Legends of supplementary data

Supplementary Table 1. Randomly selected DNA sequences for RNA library.

Supplementary Table 2. The sequences of 51 unique RNAs bound to Sox2 after 12 cycles of SELEX. The enriched motifs are underlined.

Supplementary Table 3. Sequences for RNAs and DNAs used for binding assays. Substituted bases are underlined.

Supplementary Table 4. Primers used for plasmid constructions. The restriction sites are underlined. Sequence for FLAG tag is highlighted in bold face.

Supplementary Table 5. Oligos used for RNA SELEX library construction. The T7 promoter sequence is underlined.

Supplementary Table 6. Primers used for ChIP-qPCR.

Supplementary Table 7. Primers used for RT-PCR and qRT-PCR.

Supplementary Figure 1. Purification of Sox2. Purified Sox2 was separated with a 12% SDS-PAGE gel and subjected to Coomassie blue staining (CS) and Western blot (WB).

Supplementary Figure 2. PAR-CLIP RNA targets of Sox2 in HEK293T cells. (**A**) Genome distribution of RNA species identified by Sox2 PAR-CLIP sequencing. (**B**) Top motifs identified by DREME in PAR-CLIP targets. The “CCCY” (where Y is A or G) motif (framed in red) was identified in intronic regions with E-value=1.2e-011.

Supplementary Figure 3. Verification of RNA binding motif of Sox2. (**A**) Comparison of the binding activities of Sox2 and Sox2-ΔRBM to the 12^{th}-24 and 12^{th}-24 Mut6 RNAs. Concentrations of Sox2 constructs are labelled above each lane. The percentages of bound RNA are indicated under each lane. Free RNA and protein-RNA complex are marked. (**B**) Schematic representation of the various His_6-Sox2 constructs used for binding assays. (**C**) Coomassie blue staining of purified His_6-Sox2 constructs. M: protein marker and sizes are shown in kDa. (**D**-**E**) Comparison of the binding capacities of Sox2 constructs to the 12^{th}-24 RNA and mutants thereof. Fractions of bound RNA (left panels) were determined by EMSAs (right panels) using densitometric analysis with the ImageQuantTL software. Concentrations of Sox2 constructs are labeled above each lane. Free RNA and protein-RNA complex are marked. Barplots represent the mean ± SD (n=3). (**F**) EMSAs to probe binding of Sox2(120-319) and Sox2(180-319) to the 12^{th}-24 RNA and mutants thereof. Concentrations of Sox2 constructs are labelled above each lane. The percentages of bound RNA are indicated under each lane. Free RNA and protein-RNA complexes are marked.

Supplementary Figure 4. Defining sequence determinants of the Sox2-RNA interaction. (**A**-**G**)

Secondary structures of the RNA constructs based on 12th-24 were predicted using Mfold (http://unafold.rna.albany.edu/?q=mfold). Mutated bases are highlighted in red. **Of note:** 12th-24 Mut6 is predicted to induce profound structural changes. EMSAs were performed with increasing concentrations of Sox2 (**H**) or Sox2-ΔRBM (**I**) and the 12th-24 RNA, Mut2, Mut4 and Mut5. Fractions of bound RNA are calculated. Barplots represent the mean ± SD (n=3).

Supplementary Figure 5. Characterizing exogenous constructs and pluripotency gene expression in reprogramming. (**A**) The relative transcript levels of exogenously expressed Oct4, Klf4, c-Myc, and Sox11 or Sox2 at day 4 of reprogramming were analyzed by qRT–PCR. Barplots represent the mean ± SD (n=3). (**B**) The relative transcript levels of selected pluripotency markers were analyzed by qRT–PCR. Data are represented as mean ± SD (n=3). ANOVA was used to assess significance (** $P < 0.01$).

Supplementary Figure 6. Characterizing the effects of Sox2-ΔRBM and Sox2-ΔHMG on intact Sox2 in reprogramming. (**A**) The relative transcript levels of exogenously expressed Oct4, Klf4, c-Myc, Sox2 at day 4 of reprogramming were analyzed by qRT–PCR. Barplots represent the mean ± SD (n=3). (**B**) The relative transcript levels of selected pluripotency markers were analyzed by qRT–PCR. Data are represented as mean ± SD (n=3). ANOVA was used to assess significance (* $P < 0.05$, ** $P < 0.01$).

