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

Catia Attanasio [1#], Alex S. Nord [1#], 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], Len A. Pennacchio [1,2*]

[#] These authors have equally contributed to this work.

* Correspondence should be addressed to L.A.P. (*lapennacchio@lbl.gov*)

Supplemental Figures and Tables

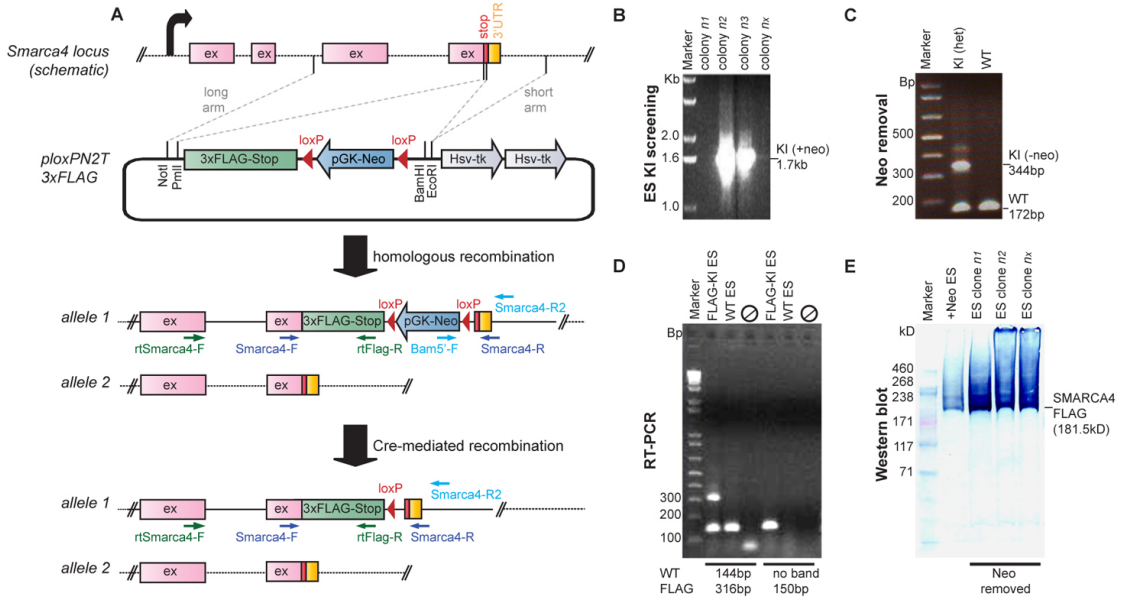
- **Supplemental Figure S1:** Generation of the *Smarca4*^{FLAG} knock-in mouse line.
- **Supplemental Figure S2:** SMARCA4-FLAG binding is consistent with endogenous SMARCA4 binding.
- **Supplemental Figure S3:** SMARCA4-interacting regions exhibit high levels of evolutionary constraint.
- **Supplemental Figure S4:** Results from *in vivo* transgenic enhancer assays by tissue.
- **Supplemental Figure S5:** Retrospective analysis of VISTA tested sequences that overlap SMARCA4-bound regions by tissue.
- **Supplemental Figure S6:** Functional term enrichment for SMARCA4-enriched TSSs by chromatin state.
- **Supplemental Figure S7**. Differential expression associated with SMARCA4-bound promoters and promoters with no SMARCA4 binding.
- **Supplemental Figure S8.** Relative co-enrichment of SMARCA4 and EP300 by chromatin state at forebrain and limb distal SMARCA4 elements.
- **Supplemental Figure S9.** Expression levels associated with SMARCA4-bound distal sites and distal sites with no SMARCA4 binding.
- **Supplemental Figure S10.** Differential chromatin state and associated expression between shared SMARCA4-enriched proximal sites found in forebrain and limb.
- **Supplemental Figure S11.** Differential chromatin state and associated expression between shared SMARCA4-enriched distal sites found in forebrain and limb.
- **Supplemental Figure S12.** SMARCA4-predicted functional classes of elements in limb reflect their regulatory activity *in vivo*.
- **Supplemental Table S1:** Forebrain-specific SMARCA4-enriched distal regions.
- **Supplemental Table S2:** Limb-specific SMARCA4-enriched regions.
- **Supplemental Table S3:** Heart-specific SMARCA4-enriched regions.
- **Supplemental Table S4:** EXCEL file listing all elements tested *in vivo*.
- **Supplemental Table S5:** Counts for forebrain and limb TSS regions by chromatin state.

- **Supplemental Table S6:** Counts for forebrain and limb distal regions by chromatin state.
- **Supplemental Table S7:** Differential chromatin state and associated functional enrichment between shared SMARCA4-enriched distal sites found in forebrain and limb.
- **Supplemental Table S8:** Retrospective analysis of tested enhancers.
- **Supplemental Table S9**: Primers used to generate and screen *Smarca4*^{FLAG} ESCs.

