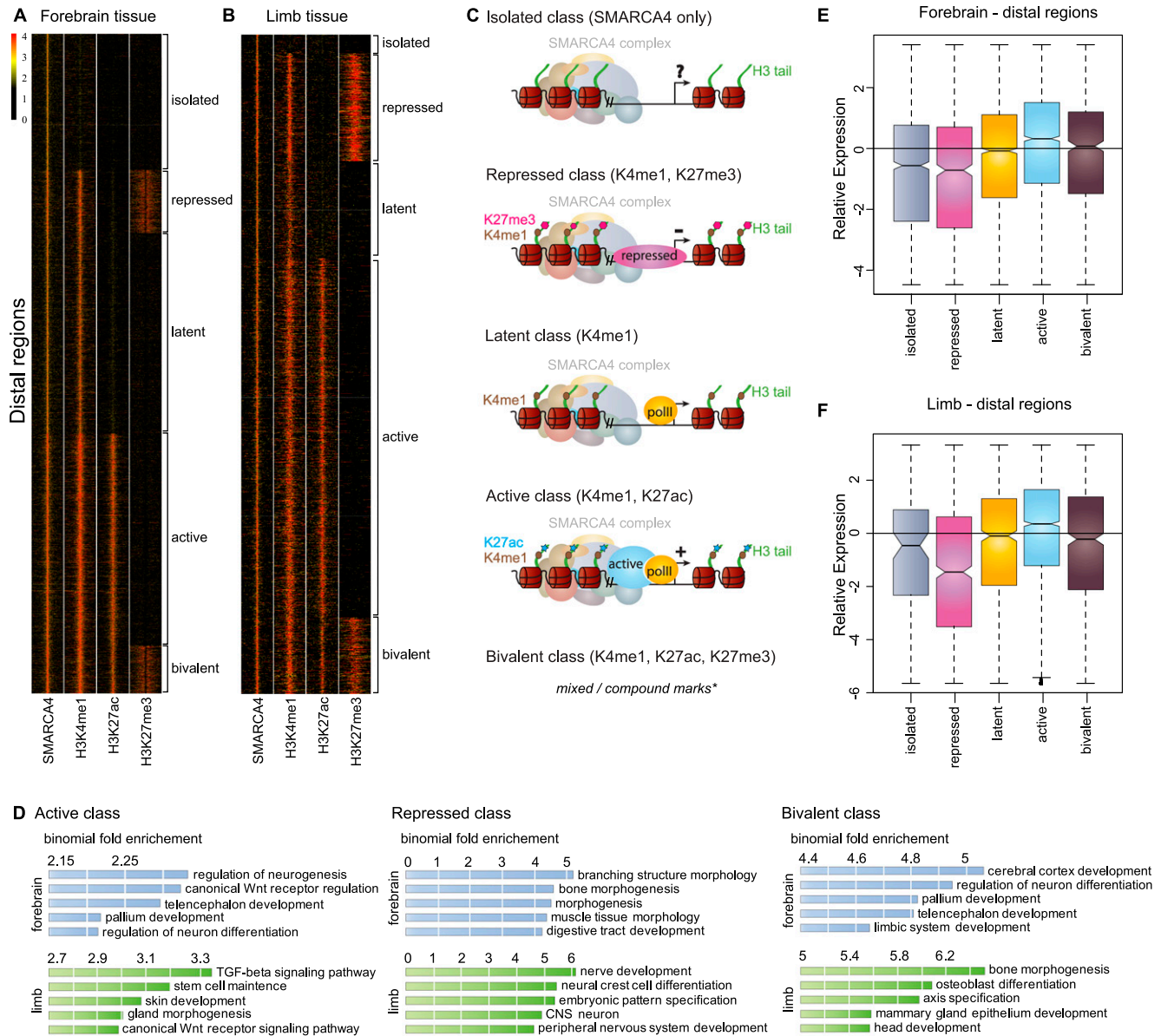
Figure 3. Proximal SMARCA4 enrichment is associated with active and bivalent promoters. (*A*,*B*) Heatmaps displaying coverage for SMARCA4 and informative histone marks (H3K4me3 and H3K27me3) for proximal regions. Each row of heatmap represents one enhancer, with coverage plotted across the 10 kb surrounding enrichment peak. SMARCA4 enrichment at proximal regions for forebrain (*n* = 10,737) (*A*) and limb (*n* = 10,193) (*B*). (*C*,*D*) Gene expression profile associated with SMARCA4 proximal regions classified as active or bivalent. RNA-seq data were generated for E11.5 forebrain and limb and the distribution of expression levels of the nearest gene was compared for SMARCA4-enriched regions in forebrain (*C*) and limb (*D*).

