Supporting Information

Lee et al. 10.1073/pnas.1308758111

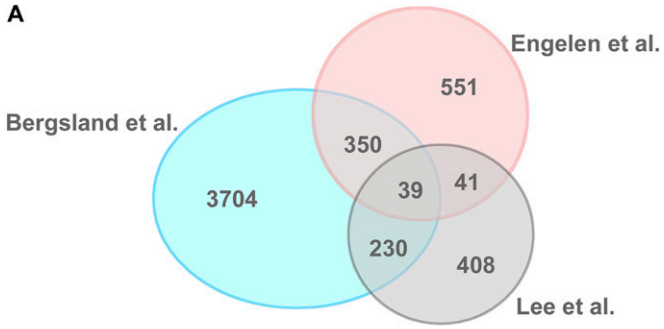
SI Materials and Methods

Array Scan and Data Extraction from ChIP-on-Chip Assay. Slides were scanned using an Agilent DNA microarray scanner. Photomultiplier tube settings were set manually to normalize bulk signal in the Cy3 and Cy5 channels. For efficient batch processing of scans, we used GenePix 4000B software. Scans were automatically aligned and manually examined for abnormal features. Intensity data were then extracted from each batch. Feature extraction is the process by which information from the probe feature is extracted from microarray scan data, allowing investigators to measure gene expression in the experiments. We used DNA Analytics software for the feature extraction. We selected statistically significant peaks per probe, detected continuously in more than two probes and using P value cutoff of <0.01, and then combined the peaks into a region, which thus represented the Sox2-binding region. The P value is a probability of binding events that will occur on each probe, and the Rosetta error model was used to establish appropriate relative weights, which corrects for the uncertainties in individual probe measurements. Consequently, we identified total 811 Sox2-binding regions. The Sox2 ChIP-on-chip data have been made publicly available in the Gene Expression Omnibus repository at the National Center for Biotechnology Information (accession no. GSE34395).

Motif Analysis of Sox2. Sequence motif representing the binding specificity of Sox2 was estimated from the ChIP-on-chip data using the Regulatory Sequence Analysis Tools (RSAT: http://rsat.bigre.ulb.ac.be/rsat/). We conducted RSAT oligo-analysis on the 811 Sox2-binding regions with the checking "purge sequences" option, setting "oligomer length" to 8 and keeping other parameter values unchanged. Then, we used predefined background frequencies as a background probabilistic model to estimate the expected frequency of each oligonucleotide. A variety of statistical methods have been proposed to compare observed and expected frequencies such as observed/expected ratio, z-score, binomial, Poisson, and compound Poisson. The probability of observing successes by chance is given by the inverse cumulative binomial distribution. This probability, also called nominal P value, indicates the risk for a given oligonucleotide to be considered as significant when in fact it is not. Because the same statistical test is applied to all possible oligonucleotides, this program performs a correction for multi-testing by multiplying the nominal P value by the number of oligonucleotides tested to obtain an expected value (E-value). The E-value represents the number of false positives expected by chance at a given level of P value. The significance score is a minus log transformation of this E-value. To estimate the statistical significance of the appearance of 5-base-long core Sox-binding elements (5′-$^{A}/_{T}$TTGT-3′ or its complement sequence), we again used the RSAT program to first estimate the expected occurrence for each of the core Sox-binding elements from a randomly selected human genome sequence of equal length to the pooled Sox2-binding regions from 811 targets. Subsequently, Fisher's combined probability test was used to estimate the significance of the combined observed number of the core elements.