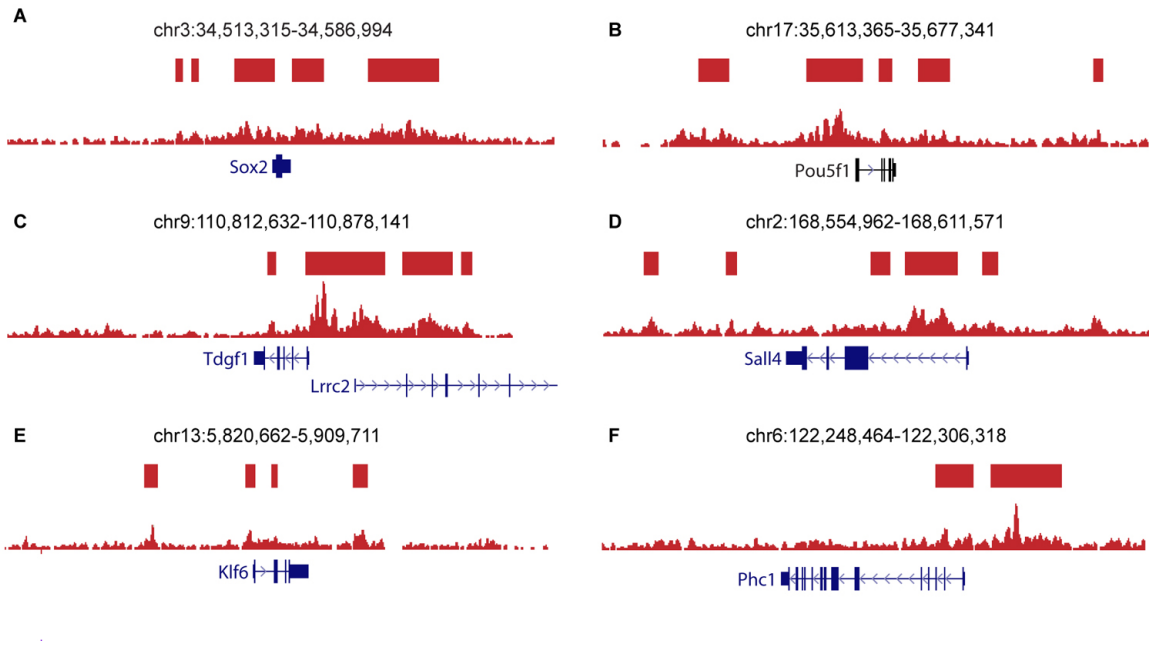
Additional Supplemental Files/Material

1. **Supplemental Bed Files S1-6**: (all coordinates mm9)

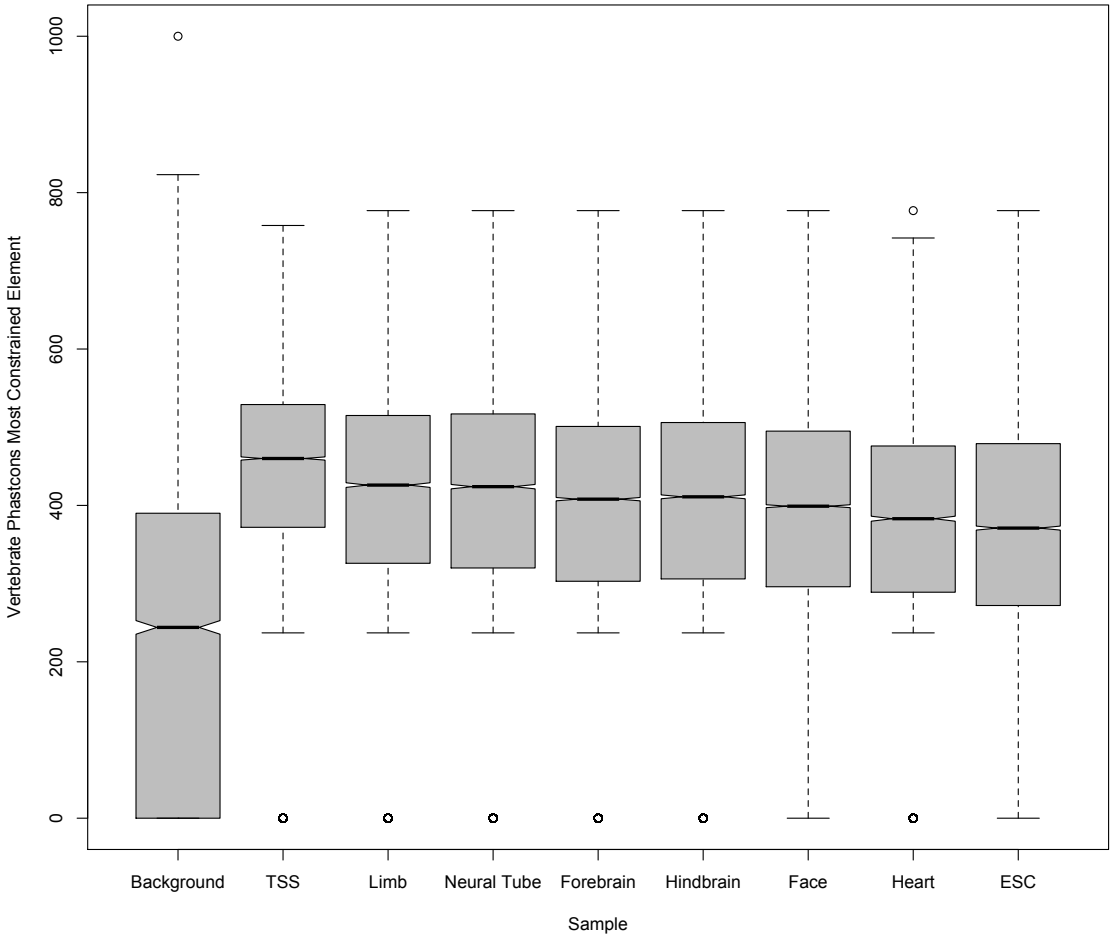
    -SMARCA4_Distal.bed: Distal SMARCA4 elements identified across six tissues plus ESCs annotated with SMARCA4 tissue enrichment calls.
    -SMARCA4_Proximal.bed: Proximal SMARCA4 elements identified across six tissues plus ESCs annotated with SMARCA4 tissue enrichment calls.
    -SMARCA4_Forebrain_State_Distal.bed: Distal SMARCA4 elements in forebrain annotated with histone mark enrichment calls (H3K4me1, H3K27ac, H3K27me3).
    -SMARCA4_Forebrain_State_Proximal.bed: Proximal SMARCA4 elements in forebrain annotated with histone mark enrichment calls (H3K4me3, H3K27me3).
    -SMARCA4_Limb_State_Distal.bed: Distal SMARCA4 elements in limb annotated with histone mark enrichment calls (H3K4me1, H3K27ac, H3K27me3).
    -SMARCA4_Limb_State_Proximal.bed: Proximal SMARCA4 elements in limb annotated with histone mark enrichment calls (H3K4me3, H3K27me3).