83% combined), suggesting that SMARCA4 distal regions are strongly enriched for regulatory regions. These sequences are further characterized by local enrichment patterns for H3K27ac and H3K27me3, and can be grouped into five classes: (1) isolated SMARCA4 binding with no associated histone marks, (2) latent class (H3K4me1), (3) active class (H3K4me1 with H3K27ac), (4) repressed class (H3K4me1 with H3K27me3), and (5) bivalent class (H3K4me1 with H3K27ac and H3K27me3) (Fig. 4A–C; Supplemental Table S6). Ontology analysis (McLean et al. 2010) revealed that active, repressed, and bivalent classes are strongly enriched for tissue-specific activated or repressed biological processes characteristic of forebrain or limb development (Fig. 4D). Consistent with expectations based on histone modifications, expression of the nearest transcript is increased for the active class and decreased for the repressed class across both tissues. Isolated SMARCA4-enriched elements are also associated with significantly (forebrain) or suggestively (limb) decreased levels of expression (Fig. 4E,F). Expression levels for latent and bivalent classes are not significantly different from mean levels in either tissue. We additionally examined classi-

fied distal forebrain and limb SMARCA4 regions for the presence of the enhancer-associated transcriptional co-activator EP300 using previously published data sets (Visel et al. 2009). In both tissues, the maximum enrichment levels for EP300 are observed in the active and bivalent classes (Supplemental Fig. S8), consistent with previous findings that EP300 marks tissue-specific in vivo enhancers (Visel et al. 2009) and with the predicted activity of the active (and bivalent) SMARCA4 classes.

In contrast to the overall SMARCA4 enrichment at promoter regions, SMARCA4 enrichment at distal putative regulatory elements, as identified by histone modification profiling, is present at lower frequency, with only 56% of forebrain and 21% of limb putative regulatory elements also enriched for SMARCA4. The difference between forebrain and limb is largely driven by an increased number of latent sites (H3K4me1) without evidence of SMARCA4 binding in limb (91%) versus forebrain (55%), which may be due to biological or technical differences. While there is suggestive evidence that within some classes SMARCA4-bound elements (e.g., bivalent limb elements) are associated with differ-

Figure 4. Histone marks associated with activation and repression co-localize with SMARCA4. (*A,B*) Heatmaps displaying coverage for SMARCA4 and informative histone marks for proximal (H3K4me3 and H3K27me3) and distal (H3K4me1, H3K27ac, H3K27me3) regions. Each row of heatmap represents one enhancer, with coverage plotted across the 10 kb surrounding enrichment peak. SMARCA4 enrichment at distal regions for forebrain (*n* = 12,269) (*A*) and limb (*n* = 6918) (*B*). (*C*) Classified regulatory groups based on histone signatures. (*) Bivalent SMARCA4 elements show a mix of enrichment patterns. (*D*) Differential enrichment for functional annotation terms associated with distal SMARCA4 regions categorized by histone signature. Shown are the top five enriched "Biological process terms" specific to distal forebrain (blue) and limbs (green) differentially marked SMARCA4 regions. (*E,F*) Transcription associated with SMARCA4 regions that can be classified as active or repressed based on co-occurring histone marks. RNA-seq data were generated for E11.5 forebrain and limb and the distribution of expression levels of the nearest gene was compared for SMARCA4-enriched regions in forebrain (*E*) and limb (*F*) classified by histone signature.

ences in expression or functional annotation compared to sites with similar histone modifications but no SMARCA4 binding, no consistent significant differences were observed (Supplemental Fig. S9).

Intriguingly, local H3K27me3 enrichment profiles at both proximal and distal sites differed between forebrain and limb, with forebrain H3K27me3 binding exhibiting depletion directly at the SMARCA4 peak, whereas in limb no such localized depletion is evident (Figs. 3A,B, 4A,B). Despite this difference, expression and

ontology analysis of H3K27me3-enriched chromatin states were comparable between the tissues. For both TSS and distal regions, we observe significant differences in chromatin state at common forebrain and limb SMARCA4-enriched regions (Supplemental Figs. S10A,C, S11A,C). Where chromatin state differs between forebrain and limb at shared SMARCA4 sites, we observe differential gene expression and tissue-specific functional term enrichment consistent with the active/repressed state model (Supplemental Figs. S10B,D, S11B,D; Supplemental Table S7).

To test the predictive value of our functional SMARCA4 classification in vivo, we overlapped all distal SMARCA4-enriched regions identified in the forebrain and limb with a set of 1747 mammalian noncoding sequences that were previously tested for in vivo enhancer activity in transgenic mice (http://enhancer.lbl.gov). A total of 220 forebrain and 180 limb distal SMARCA4-bound regions overlapped tested sequences representing all histone-derived regulatory classes (Supplemental Table 8). The predictive value of our model was supported by the observed correlation of chromatin state at SMARCA4-enriched regions and tissue-specific activity pattern of the previously identified enhancers for both forebrain and limb tissues (Fig. 5A; Supplemental Fig. S12). For example, among the 73 previously characterized in vivo enhancers overlapping predicted forebrain-active SMARCA4 elements, 35 (48%) indeed act as forebrain enhancers in vivo, while only 6/21 (29%) are active in forebrain in the predicted repressed SMARCA4 class. Overall, of the active and bivalent classes, all but bivalent forebrain enhancers (P-value = 0.06) were significantly enriched (P-value < 0.0005) for in vivo tissue-specific activity relative to background rates, whereas none of the latent or repressed classes significantly differed from background (Fisher's exact test for all significance estimates). Furthermore, in the retrospective analysis of the 54 elements initially tested for this study, only elements categorized as active or bivalent in forebrain or limb drove expression in the respective tissue.