Supplementary Figure 7. Evaluation of the effects of RBM on Sox2-chromatin and Sox2-Oct4 interactions. (**A**) Chromatin immunoprecipitation (ChIP) analysis for Sox2 and Sox2-ΔRBM on core pluripotent genes in reprogramming MEFs at day 4. Relative enrichments to IgG-ChIP control (upper panel) are shown. ChIP-Seq binding view of the above genes in MEFs after 48-hour infected with Oct4, Sox2, Klf4 and c-Myc, pre-iPSCs-1, pre-iPSCs-2 and mESCs are also shown [1]. Images were created with the Integrative Genomic Viewer (IGV) [2]. Data are normalized to total counts, and the scale rage is 0-20. The regions used for qPCR testing are highlighted under each panel. (**B**) MEF cells were transfected with $FLAG_3$-Sox2 along with Oct4, Klf4 and c-Myc. Co-immunoprecipitations were performed with IgG or anti-FLAG antibodies as shown on the top of each panel and immunoblots were done with anti-FLAG, anti-Sox2 and anti-Oct4.

Supplementary Figure 8. Comparison of gene expression and alternative splicing between reprogramming cells transduced with OKMS, OKMΔHMG and OKMΔRBM at day 12. (**A**) Heatmap depicting different expression of early ecto-, endo-, and mesodermal genes during reprogramming with Sox2, Sox2-ΔHMG, and Sox2-ΔRBM alongside OKM. (**B**) The Venn diagram comparing the alternatively spliced genes between OKMS versus OKMΔRBM and OKMS versus OKMΔHMG. (**C**) Scatterplot comparing exon inclusion changes in OKMS versus OKMΔHMG to those in OKMS versus OKMΔRBM. Pearson correlation coefficient value is shown in top right corner.

Supplementary Figure 9. RIP of indicated transcripts in mESC using anti-Sox2 antibody. RIP enrichment was measured by qRT–PCR, and values were compared to background immunoprecipitation measured by isotype IgG. ANOVA was used to assess statistical

significance (* $P < 0.05$, ** $P < 0.01$).

Supplementary Figure 10. Characterization of nucleotide composition around 5' splice sites. Nucleotide frequencies around the 5' splice sites of exons affected (n= 749) and unaffected by AS (n= 14,342). The exon sequences used are from **Figure 7E**.

Supplementary Figure 11. EMSAs to compare the binding activities of Sox2(180-319) (**A**), Sox2(120-319) (**B**) and Sox11 (**C**) to RNA probes derived from *Dnmt3b* and *Dicer1* mRNAs. Concentrations of Sox2 constructs are labelled above each lane. Free RNA and protein-RNA complex are marked.

Supplementary Table 1. Randomly selected DNA sequences from the RNA library.

Number	Sequence
1	GCGGGATGGTTAGGTGCGCCCTGTT
2	CGAGGTACACACTTAGGTTATATCG
3	AGTCTATCTTTCGTTCAAGATATGC
4	CATACCTCGCTCAAACTTGCCTCTG
5	TGCATCGGTCCTGCCTTGTCATCGA
6	TTTCCTAGCCCACTCCGAGCCTGGT
7	TTGGTATTCTTGATACATACATATT
8	CTAGGAGCCGACGAAGCATAAACGC
9	AGATGTTTGCTGCAATGGATCCGTT
10	TAGCGAATCTCCAAATCGTATTAGG
11	TTGACGAGCGCCTCGCAAAATCTCT
12	TAGCGTTTGTGTAAGGGCATCGGGC
13	TTTCAGGGGGGCGGGTCAGTAGTTT
14	CGAATTGTGATGGCCTGTAACATTA
15	CATCTATCGGACGCATTGCAGGCGC
16	ACGCGTTGAGCCCCCGCCCACGCTT
17	TCCTAAGCTCCGTTATCTTCCCTCA

Supplementary Table 2. The sequences of 51 unique RNAs bound to Sox2 after 12 cycles of SELEX.