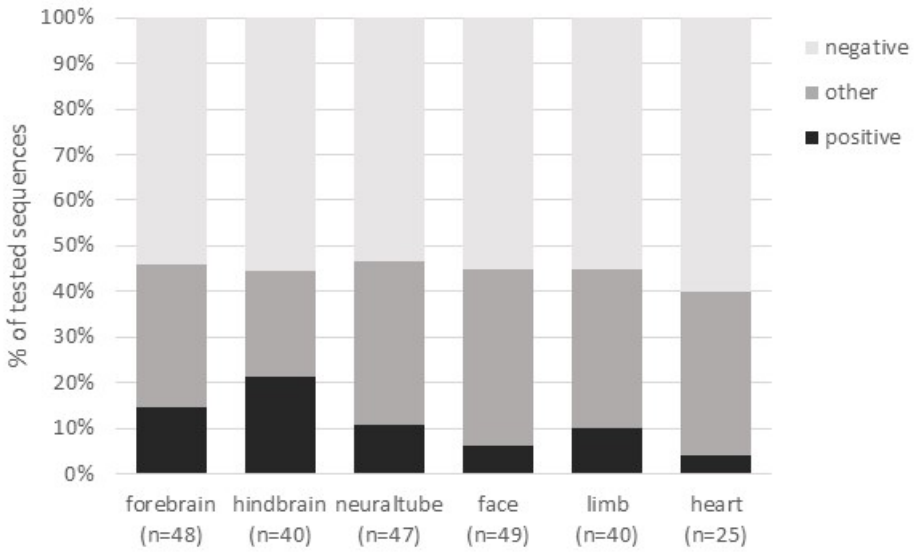
Supplemental Figure S1. Generation of the *Smarca4*^{FLAG} knock-in mouse line. A. Targeting strategy (schematic, not to scale). The endogenous *Smarca4* locus is targeted in ESCs with an 'homology arms' targeting vector (ploxPN2T3xFLAG) containing a Neomycin resistance cassette (+ Neo locus) flanked by loxP sites (red triangles). After recombination of the FLAG cassette into the carboxyl terminus of the endogenous SMARCA4 locus, a Cre-recombinase-expressing plasmid (kind gift from Dr. Timothy Ley of the Embryonic Stem Cell Core of the Siteman Cancer Center, Washington University Medical School) was used to remove the Neomycin cassette from the genome (- Neo locus). **B**. PCR on ESC colonies to identify FLAG recombined cells. **C**. Neomycin removal in ESCs. KI (het), *Smarca4* FLAG/+ embryonic stem cells. WT, non-targeted ESCs. **D**. RNA expression analysis of *Smarca4*-FLAG by reverse-transcribed PCR demonstrates mRNA expression of the knock-in allele. **E**. Western blot confirming the expression of the SMARCA4-FLAG protein. A list of all primers used is provided in **Supplemental Table S9**.



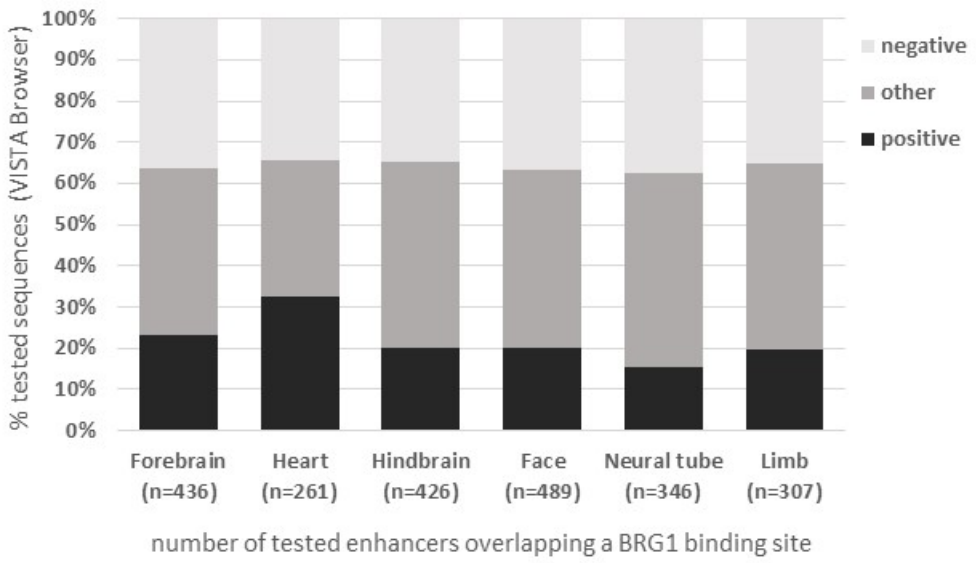
Supplemental Figure S2. SMARCA4-FLAG binding is consistent with endogenous SMARCA4 binding. A-F. Anecdotal examples of SMARCA4-enriched regions in mouse ESCs at previously reported SMARCA4-enriched regions (Kidder et al. 2009). Coordinates are provided in mm9.



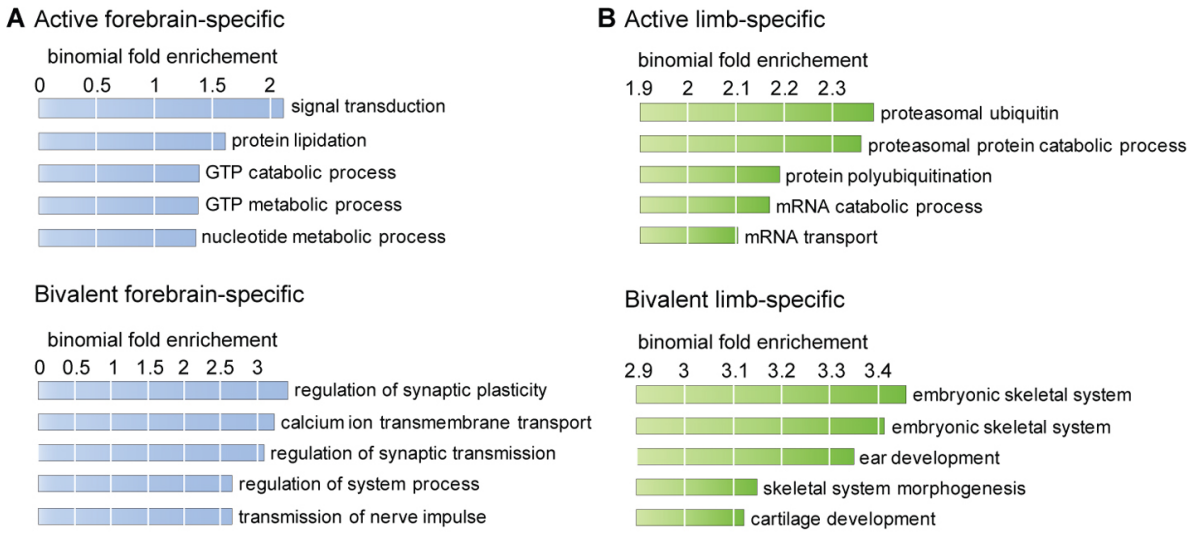
Supplemental Figure S3. SMARCA4-interacting regions exhibit high levels of evolutionary constraint. Background distribution represents 5000 size-matched random non-TSS genomic regions. TSSs are SMARCA4-bound transcription start sites. Data for each tissue represents all distal SMARCA4-bound regions in tissue at E11.5. Constraint represents the distribution of the most conserved overlapping phastCons Vertebrate Element.