To examine whether this retrospective analysis is consistent with forward predictions, we selected nine regions for testing in this transgenic reporter assay that exhibited SMARCA4 enrichment in forebrain and limb, but were predicted to be active in the forebrain and repressed in the limb. Of the nine sequences tested, four (44%) drove patterns in the forebrain and none drove expression in the limb (Supplemental Table 4). Representative SMARCA4-enriched distal regions exhibiting differential histone signatures in forebrain and limb are shown in Figure 5B. Taken together, this combination of retrospective and forward testing of in vivo functional activity corroborates the regulatory activity model derived from genome-wide analysis of SMARCA4.

Discussion

The SMARCA4 chromatin remodeling factor is required for normal development since its targeted deletion in mice causes embryonic lethality (Bultman et al. 2000), yet little has been known about its in vivo genomic targets and involvement in regulatory pathways during tissue ontogenesis. Here we report the generation of a knock-in mouse line where *Smarca4* is tagged with a FLAG epitope, providing a novel tool for functional studies of SMARCA4 in vivo. Using tissues collected from a *Smarca4*[FLAG] mouse line, we show that SMARCA4 interacts with gene promoters and distal enhancer sequences, with predominant binding to distal sequences as previously reported in SMARCA4 cell line studies (Ho et al. 2009; Euskirchen et al. 2011). SMARCA4-bound promoters are generally consistent across tissues, though we observed differential active and bivalent promoter chromatin signatures at proximal SMARCA4 regions that were correlated with gene function and expression. In contrast, distal SMARCA4 binding is highly dynamic and tissue-specific during development. We show that distal SMARCA4 binding occurs primarily at regulatory elements, and that these regions fall into characteristic chromatin states that are strongly associated with the activation or repression of tissue-specific developmental processes in vivo. We further demonstrate that expected tissue-specific active or inactive enhancer state corresponds to in vivo enhancer activity in transgenic mice. Unlike epigenomic markers that are singularly associated with active or
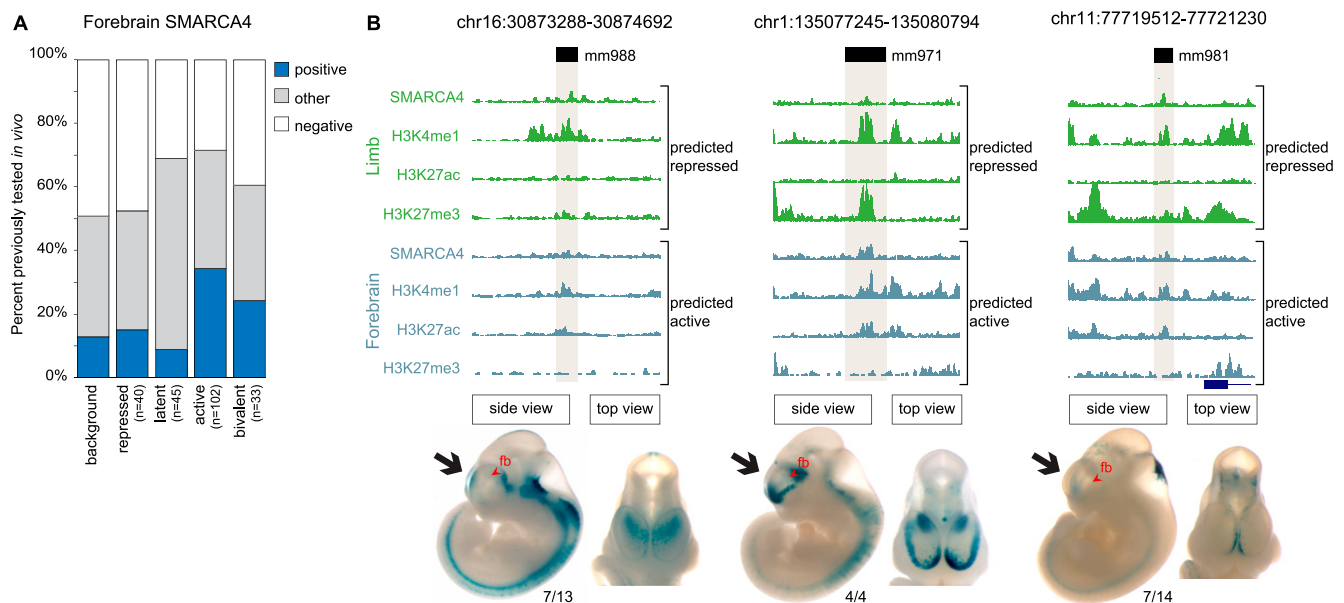
Figure 5. SMARCA4 marks developmental enhancers in an active or repressed state. (*A*) Intersection of forebrain SMARCA4-enriched regions with 1747 published mouse sequences tested in the transgenic assay shows that SMARCA4 marks enhancers and the histone signature at SMARCA4-enriched loci predicts tissue-specific activity. (Positive) Reproducible enhancer activity in forebrain; (other) reproducible enhancer activity in a tissue other than forebrain; (negative) no detectable enhancer activity in vivo. (*B*) Representative examples of regions with characteristic forebrain and limb chromatin state showing predicted tissue specificity in transgenic assays. One representative embryo is displayed for each region and the reproducibility of the observed pattern is indicated (number of observed/total number of embryos) for each. The predicted activity in forebrain (fb) is shown both in the *side* views (red arrows) and *top* views (the angle from which the image was captured is indicated on the *side* views by a black arrow).