Reporter Assay. The promoter regions of human *Sox6* containing the potential five SOX-binding sites (−655 to +43) or none of the binding sites (−358 to +43) were obtained by PCR amplification using a BAC clone (RP11-635A4) purchased from the Children's Hospital Oakland Research Institute as the template and subsequently cloned into pGL3-basic vector (Promega). Mutagenesis of the potential binding site was performed using the QuikChange II Site-Directed Mutagenesis Kit (Stratagene). Oligonucleotide primer pairs used to generate mutant plasmids are as follows: for pGL3-Sox6-MT1, 5′-GAGGCCCAGTCTTAGAATAAGACTAATGAGACCAT-3′ and 5′-ATGGTCTCAT TAGTCTTATTCTAAGACTGGGCCTC-3′; for pGL3-Sox6-MT2, 5′-GCTAAACAGAGAACTTAAGTCTTCTGGTCTTGTCC-3′ and 5′-GGACAAGACCAGAAGACTTAAGTTCTCTGTTTAGC-3′; for pGL3-Sox6-MT3, 5′-GAATCAACAACCTAGTCTTGTGGACTTG-3′ and 5′-CAAGTCCACAAGACTAGGTTGTTGATTC-3′; for pGL3-Sox6-MT4, 5′-CCGGCACTCTGTAGAAGTCTTTTCTTTTGAAACTCTTG-3′ and 5′-CAAGAGTTTCAAAAGAAAAGACTTCTACAGAGTGCCGG-3′; and for pGL3-Sox6-MT5, 5′-CATAAAAGGCGAGTCTCTGGATAAAAAGTCC-3′ and 5′-GGACTTTTTATCCAGAGACTCGCCTTTTATG-3′. For the reporter assay, 3×10^4 HEK293T cells were plated 24 h before transfection. Each of the pGL3-*Sox6* reporter plasmids (200 ng) was cotransfected with pCMV β-gal plasmid (50 ng) and indicated amounts of human *Sox2* expression plasmids into HEK293T cells using Effectene transfection reagent (Qiagen) following the manufacturer's protocols. Human *Sox2* expression plasmid contains the entire ORF amplified from a BAC clone (RP11-947C12), which was subsequently cloned into pcDNA3.1A. After 24 h, cells were harvested, and luciferase activity was measured and normalized using the Luciferase Assay System (Promega). Details are available upon request.

RNA in Situ Hybridization for Mouse Embryos. The probe for mouse *Sox2* corresponds to the nucleotide 1282–2029 region of a mouse *Sox2* cDNA (NM_011443). The template for mouse *Sox6* probe was obtained by PCR amplification from a cDNA plasmid obtained from Open Biosystems (clone ID: 30094413) and corresponds to the nucleotide 1–1378 region.



B

ABTB2	HMGB1	RFX4
ADAR	HNRNPL	RNF144A
ARRDC3	INSM1	SALL1
BTG1	ITGB8	SEMA6D
CAPRIN1	KIAA0182	SOX2
CUEDC1	MAF	SOX5
CYR61	MARCKS	SOX6
DDR1	MTF2	SP3
DUSP6	OLIG1	SPRED2
EIF4G2	PIPOX	SYNE2
ETV1	PTCH1	TUBA1A
GADD45G	PTPRZ1	ZNF532
GLI3	RARB	ZNF608

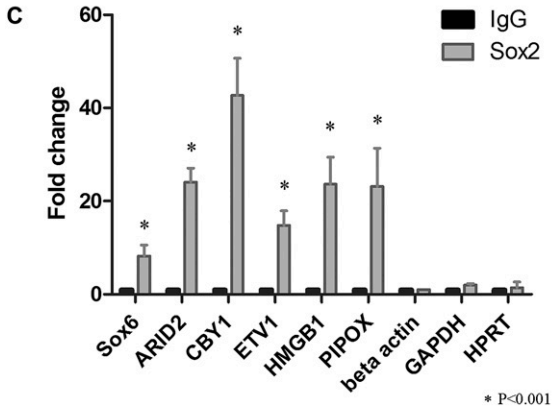


Fig. S1. (*A*) A Venn diagram showing the numbers of common and exclusive candidate direct target genes of Sox2 from three studies (*Discussion*). (*B*) List of candidate direct target genes of Sox2 common to all three studies. (*C*) Interaction of Sox2 with potential binding sites of the candidate targets was analyzed by ChIP assay using anti-Sox2 antibody and control IgG antibody. Genomic sites near *GAPDH*, *hypoxanthine phosphoribosyltransferase 1* (*HPRT*), and *β-actin* loci were used as controls, and normalization was carried out using *β-actin*. Data are the average of three independent experiments, and error bars represent SDs. (*$P < 0.001$.) Oligonucleotide primers used for PCR are listed in Table S1.