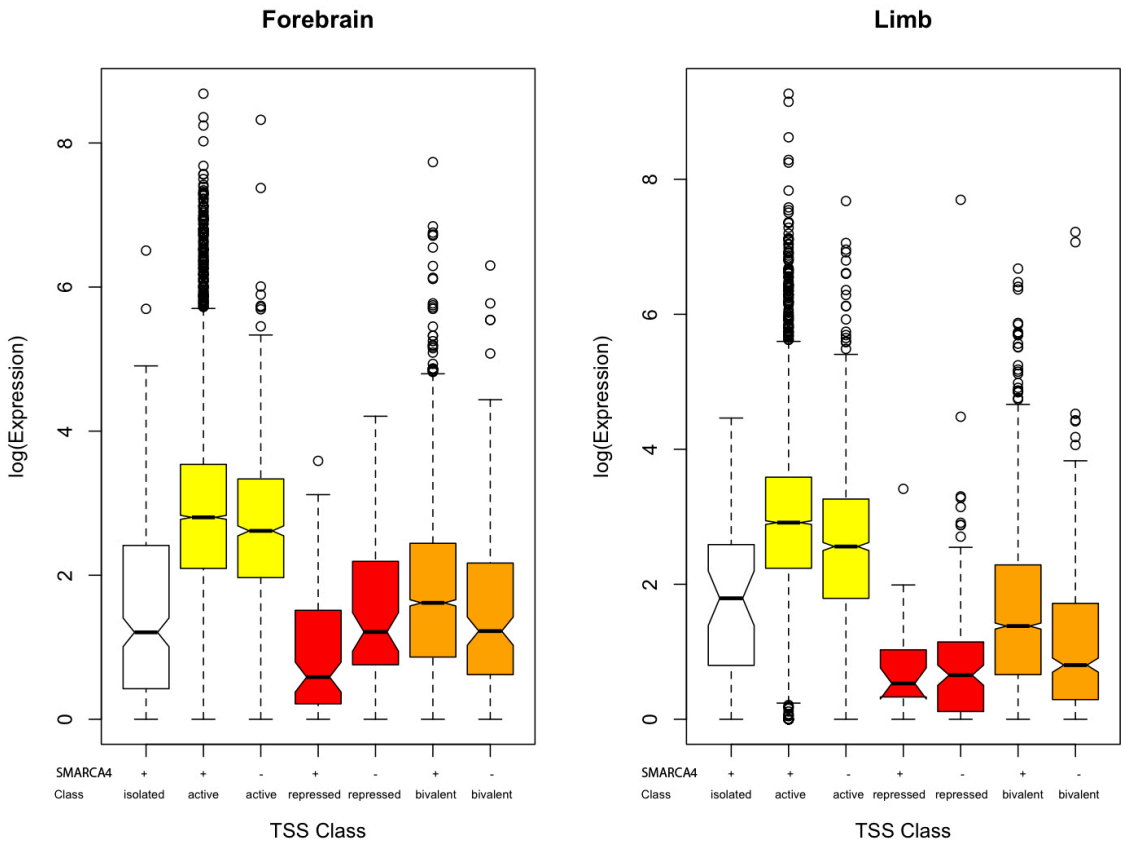
Supplemental Figure S4. Results from *in vivo* transgenic enhancer assays by tissue. In each tissue, the proportion of SMARCA4 enhancers approximate 45% of the tested sequences. The tissue specificity of these enhancers is, however, lower than previously observed for other enhancers marks such as EP300 (5-20% of the positive enhancers depending on the analyzed tissue). Positive, sequence with enhancer activity in the predicted tissue. Other, sequence with enhancer activity in a different tissue than predicted by SMARCA4 binding. Negative, sequence with no *in vivo* enhancer activity at E11.5.



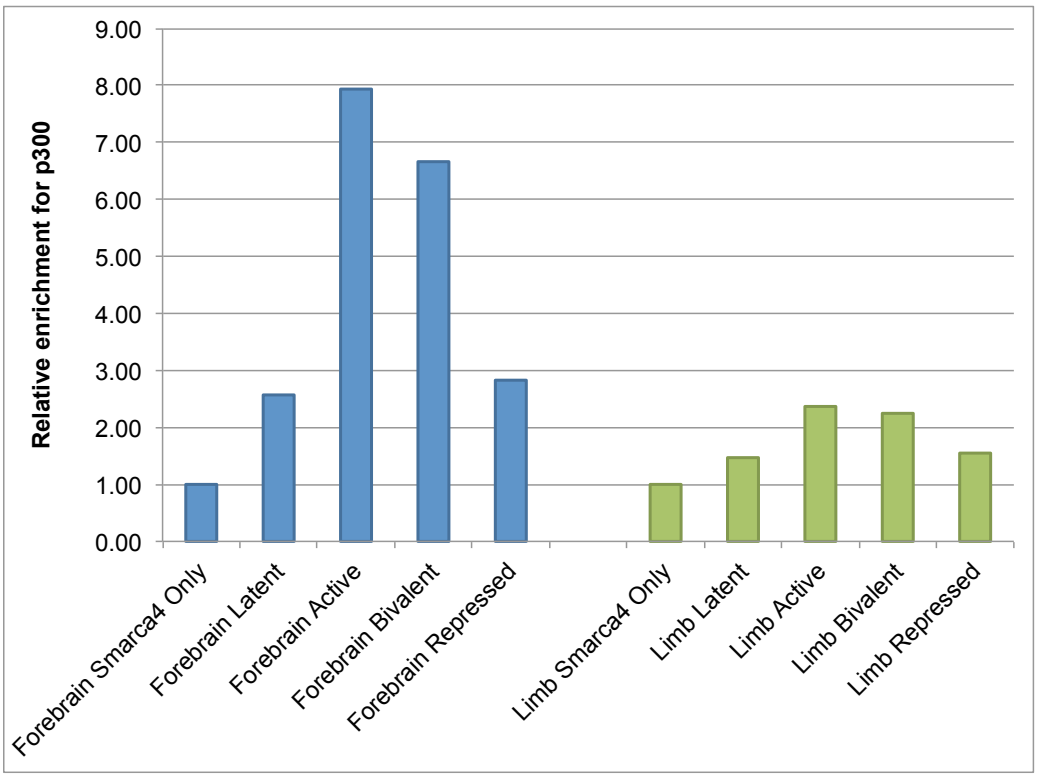
Supplemental Figure S5. Retrospective analysis of VISTA tested sequences that overlap SMARCA4-bound regions by tissue. Positive, sequence with enhancer activity in the predicted tissue. Other, sequence with enhancer activity in a different tissue than predicted by SMARCA4 binding. Negative, sequence with no *in vivo* enhancer activity at E11.5.