repressed regulatory elements or chromatin states, such as EP300 or H3K27me3, profiling SMARCA4-interacting regions identifies a complex combination of regulatory elements representing inactive, active, and repressed states. The integrative analysis presented here demonstrates that the combination of genome-wide SMARCA4 and histone modification data can be used to functionally characterize SMARCA4-bound regions, suggesting that the complexity of SMARCA4 activity during development can be made biologically interpretable via this approach.

In this study, direct interrogation of heterogeneous tissues during development necessarily generates results across a variety of cell types, which can make interpretation of tissue-derived data difficult. However, our results suggest that this approach captures processes not recapitulated by examination of individual cell lines. The genomic patterns observed here were robust, even though we were not able to perform analysis of biological replicates due to the large amount of *Smarca4*^{FLAG} embryonic tissue initially required per experiment (Methods). Conclusions made from these genomic data sets represent initial findings that will be informed by future studies on the mechanism of *Smarca4* function in developing tissues. Additionally, SMARCA4 enrichment is associated with turnover in histone modifications, as predicted to occur during transitions between active and repressed states. Combined with uncertainty regarding regulatory target genes, these factors could further confound relationships between gene expression, functional enrichment, and transgenic enhancer activity for chromatin-based classification of SMARCA4-bound elements. Despite these limitations, the patterns of embryonic SMARCA4 binding and correlated enhancer-associated histone modifications suggest that SMARCA4 is substantially involved both in activation and repression pathways during lineage differentiation and development.

While it remains to be demonstrated how exactly SMARCA4 binding is related to establishing and maintaining different chromatin states, the strong association of SMARCA4 with active, repressed, and bivalent chromatin signatures and corresponding functional gene classes and expression levels raises the possibility that SMARCA4 is required in these processes during development. Further studies will be required to resolve the mechanism and interacting proteins via which SMARCA4 contributes to establishing and maintaining chromatin states, to clarify whether SMARCA4-interacting distal regulatory targets lose or gain functionality in the absence of SMARCA4, and whether SMARCA4 itself plays a direct role in activation or repression. The early embryonic lethality associated with loss of SMARCA4 function suggests, however, that large-scale alterations of gene activation and repression drive genome-wide disruption of chromatin state mechanics associated with lineage differentiation. Understanding the tissue-specific genome-wide changes in chromatin structure driven by SMARCA4 activity, over time during development, will thus be critical toward understanding the role of SMARCA4 in development and disease (Nord et al. 2013). The genome-wide maps of in vivo SMARCA4 binding sites during development generated in this study provide a foundation for understanding the tissue-specific regulatory activity of this key protein.

Methods

All procedures of this study involving animals were reviewed and approved by the Animal Welfare and Research Committee at Lawrence Berkeley National Laboratory.

Smarca4 knock-in mice

To study the genome-wide binding profile of the SMARCA4 transcriptional coactivator during mouse development we generated a knock-in mouse strain in which 3×FLAG tags were inserted into *Smarca4* through homologous recombination in mouse ESCs. Briefly, we first introduced a "3×FLAG-stop codon" cassette into a ploxPN2T targeting vector (Ahituv et al. 2007). Subsequently, a 5′ long homology arm and a 3′ short homology arm were cloned upstream and downstream, respectively, of the 3×FLAG-stop codon cassette, to generate an in-frame *Smarca4*^{FLAG} allele (see primer list in Supplemental Table 9). The "homology arms" 3×FLAG vector was then electroporated into W4/129S6 ES cells (Taconic) and recombination events were selected by treatment of cells with G418 (150 μg/mL) and 0.2 μM FIAU for 7 d. PCR screening for correct targeting was carried out on the surviving ES colonies as previously described (Ahituv et al. 2007), followed by Neomycin cassette removal (Supplemental Fig. S1B,C). To verify accurate mRNA expression from the *Smarca4*^{FLAG} allele, RT-PCR was performed on *Smarca4*^{FLAG} +/− colonies and resulting RNA was checked by Sanger sequencing (Supplemental Fig. S1D). Finally, *Smarca4*^{FLAG} positive ESCs were injected into C57BL/6J blastocysts and chimeras were obtained. Further breeding was carried out to ultimately obtain homozygous *Smarca4*^{FLAG} mice.