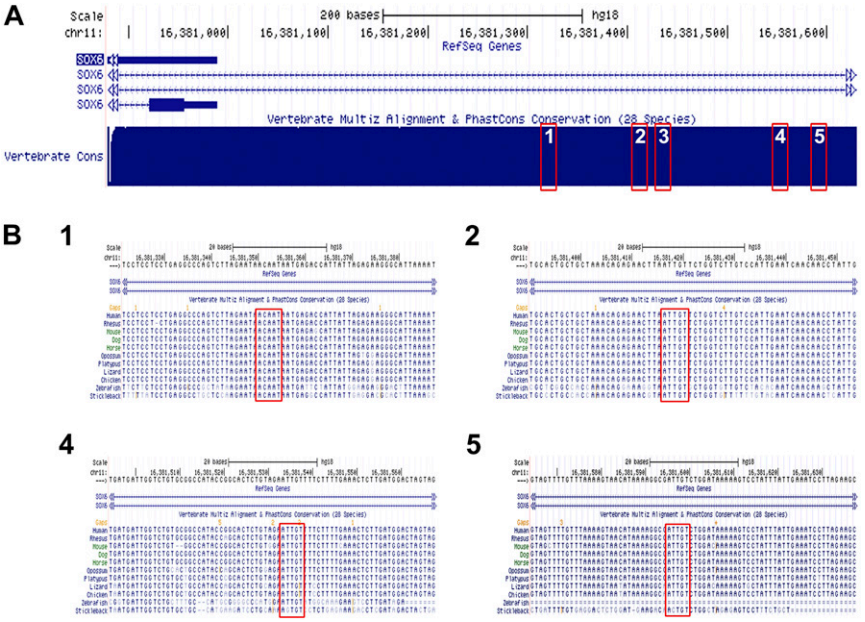


Fig. S2. Conserved SOX-binding sites near *Sox6* promoter. (*A*) A portion of the *Sox6* locus on human chromosome 11 (adapted from the University of California at Santa Cruz Genome Browser). The locations of five potential Sox2-binding sites near the ChIP signal are indicated by red boxes. (*B*) The nucleotide sequences surrounding each of the binding sites are shown in detail. The number next to each panel matches the number above the red box in *A*. Within each of the red boxes, the conserved core sequence, 5′-ATTGT-3′, or the complementary sequence, 5′-ACAAT-3′, is found. The third binding site is shown in the Fig. 1*E*.