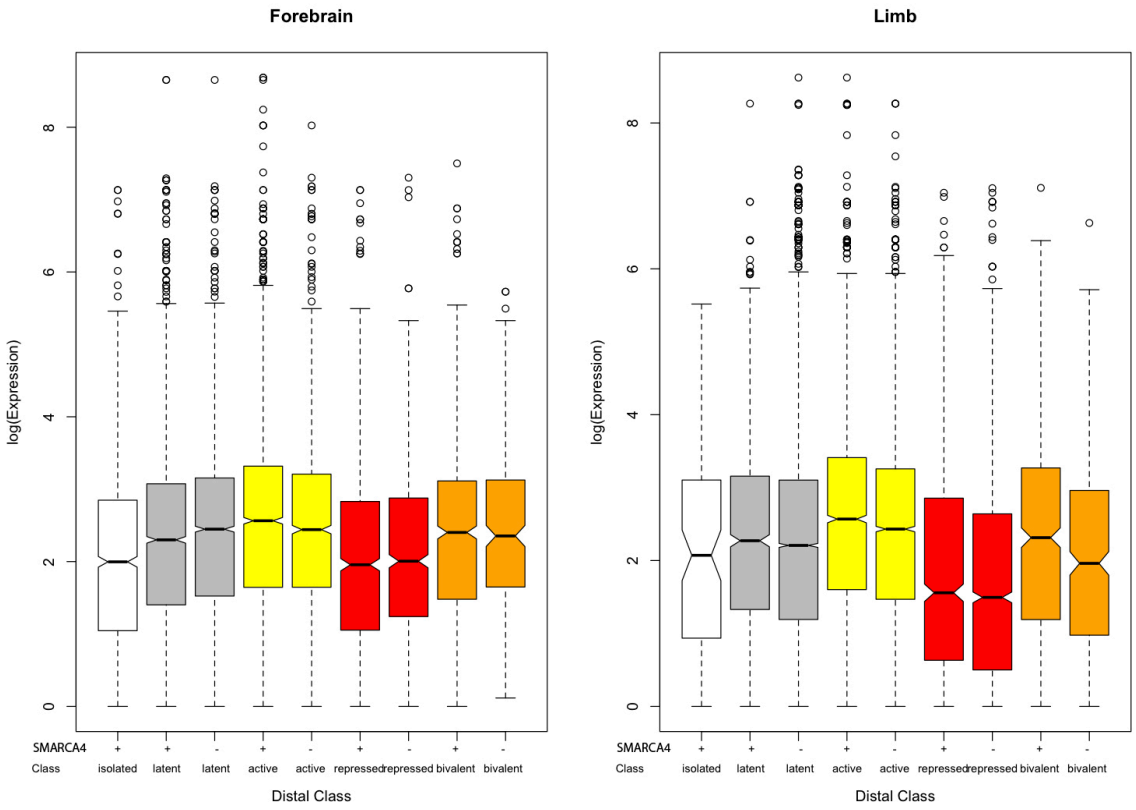
Supplemental Figure S6. Functional term enrichment for SMARCA4-enriched TSSs by chromatin state. A,B. Differential enrichment for functional annotation terms associated with proximal SMARCA4 binding categorized by histone signature. Shown are the top five enriched 'Biological process terms' (McLean et al.) for forebrain- (**A**) and limb-enriched (**B**) SMARCA4 proximal regions.