SMARCA4 Western blot

Smarca4^{FLAG} ESCs or mouse tissues were lysed in RIPA buffer (Sigma) in the presence of Protease Inhibitor cocktail (Sigma) for 1–2 h at 4°C with constant agitation. Standard Bradford protocol was then used to determine the protein concentration in each lysate. Protein was denatured and loaded on NuPAGE Novex 3%–8% Tris-Acetate gradient gel (Invitrogen) following the manufacturer's protocol. Following electrophoresis, protein transfer on PVDF membranes was achieved with the XCell II Blot Module (Invitrogen). FLAG detection was obtained by incubation of blocked PVDF membranes with ANTI-FLAG M2 monoclonal antibody-peroxidase conjugate (Sigma, A8592) for 30–60 min and staining with TMB Liquid Substrate solution (Sigma).

FLAG ChIP-seq

FLAG ChIP-seq was performed on *Smarca4*^{FLAG} ESCs and E11.5 mouse embryonic tissues as previously described (Kim et al. 2007). Briefly, ESCs/tissues were cross-linked with 1% formaldehyde at room temperature for 15 min, incubated in lysis buffer and sonicated with a Bioruptor sonicator (Diagenode). Solubilized chromatin was incubated overnight at 4°C with 20 μg of mouse monoclonal anti-FLAG M2 antibody (Sigma, F1804) coupled to M-280 Sheep anti-Mouse IgG Dynabeads (Invitrogen). Immunoprecipitated chromatin was then purified by standard phenol-chloroform extraction and DNA libraries were prepared with the Illumina TruSeq kit according to the manufacturer's protocol. Libraries were sequenced on an Illumina GA II or HiSeq 2000 instrument. DNA sequencing and sequence base calls were made using standard Illumina methods. Resulting 1×36-bp sequences were filtered to remove sequencing artifacts and adaptors and then mapped to the mouse genome (mm9) using the BWA algorithm (Li and Durbin 2009).

BWA call
bwa aln -t 6 -l 25 mm9 sample.fastq.gz

Peaks were called using MACS (version 1.4) (Zhang et al. 2008), with settings optimized for broad peaks.

MACS call

```
macs14 -t chip.bam –control=input.bam –name=chip_output
   –format=BAM –gsize=mm –tsize=36 –bw=300 –mfold=10,30
   –nolambda –nomodel –shiftsize=150 -p 0.00001
```

Input control libraries were generated and used for peak calling and downstream analysis for all except three data sets (hindbrain, face, and ESCs), which were run using the same parameters but with no input data set. Due to the substantial mouse breeding and tissue collection requirement for E11.5 FLAG ChIP-seq experiments, biological replicates were not performed. However, specific tissues were pooled to reduce effects of inter-individual biological variation, with an average of ~70–250 E11.5 embryos included per experiment depending on the analyzed tissue.

Histone ChIP-seq

ChIP-seq data sets for four histone marks (H3K4me1, H3K4me3, H3K27ac, and H3K27me3—Abcam ab8895, ab8580, ab4729, and Millipore 07-449, respectively) were generated for E11.5 limbs, while for E11.5 forebrain we produced ChIP-seq for H3K4me1, H3K4me3, and H3K27ac and used a previously generated H3K27ac data set (Nord et al. 2013). Histone ChIP-seq was carried out using previously described protocols (Bernstein et al. 2006). DNA libraries were prepared, sequenced, and analyzed as described for FLAG ChIP-seq. All histone data sets were generated with an input control.

RNA-seq

Mouse E11.5 forebrain and limb tissues were collected from normal CD-1 embryos and frozen. Poly(A) RNA was then extracted by the combined use of Ambion RNAqueous Kit (total RNA extraction) and MicroPoly(A)Purist [poly(A) RNA extraction]. RNA libraries were prepared with the Illumina TruSeq protocol according to the manufacturer's protocol for RNA-seq libraries. Short-read data were generated using the Illumina GA II sequencing platform. Reads were mapped to the mouse genome (NCBI37/mm9) using the TopHat pipeline with default parameters (Trapnell et al. 2012), mapping both spliced and unspliced reads. Gene FPKM (fragments per kilobase of transcript per million reads sequenced) generated using Cufflinks, with all transcripts mapped back to known mm9 genes.