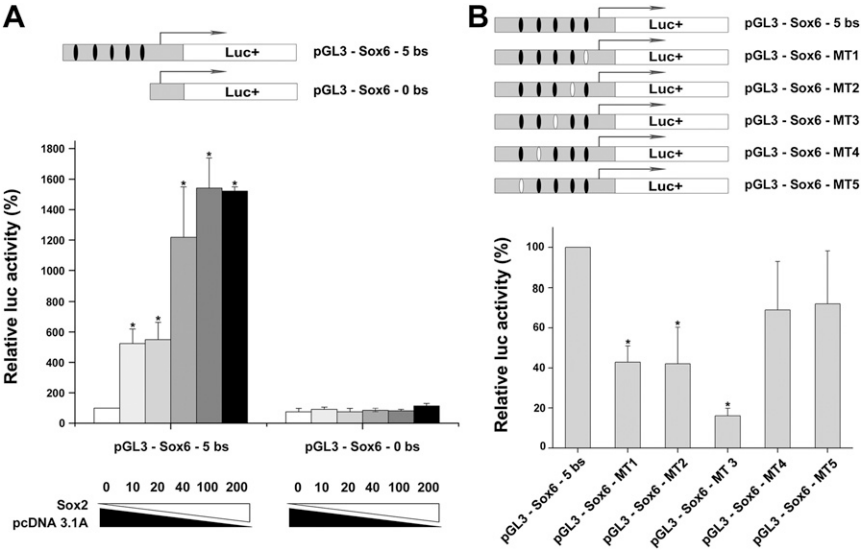


Fig. S3. Reporter assay. (*A*) (*Upper*) The reporter constructs. pGL3-*Sox6*-5bs contains the wild-type *Sox6* promoter/enhancer region with all of the five potential Sox2-binding sites intact whereas pGL3-*Sox6*-0bs contains none of them. (*Lower*) Graph shows normalized relative levels of luciferase activity with the value from pGL3-*Sox6*-5bs in the absence of Sox2 as 100%. The combined amounts of Sox2 and pcDNA vector plasmid were equal in all cases. Data are the average of three independent experiments, and error bars represent SDs. Significant Sox2 dose-dependent induction (*$P < 0.05$ by Student *t* test) was seen only in the case of pGL3-*Sox6*-5bs. (*B*) Reporters with each of the five Sox2-binding sites mutated were examined. The white oval represents the mutated binding site. The numeral designations of mutants match those used in Fig. 1*E* and Fig. S2. (*Lower*) Graph shows normalized relative levels of luciferase activity with the value from pGL3-*Sox6*-5bs with 40 ng of Sox2 expression plasmid as 100%. These data are representative of four independent experiments (*$P < 0.05$ by Student *t* test). Binding sites 1, 2, and 3 appear to mediate the activity of Sox2 significantly.

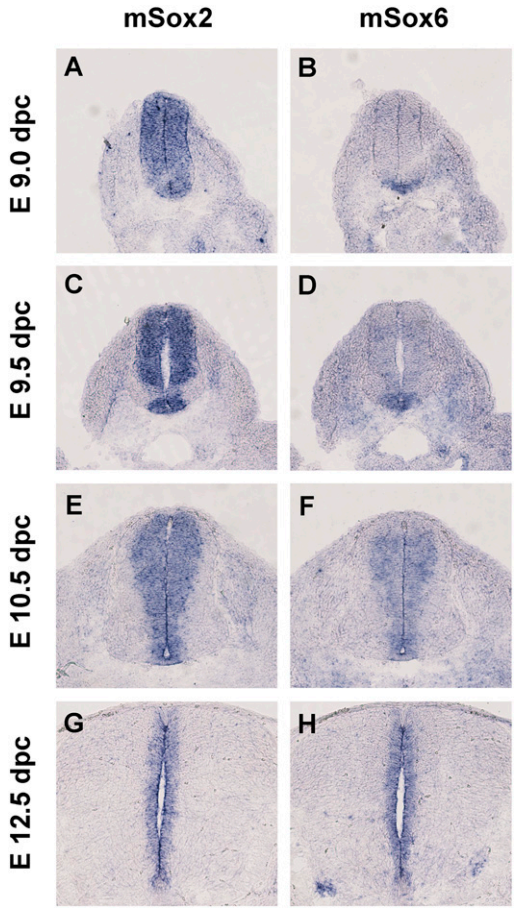


Fig. S4. Expression of *Sox2* and *Sox6* in the developing mouse neural tube. Neural tubes at the indicated embryonic stages were examined for the expression of *Sox2* (*A*, *C*, *E*, and *G*) and *Sox6* (*B*, *D*, *F*, and *H*) by RNA in situ hybridization. Adjacent sections at the forelimb level were used for comparison.