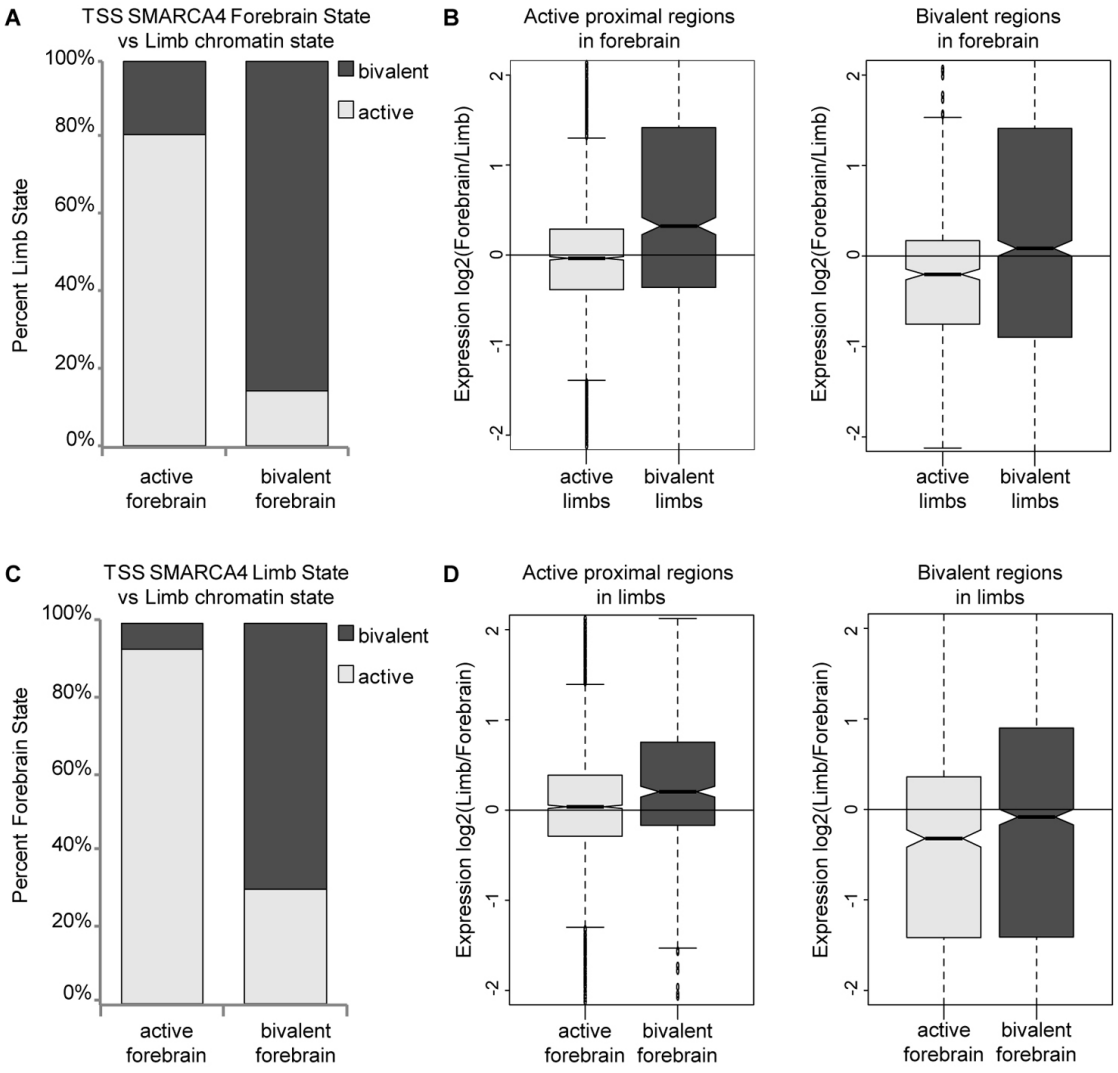
Supplemental Figure S7. Differential expression associated with SMARCA4-bound promoters and promoters with no SMARCA4 binding. Expression levels of genes that had characteristic chromatin states were separated based on presence or absence of SMARCA4 binding. SMARCA4-bound active and bivalent promoters exhibit increased expression associated with SMARCA4 binding, while repressed promoters (marked only by H3K27me3) do not show a consistent effect.



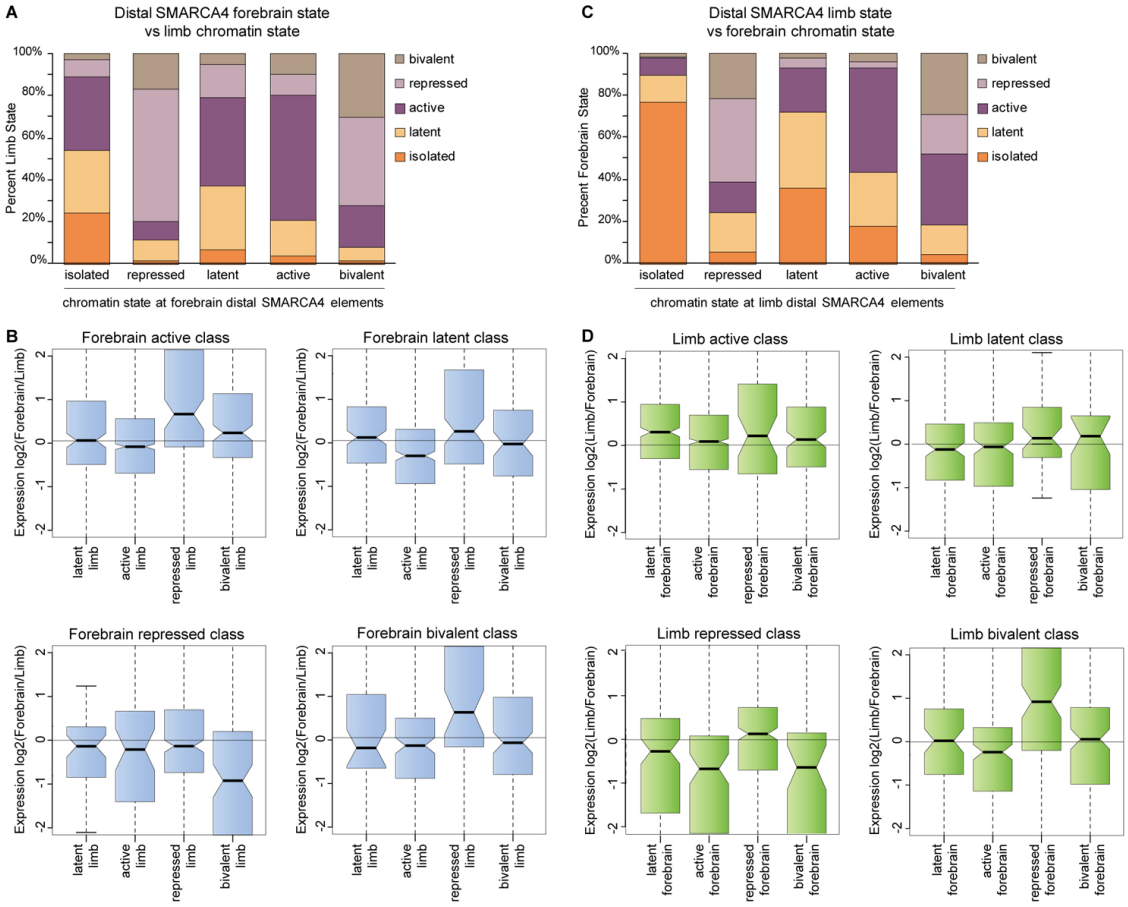
Supplemental Figure S8. Relative co-enrichment of SMARCA4 and EP300 by chromatin state at forebrain and limb distal SMARCA4 elements. EP300 enrichment from forebrain and limb ChIP-seq datasets was analyzed for co-enrichment at distal SMARCA4 binding sites. SMARCA4 elements were separated by chromatin-state based classification. Relative EP300 enrichment is defined as the proportion enriched in a given class versus the proportion of SMARCA4-only elements enriched in the tissue.