In vivo *lacZ* transgenesis

Candidate SMARCA4-enriched sequences were PCR amplified from mouse genomic DNA by PCR and cloned using the Gateway technology (Invitrogen) upstream of a minimal Hsp68 promoter driving the expression of a *lacZ* reporter (Kothary et al. 1988; Pennacchio et al. 2006). Each construct was then injected into mouse fertilized eggs, implanted into pseudopregnant foster females, and F_0 embryos were collected at E11.5. Whole embryos were stained for *lacZ* activity and resulting patterns were evaluated for reproducibility: Only patterns observed in at least three different transgenic embryos were considered reproducible enhancers. Details of all transgenic experiments, including images of all transgenic embryos displaying annotated expression patterns, will be available at the VISTA Enhancer Browser (http://enhancer.lbl.gov/; Visel et al. 2007) and are reported in Supplemental Table 4. For retrospective analysis, background rates of enhancer activity in any E11.5 tissue, and specifically in forebrain and limb, were estimated based on overall rates across the complete set of 1747 tested elements. In this retrospective analysis, estimated success rates should be interpreted on a relative scale rather than as an accurate absolute estimation, as these elements were selected based on other criteria (EP300 binding in tissues and/or high levels of evolutionary conservation).

Computational analyses of SMARCA4-enriched regions

Peaks identified across ESCs and the six E11.5 tissues (forebrain, limb, heart, neural tube, hindbrain, face) were merged, with the final regions representing the outermost summits of all component peaks (base position with the highest SMARCA4 enrichment) plus 1-kb flanking on each side. Enriched regions were classified as promoter-proximal if the region boundaries were within 1 kb of a TSS from the combined collection of UCSC and RefSeq genes (mm9). Within each enriched region, individual ChIP-seq data sets were examined for local enrichment. Where ChIP-seq coverage exceeded the 95th percentile of genomic background coverage across at least 250 bp of the enriched region and immediate flanking sequence, the specific region was scored as enriched. Use of different criteria for enrichment level and enriched region size produced generally consistent results. Genomic background was calculated via random sampling of mappable regions across 1 Mb total. Where available, coverage was corrected using matched input control data. For ChIP-seq data sets lacking an input control (ESCs, hindbrain, and face), SMARCA4 binding estimates may be overestimated and a higher false discovery rate of tissue-specific SMARCA4 elements is possible. Coverage heatmaps were generated by centering and scaling corrected ChIP-seq coverage using mean and standard deviation, and plotting normalized coverage across 10 kb centered on the SMARCA4-enriched region. Analysis was performed using standard computational tools (Li and Durbin 2009; Quinlan and Hall 2010) and custom R scripts for ChIP-seq enrichment analysis, which are available on request. Functional enrichment was performed using the GREAT tool (McLean et al. 2010) with default settings for distal regions and using the nearest TSS for proximal regions. Functional enrichment analysis showed strong enrichment for matched tissue-specific processes for all ChIP-seq data sets, including those lacking input controls.

We observed locally overlapping enrichment of both H3K27me3 and H3K27ac, two histone marks previously reported to be mutually exclusive (Rada-Iglesias et al. 2011), at the majority of bivalent SMARCA4 elements in the limb but not forebrain (Figs. 3A,B, 4A,B). We suspect that increased cellular heterogeneity of E11.5 limb tissue may drive this apparent difference, as we observe similar patterns for H3K27me3 at bivalent promoters and repressed distal sites. Regardless of the differences between forebrain and limb with respect to H3K27me3, expression and functional analysis of the relevant states (proximal bivalent and distal repressed and bivalent) is generally consistent between the tissues (Figs. 3, 4; Supplemental Figs. S10, S11).

Data access

All ChIP-seq data sets (FLAG and histones) have been submitted to the NCBI Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) under accession number GSE37151.

Acknowledgments

References

Ahituv N, Zhu Y, Visel A, Holt A, Afzal V, Pennacchio LA, Rubin EM. 2007. Deletion of ultraconserved elements yields viable mice. *PLoS Biol* **5:** e234.

Armstrong JA, Bieker JJ, Emerson BM. 1998. A SWI/SNF-related chromatin remodeling complex, E-RC1, is required for tissue-specific transcriptional regulation by EKLF in vitro. *Cell* **95:** 93–104.

Battaglioli E, Andres ME, Rose DW, Chenoweth JG, Rosenfeld MG, Anderson ME, Mandel G. 2002. REST repression of neuronal genes requires components of the hSWI.SNF complex. *J Biol Chem* **277:** 41038–41045.

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.

Bilodeau S, Vallette-Kasic S, Gauthier Y, Figarella-Branger D, Brue T, Berthelet F, Lacroix A, Batista D, Stratakis C, Hanson J, et al. 2006. Role of Brg1 and HDAC2 in GR *trans*-repression of the pituitary POMC gene and misexpression in Cushing disease. *Genes Dev* **20:** 2871–2886.

Blow MJ, McCulley DJ, Li Z, Zhang T, Akiyama JA, Holt A, Plajzer-Frick I, Shoukry M, Wright C, Chen F, et al. 2010. ChIP-Seq identification of weakly conserved heart enhancers. *Nat Genet* **42:** 806–810.

Bultman S, Gebuhr T, Yee D, La Mantia C, Nicholson J, Gilliam A, Randazzo F, Metzger D, Chambon P, Crabtree G, et al. 2000. A *Brg1* null mutation in the mouse reveals functional differences among mammalian SWI/SNF complexes. *Mol Cell* **6:** 1287–1295.