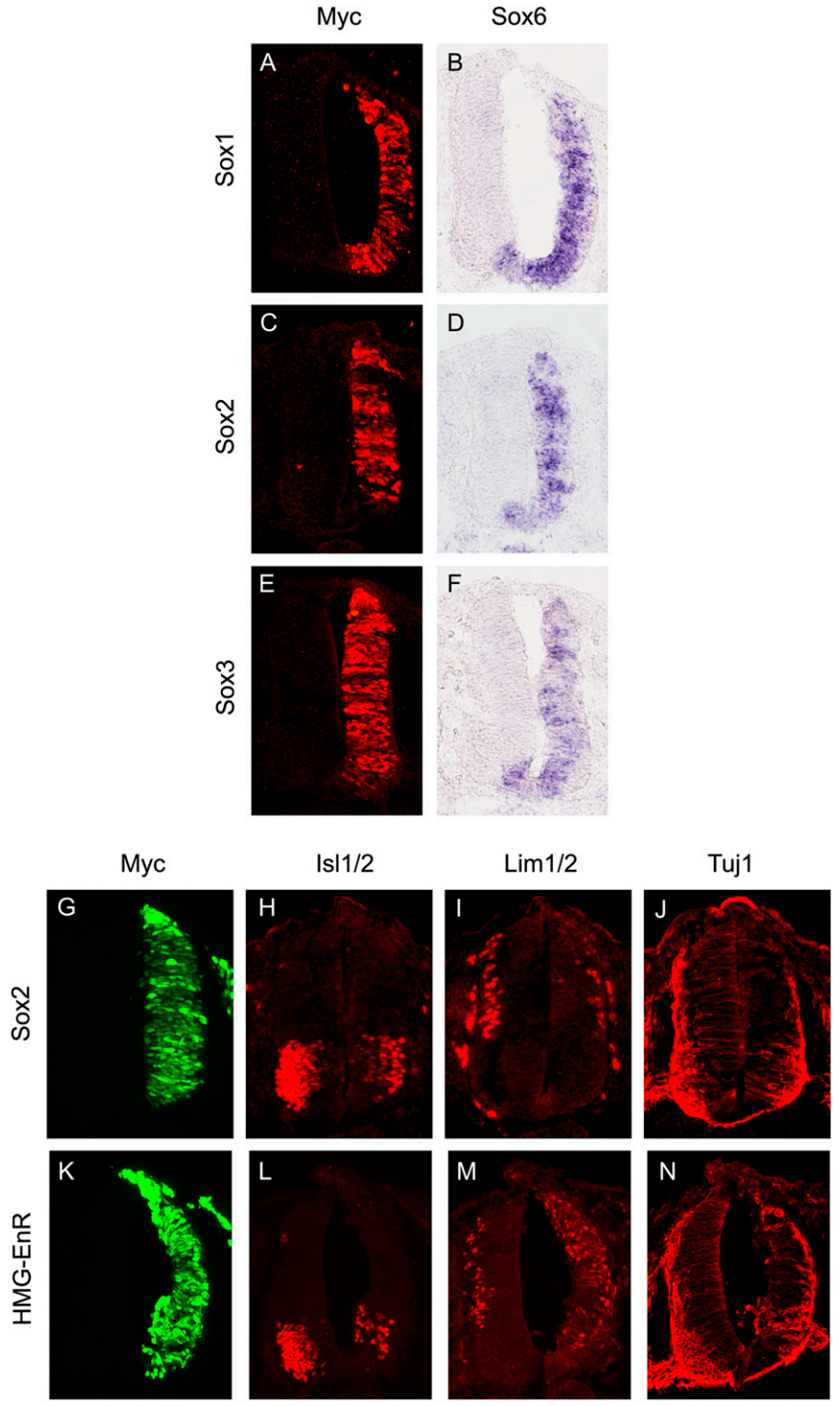


Fig. S5. Sox2 inhibits neuronal differentiation. Myc-tagged SoxB1 genes were overexpressed in chicken embryo neural tube (*A*, *C*, and *E*), and the expression of *Sox6* was examined by RNA in situ hybridization (*B*, *D*, and *F*). All three SoxB1 genes redundantly induce *Sox6* expression. (*G–J*) Sox2 expression inhibits neuronal differentiation. Sox2 expression is confirmed by myc staining (*G*). Differentiation of Islet-1/2+ (Isl1/2+) (*H*), Lim1/2+ (*I*), and Tuj1+ (*J*) neurons is inhibited. (*K–N*) HMG-EnR induces precocious neuronal differentiation. HMG-EnR expression is confirmed by myc staining (*K*). Isl1/2+ (*L*), Lim1/2+ (*M*), and Tuj1+ (*N*) neurons are seen in the presumptive ventricular zone (VZ) and not confined in the marginal zone.