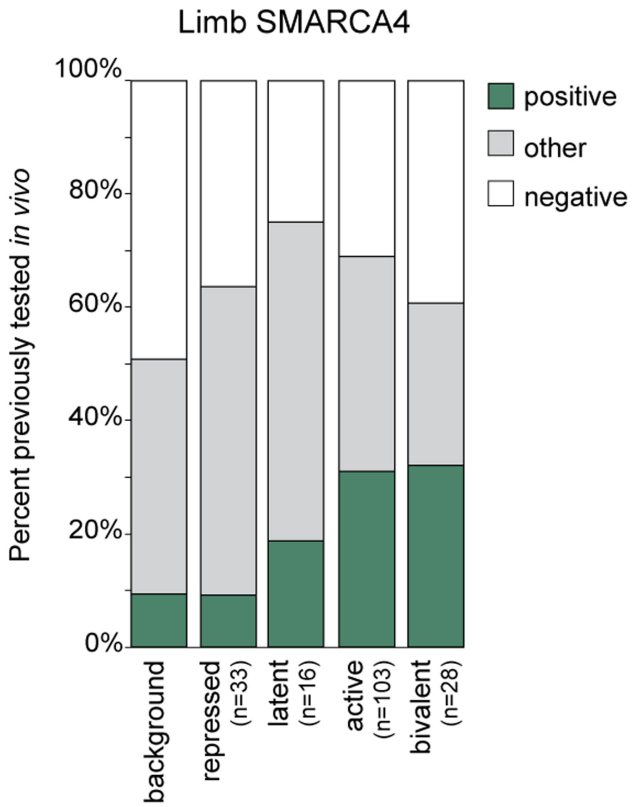
Supplemental Figure S9. Expression levels associated with SMARCA4-bound distal sites and distal sites with no SMARCA4 binding. Distal sites with characteristic chromatin states were separated into SMARCA4-enriched and not enriched and expression was compared within groups. Subtle but significant differences are present, but are weaker than across group differences with the exception of limb bivalent elements.



Supplemental Figure S10. Differential chromatin state and associated expression between shared SMARCA4-enriched proximal sites found in forebrain and limb. Shared proximal SMARCA4-enriched regions were classified for state in forebrain and limb using histone marks (H3K4me3 and H3K27me3). **A.** Forebrain states subdivided by proportion of corresponding limb state (e.g. the proportion of forebrain active promoters that are as either limb active or limb bivalent). **B.** Relative forebrain expression (log2(forebrain/limb)) for differential state combinations (e.g. top left panel shows that promoters classified as a combined state of forebrain active and limb bivalent exhibit significantly increased relative forebrain expression). **C.** Limb states subdivided by proportion forebrain state. **D.** Relative limb expression (log2(limb/forebrain)) for differential state combinations.



Supplemental Figure S11. Differential chromatin state and associated expression between shared SMARCA4-enriched distal sites found in forebrain and limb. Shared distal SMARCA4-enriched regions were classified for state in forebrain and limb using histone marks (H3K4me1, H3K27ac, and H3K27me3). **A,C.** Forebrain (**A**) and limb (**C**) states subdivided by proportion of corresponding limb state and forebrain state respectively (e.g. the proportion of forebrain-active elements that are classified as each of the five states in limb). **B.** Relative forebrain expression (log2(forebrain/limb)) for differential state combinations (e.g. top left panel shows that elements classified as a combined state of forebrain active and limb repressed exhibit significantly increased relative forebrain expression). **D.** Relative limb expression (log2(limb/forebrain)) for differential state combinations.



Supplemental Figure S12. SMARCA4-predicted functional classes of elements in limb reflect their regulatory activity *in vivo*. Limb SMARCA4-enriched regions were intersected with 1,747 mouse sequences previously tested in mouse transgenic assays. As shown in the bar graph, SMARCA4 marks enhancers and the histone signature at SMARCA4-enriched loci predicts tissue-specific activity. Positive, reproducible enhancer activity in limb; other, reproducible enhancer activity in a tissue other than limb; negative, no detectable enhancer activity *in vivo*.

Supplemental Tables

Rank	Phenotype Term	Binomial P-value	Binomial Fold Enrichment
Biological processes			
2	sensory organ development	4.2e-62	2.0
7	central nervous system neuron differentiation	9.9e-43	2.4
12	forebrain neuron differentiation	2.1e-36	4.5
14	spinal cord association neuron differentiation	1.2e-35	7.1
15	forebrain generation of neurons	3.2e-35	3.8
19	dorsal spinal cord development	1.1e-30	5.0
Mouse phenotypes			
1	abnormal basisphenoid bone morphology	1.1e-36	3.1
2	abnormal basicranium morphology	4.1e-36	2.8
3	abnormal forebrain development	1.5e-32	2.1
4	abnormal neuronal precursor proliferation	8.5e-28	2.8
5	abnormal sphenoid bone morphology	9.7e-28	2.4

Supplemental Table S1. Forebrain-specific SMARCA4-enriched distal regions (5,328 regions). Six of the top twenty ranked biological processes are linked to forebrain development (only relevant terms shown). Additional terms (not shown in table) are associated to stem cell maintenance (2/20), cell fate/tissue differentiation (7/20) and eye development (5/20) processes.

Rank	Phenotype Term	Binomial P-value	Binomial Fold Enrichment
Biological processes			
1	skeletal system development	7.2e-35	2.8
5	embryonic morphogenesis	7.1e-26	2.2
6	limb morphogenesis	3.4e-22	2.9
7	limb development	3.7e-22	2.8
9	embryonic organ morphogenesis	3.9e-20	2.6
10	embryonic organ development	7.8e-20	2.3
Mouse phenotypes			
1	abnormal skeleton development	1.9e-29	2.2
3	abnormal limb morphology	3.3e-24	2.1
6	abnormal skeleton extremities morphology	3.8e-23	2.2
7	abnormal appendicular skeleton morphology	7.0e-23	2.2
9	abnormal long bone morphology	6.2e-21	2.2

Supplemental Table S2. Limb-specific SMARCA4-enriched regions (1,488 regions). Six of the 10 top ranked biological processes are associated to limb development (only relevant terms shown). Other terms (not shown in table) involve negative gene expression regulation (3/10) and epithelium development (1/10). Among the top 10 ranked mouse phenotypes, 5 are relevant to limb development, 3 are involved into abnormal craniofacial morphology and 1 is linked to thoracic morphology.