Bultman SJ, Gebuhr TC, Magnuson T. 2005. A Brg1 mutation that uncouples ATPase activity from chromatin remodeling reveals an essential role for SWI/SNF-related complexes in β-globin expression and erythroid development. *Genes Dev* **19:** 2849–2861.

Chandrasekaran R, Thompson M. 2007. Polybromo-1-bromodomains bind histone H3 at specific acetyl-lysine positions. *Biochem Biophys Res Commun* **355:** 661–666.

Chi TH, Wan M, Zhao K, Taniuchi I, Chen L, Littman DR, Crabtree GR. 2002. Reciprocal regulation of CD4/CD8 expression by SWI/SNF-like BAF complexes. *Nature* **418:** 195–199.

De S, Wurster AL, Precht P, Wood WH III, Becker KG, Pazin MJ. 2011. Dynamic BRG1 recruitment during T helper differentiation and activation reveals distal regulatory elements. *Mol Cell Biol* **31:** 1512–1527.

de la Serna IL, Ohkawa Y, Imbalzano AN. 2006. Chromatin remodelling in mammalian differentiation: lessons from ATP-dependent remodellers. *Nat Rev Genet* **7:** 461–473.

Decristofaro MF, Betz BL, Rorie CJ, Reisman DN, Wang W, Weissman BE. 2001. Characterization of SWI/SNF protein expression in human breast cancer cell lines and other malignancies. *J Cell Physiol* **186:** 136–145.

Euskirchen GM, Auerbach RK, Davidov E, Gianoulis TA, Zhong G, Rozowsky J, Bhardwaj N, Gerstein MB, Snyder M. 2011. Diverse roles and interactions of the SWI/SNF chromatin remodeling complex revealed using global approaches. *PLoS Genet* **7:** e1002008.

Griffin CT, Brennan J, Magnuson T. 2008. The chromatin-remodeling enzyme BRG1 plays an essential role in primitive erythropoiesis and vascular development. *Development* **135:** 493–500.

Hang CT, Yang J, Han P, Cheng HL, Shang C, Ashley E, Zhou B, Chang CP. 2010. Chromatin regulation by Brg1 underlies heart muscle development and disease. *Nature* **466:** 62–67.

Hassan AH, Prochasson P, Neely KE, Galasinski SC, Chandy M, Carrozza MJ, Workman JL. 2002. Function and selectivity of bromodomains in anchoring chromatin-modifying complexes to promoter nucleosomes. *Cell* **111:** 369–379.

Ho L, Crabtree GR. 2010. Chromatin remodelling during development. *Nature* **463:** 474–484.

Ho L, Ronan JL, Wu J, Staahl BT, Chen L, Kuo A, Lessard J, Nesvizhskii AI, Ranish J, Crabtree GR. 2009. An embryonic stem cell chromatin remodeling complex, esBAF, is essential for embryonic stem cell self-renewal and pluripotency. *Proc Natl Acad Sci* **106:** 5181–5186.

Hon GC, Hawkins RD, Ren B. 2009. Predictive chromatin signatures in the mammalian genome. *Hum Mol Genet* **18:** R195–R201.

Hu G, Schones DE, Cui K, Ybarra R, Northrup D, Tang Q, Gattinoni L, Restifo NP, Huang S, Zhao K. 2011. Regulation of nucleosome landscape and transcription factor targeting at tissue-specific enhancers by BRG1. *Genome Res* **21:** 1650–1658.

Kidder BL, Palmer S, Knott JG. 2009. SWI/SNF-Brg1 regulates self-renewal and occupies core pluripotency-related genes in embryonic stem cells. *Stem Cells* **27:** 317–328.

Kim TH, Barrera LO, Ren B. 2007. ChIP-chip for genome-wide analysis of protein binding in mammalian cells. *Curr Protoc Mol Biol* **79:** 21.13.1–21.13.22.

Kim SI, Bresnick EH, Bultman SJ. 2009. BRG1 directly regulates nucleosome structure and chromatin looping of the α globin locus to activate transcription. *Nucleic Acids Res* **37:** 6019–6027.

Kothary R, Clapoff S, Brown A, Campbell R, Peterson A, Rossant J. 1988. A transgene containing lacZ inserted into the dystonia locus is expressed in neural tube. *Nature* **335:** 435–437.

Lee S, Lee JW, Lee SK. 2012. UTX, a histone H3-lysine 27 demethylase, acts as a critical switch to activate the cardiac developmental program. *Dev Cell* **22:** 25–37.

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

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

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.

Nord AS, Blow MJ, Attanasio C, Akiyama JA, Holt A, Hosseini R, Phouanenavong S, Plajzer-Frick I, Shoukry M, Afzal V, et al. 2013. Rapid and pervasive changes in genome-wide enhancer usage during mammalian development. *Cell* **155:** 1521–1531.