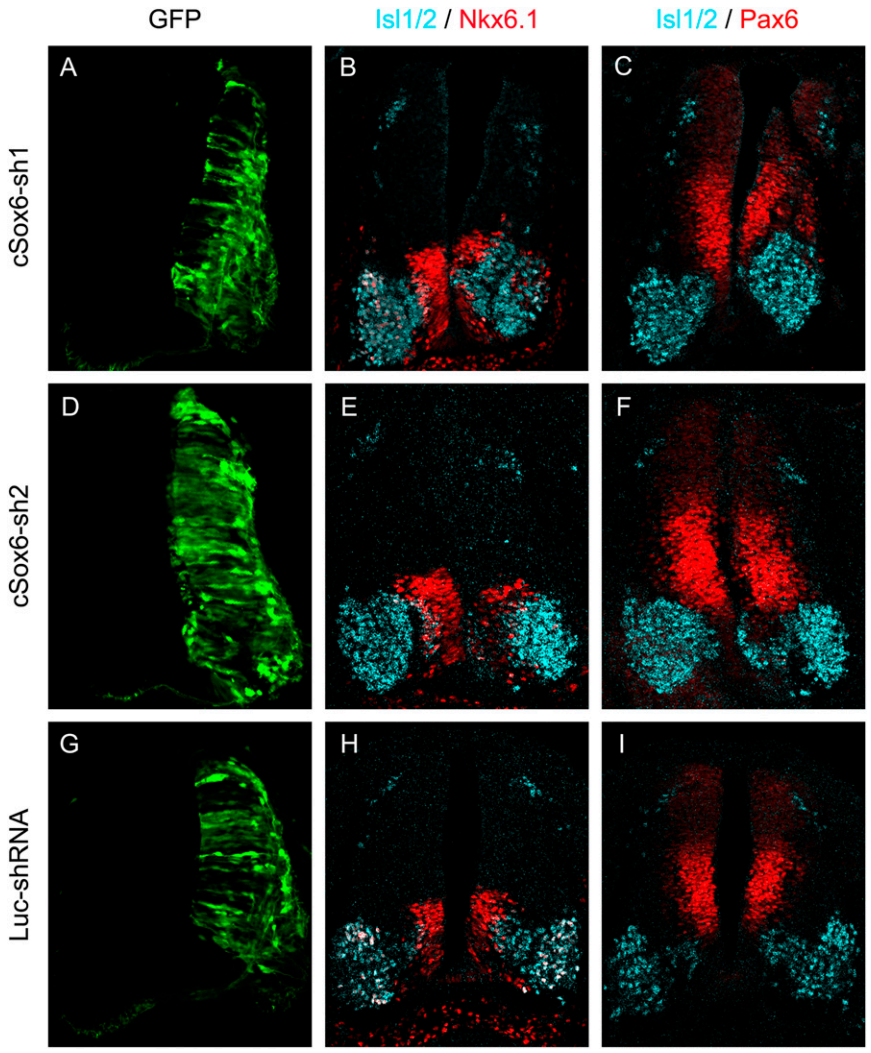


Fig. S6. Knockdown of *Sox6* inhibits expression of NK6 homeobox 1 (Nkx6.1), a VZ marker. c*Sox6*-sh1, c*Sox6*-sh2, and Luc-shRNA are expressed in chicken embryo neural tube, and expression of VZ markers Nkx6.1 and Pax6 is examined by immunohistochemical staining. Electroporation of c*Sox6*-sh1 (*A*), c*Sox6*-sh2 (*D*), and Luc-shRNA (*G*) are confirmed by GFP expression. Coimmunostaining for Isl1/2 and Nkx6.1 show that expression of c*Sox6*-sh1 (*B*) and c*Sox6*-sh2 (*E*) inhibits Nkx6.1 expression while inducing differentiation of Isl1/2+ neurons in the VZ, which is not seen with the control Luc-shRNA expression (*H*). In contrast, expression of c*Sox6*-sh1 (*C*) and c*Sox6*-sh2 (*F*) has a minor effect on Pax6 expression as is the case with Luc-shRNA expression (*I*).

Table S1. Oligonucleotides used for ChIP assay

No.	Gene	Forward	Reverse
1	GAPDH	AAG GTC ATC CCT GAG CTG AA	TGC TGT AGC CAA ATT CGT TG
2	HPRT	ACA GCT TGC TGG TGA AAA GG	ACT GGC AAA TGT GCC TCT CT
3	β-Actin	TCG TGC GTG ACA TTA AGG AG	AGG AAG GAA GGC TGG AAG AG
4	Sox6	AGG GCA TTA AAA TGC TGC AC	TTC TAC AGA GTG CCG GTA TGG
5	PIPOX	GGG TCT TAT TCT GGG CAT CG	CTT CCT TTG CCA CCT TCC TG
6	HMGB1	TGC AAT GGC TGT GAG AGC G	ACA TGC ACA CAT ACA CCA TAG AGC
7	ETV1	GCG AGG TTC TTC CCG CAG	CCA ACC ACG TGA CCA AGA AG
8	CBY1	TGC TAC TCT GAG TCC AGG CG	GAC CAG GAC CTG CAT TCA CC
9	ARID2	TGT GCG GCA CAT TGC AGA C	CTG GCA GCC AGT GCC ACA C

Dataset S1. Complete list and key features of the 811 significant peaks obtained from ChIP-on-chip assay

Dataset S1