Rank	Phenotype Term	Binomial P-value	Binomial Fold Enrichment
Biological processes			
1	actin filament-based process	1.2e-47	2.9
2	actin cytoskeleton organization	8.1e-47	2.9
3	cytoskeleton organization	2.8e-42	2.2
5	actin filament organization	3.5e-27	3.7
6	cardiac chamber development	1.4e-21	2.3
7	muscle cell differentiation	1.7e-21	2.1
8	cardiac ventricle development	2.0e-21	2.4
9	cardiac muscle tissue development	6.6e-20	2.3
Mouse phenotypes			
4	abnormal pericardium morphology	1.4e-39	2.1
5	pericardial effusion	1.8e-38	2.1
6	abnormal heart layer morphology	9.9e-37	2.5
7	abnormal myocardium layer morpholgoy	8.2e-35	2.0
8	abnormal vitelline vasculature morphology	5.7e-33	3.3

Supplemental Table S3. Heart-specific SMARCA4-enriched regions (3,921 regions). Nine of the ten top ranked biological processes are linked to heart development. The other associated term (not shown in table) is related to small GTPase-mediated signal transduction. Among the top 8 ranked mouse phenotypes, 5 are relevant to heart development and 3 (not shown in table) are linked to general abnormal embryonic development.

Class	Forebrain Count	Forebrain %	Limb Count	Limb %
Active	6706	66.7%	5936	59.3%
Bivalent	3345	33.3%	4080	40.7%
Total	10051	100.0%	10016	100.0%

Supplemental Table S5. Counts for forebrain and limb TSS regions by chromatin state.

Class	Forebrain Count	Forebrain %	Limb Count	Limb %
Isolated	2136	18.6%	139	2.1%
Latent	3476	30.3%	987	14.9%
Active	3779	33.0%	3403	51.3%
Repressed	1158	10.1%	1206	18.2%
Bivalent	919	8.0%	905	13.6%
Total	11468	100.0%	6640	100.0%

Supplemental Table S6. Counts for forebrain and limb distal regions by chromatin state.

Forebrain Bivalent, Limb Repressed

Mouse Phenotype	Binom Rank	Binom Raw P-Value	Binom FDR Q-Val	Binom Fold Enrichment
abnormal forebrain morphology	1	4.4e-14	3.2e-10	2.6
abnormal forebrain development	4	5.2e-13	9.6e-10	4.7
abnormal dentate gyrus morphology	5	6.6e-13	9.6e-10	7.0
abnormal neuron morphology	6	6.8e-13	8.3e-10	2.1
abnormal telencephalon morphology	7	3.3e-12	3.4e-09	2.7
abnormal nervous system development	8	1.5e-11	1.4e-08	2.2
absent dentate gyrus	9	3.4e-11	2.8e-08	13.4
abnormal neuron differentiation	10	5.5e-11	4.0e-08	3.3
abnormal olfactory lobe morphology	11	8.9e-11	5.9e-08	4.4
abnormal brain morphology	13	2.0e-10	1.1e-07	2.0

Limb Bivalent, Forebrain Repressed

Mouse Phenotype	Binom Rank	Binom Raw P-Value	Binom FDR Q-Val	Binom Fold Enrichment
abnormal autopod morphology	1	4.18E-21	3.06E-17	6.37
abnormal limbs/digits/tail morphology	2	5.71E-20	2.09E-16	3.80
lethality during fetal growth through weaning	3	3.56E-19	8.67E-16	2.36
small limb buds	4	6.97E-19	1.27E-15	43.11
abnormal digit morphology	5	5.40E-18	7.89E-15	6.40
Oligodactyly	6	8.50E-18	1.04E-14	19.15
abnormal limb morphology	7	1.27E-17	1.33E-14	4.30
abnormal craniofacial morphology	8	1.99E-17	1.82E-14	3.24
abnormal appendicular skeleton morphology	9	7.82E-17	6.35E-14	4.53
abnormal skeleton extremities morphology	10	1.27E-16	9.26E-14	4.59

Supplemental Table S7. Differential chromatin state and associated functional enrichment between shared SMARCA4 enriched distal sites found in forebrain and limb.

Class	Tested	Positive	Tissue
Forebrain Active	102	73	35
Forebrain Bivalent	33	20	8
Forebrain Latent	45	31	4
Forebrain Repressed	40	21	6
Limb Active	103	71	32
Limb Bivalent	28	17	9
Limb Latent	16	12	3
Limb Repressed	33	21	3

Background Positive (n=887)
Background Limb (n=164)
Background Forebrain (n=226)

Supplemental Table S8. Retrospective analysis of tested enhancers

	Primer name	5'-3' Sequence	Product Size (bp)
Short homology	Smarca4S-F	TGAACCAGACATTCCTGAGTCCTGACC	1,541bp
	Smarca4S-R	GGTTTGTTGTACAGTGTGTCTCTGTGGTA	
Long homology	Smarca4L-F	ATGGTCCGCATCCAAACTGCCTGAACA	6,239bp
	Smarca4L-R2	GTCTTCCTCACTGCCACTTCCTGAG	
Knock-in event	Smarca4-R2	CTTTCCCTCCTGGGAAGTCTCCTGT	1,671bp
	Bam5'-F	TTGGCTGGACGTAAACTCCTCTTCAG	
Neo removal screening	Smarca4-F	ACACCATGGACAAGGGAGAGTGCCTA	WT: 172 bp
	Smarca4-R	CCATCACTGCTAAGGGCTACTCCATCT	KI (+Neo): 2,294
RT-PCR (RNA expression)	rtSmarca4-F	GTTGTGAGTGACGATGACAGTGAGGAG	WT: 144 bp
	rtSmarca4-R	Same as Smarca4-R	KI: 316 bp
	rtSmarca4-F	Same as above	WT: no product
	rtFlag-R	CTTGTCATCGTCATCCTTGTAGTCGATG	KI: 150 bp

Supplemental Table S9. Primers used to generate and screen *Smarca4*^{FLAG} ESCs.