Pan G, Tian S, Nie J, Yang C, Ruotti V, Wei H, Jonsdottir GA, Stewart R, Thomson JA. 2007. Whole-genome analysis of histone H3 lysine 4 and lysine 27 methylation in human embryonic stem cells. *Cell Stem Cell* **1:** 299–312.

Pennacchio LA, Ahituv N, Moses AM, Prabhakar S, Nobrega MA, Shoukry M, Minovitsky S, Dubchak I, Holt A, Lewis KD, et al. 2006. In vivo enhancer analysis of human conserved non-coding sequences. *Nature* **444:** 499–502.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26:** 841–842.

Rada-Iglesias A, Bajpai R, Swigut T, Brugmann SA, Flynn RA, Wysocka J. 2011. A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470:** 279–283.

Robinson G, Parker M, Kranenburg TA, Lu C, Chen X, Ding L, Phoenix TN, Hedlund E, Wei L, Zhu X, et al. 2012. Novel mutations target distinct subgroups of medulloblastoma. *Nature* **488:** 43–48.

Saha A, Wittmeyer J, Cairns BR. 2006. Chromatin remodelling: the industrial revolution of DNA around histones. *Nat Rev Mol Cell Biol* **7:** 437–447.

Schneppenheim R, Fruhwald MC, Gesk S, Hasselblatt M, Jeibmann A, Kordes U, Kreuz M, Leuschner I, Martin Subero JI, Obser T, et al. 2010. Germline nonsense mutation and somatic inactivation of SMARCA4/BRG1 in a family with rhabdoid tumor predisposition syndrome. *Am J Hum Genet* **86:** 279–284.

Seo S, Richardson GA, Kroll KL. 2005. The SWI/SNF chromatin remodeling protein Brg1 is required for vertebrate neurogenesis and mediates transactivation of Ngn and NeuroD. *Development* **132:** 105–115.

Simone C, Forcales SV, Hill DA, Imbalzano AN, Latella L, Puri PL. 2004. p38 pathway targets SWI-SNF chromatin-remodeling complex to muscle-specific loci. *Nat Genet* **36:** 738–743.

Stankunas K, Hang CT, Tsun ZY, Chen H, Lee NV, Wu JI, Shang C, Bayle JH, Shou W, Iruela-Arispe ML, et al. 2008. Endocardial Brg1 represses *ADAMTS1* to maintain the microenvironment for myocardial morphogenesis. *Dev Cell* **14:** 298–311.

Takeuchi JK, Lou X, Alexander JM, Sugizaki H, Delgado-Olguin P, Holloway AK, Mori AD, Wylie JN, Munson C, Zhu Y et al. 2011. Chromatin remodelling complex dosage modulates transcription factor function in heart development. *Nat Commun* **2:** 187.

Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* **7:** 562–578.

Visel A, Minovitsky S, Dubchak I, Pennacchio LA. 2007. VISTA Enhancer Browser—a database of tissue-specific human enhancers. *Nucleic Acids Res* **35:** D88–D92.

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.

Wilson BG, Roberts CW. 2011. SWI/SNF nucleosome remodellers and cancer. *Natl Rev* **11:** 481–492.

Yu Y, Chen Y, Kim B, Wang H, Zhao C, He X, Liu L, Liu W, Wu LM, Mao M, et al. 2013. Olig2 targets chromatin remodelers to enhancers to initiate oligodendrocyte differentiation. *Cell* **152:** 248–261.

Zhan X, Shi X, Zhang Z, Chen Y, Wu JI. 2011. Dual role of Brg chromatin remodeling factor in Sonic hedgehog signaling during neural development. *Proc Natl Acad Sci* **108:** 12758–12763.

Zhang ZK, Davies KP, Allen J, Zhu L, Pestell RG, Zagzag D, Kalpana GV. 2002. Cell cycle arrest and repression of cyclin D1 transcription by INI1/hSNF5. *Mol Cell Biol* **22:** 5975–5988.

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, et al. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9:** R137.

Zhang M, Chen M, Kim JR, Zhou J, Jones RE, Tune JD, Kassab GS, Metzger D, Ahlfeld S, Conway SJ, et al. 2011. SWI/SNF complexes containing Brahma or Brahma-related gene 1 play distinct roles in smooth muscle development. *Mol Cell Biol* **31:** 2618–2631.

Zhao K, Wang W, Rando OJ, Xue Y, Swiderek K, Kuo A, Crabtree GR. 1998. Rapid and phosphoinositol-dependent binding of the SWI/SNF-like BAF complex to chromatin after T lymphocyte receptor signaling. *Cell* **95:** 625–636.

Zhao XD, Han X, Chew JL, Liu J, Chiu KP, Choo A, Orlov YL, Sung WK, Shahab A, Kuznetsov VA, et al. 2007. Whole-genome mapping of histone H3 Lys4 and 27 trimethylations reveals distinct genomic compartments in human embryonic stem cells. *Cell Stem Cell* **1:** 286–298.