NO.	Sequence
12^{th}-1	UGACUCCCACAUUAUUCGUGGUACA
12^{th}-2	GCCAGCCCAUAUUGGCGCUCCUUGU
12^{th}-3	GCCGGCCCAGAGGCGUACAUCAGUC
12^{th}-4	UUGACGCUGCGCCCUCACGCCUUGU
12^{th}-5	GACUCGCCCCCUCACGUCCCCCUGA
12^{th}-6	CCAAGUCCCGCCGGGAUUUUGCCUA
12^{th}-7	GCUCUGCCCACGUGGCGUUUUCUAC
12^{th}-8	GCGGUUUGCCCAUGCUACGCGUGUA
12^{th}-9	GCGGUCCGCCCAUGCCUUCGCGUCA
12^{th}-10	GCGGUUCGCCCAUGCCUCGCGCCAU
12^{th}-11	GCGGUUCGCCCAGCCUCGCGUCAUA
12^{th}-12	GCGGUUCGCCCAGCCUUGCGUCAUG
12^{th}-13	GCGGUUCGCCCAGCCUCGCGUCAUG
12^{th}-14	GCAACGACGCCCUACCCUCCCUCCC
12^{th}-15	GCGGCAAGCCCAUACUCGCGUUCAA
12^{th}-16	GGCGCGGUAAGCCCCAGUCGCGCCG
12^{th}-17	GCCGGCCCGCAAUAGGCGCUACUUC
12^{th}-18	GCGCCGGCCCCCAGGCGCUAGUUUC
12^{th}-19	GCGCCGGCCCCUUCGGCGCUGGCCU
12^{th}-20	UCAGCCGGUGCCCGCACGGAUGGAA
12^{th}-21	ACGCGGGCCCCUCCGCGCUCUUAUC
12^{th}-22	GCGGUGUCCCGCUCGCGCUACUUCA
12^{th}-23	ACGCCUUGCCCAUAAGCGGCGCUAU
12^{th}-24	CGCGGCUUGCCCUAUCGCGGUGUAG
12^{th}-25	CCCCGUACCCCUCGGCGGGUAUUCA
12^{th}-26	ACGCUCGGCCCACGUCGGCGCUAAA
12^{th}-27	GCGGUACCCCCCGCGUUUGCGAAUU
12^{th}-28	GCGCGGGCCCAAGCUUCCGCGCUGG
12^{th}-29	CGCGCGGGCCCAUUGAACCGCGCAU
12^{th}-30	GCGGAAAGCCCACAUAUCGCGUCAU
12^{th}-31	GCGCGGAAAGCCCACCUUCGCGCUC
12^{th}-32	GCGCGGAAUGCCCGAAUCGCGCGCA
12^{th}-33	GCGCGGAAGUCCCAUGAUUCGCGCA
12^{th}-34	GCGCGGAAUUCCCAUGUCUCGCGCA

12th-35	CGCGGAAACCCAGACUCGCGGCCCA
12th-36	CGCGGAAGCCCUGAGUCGCGGUAGG
12th-37	CGCGGAAGCCCUGAAUCGCGGUAGG
12th-38	CGCGGAAGCCCAAUGUUCGCGGGAA
12th-39	GCGGAAACCCAUACUUCGCGUUCAA
12th-40	GCGGUGGCCCAGUAGCCGCUAAGAA
12th-41	CGCGGUACGCCCAUACUUCGCGGGG
12th-42	GCGGUAGCCCAGUUCGCGCUCUUCA
12th-43	CCGCGGAAACCCACACCUCGCGGGC
12th-44	GCGGUACGCCCCAUCGCGUUCUGAG
12th-45	CGCGGAAACCCACAUCGCGGUAUGA
12th-46	CCCGGAAGCCCAGAUCGGGGUGUUA
12th-47	CCCCGAAGCCCAGAUCGGGGUGUUA
12th-48	ACGCGGAAGCCCAGGUCGCGCUUUA
12th-49	ACGCGGAAACCCAGGUCGCGCUUUA
12th-50	GCGGUACGCCCAGCUUCGCUUCGUG
12th-51	GCGCGGUAAGCCCCAUCGCGCUACU

The enriched motifs are underlined.

Supplementary Table 3. Sequences for RNAs and DNAs used for binding assays.

Name	Sequence
12^{th}-15	5'-GCGGCAAGCCCAUACUCGCGUUCAA-3'
12^{th}-24	5'-CGCGGCUUGCCCUAUCGCGGUGUAG-3'
12^{th}-24 Mut1	5'-UAUAGCUUGCCCUAUCGCGGUGUAG-3'
12^{th}-24 Mut2	5'-CGCGGCUUGUUUUAUCGCGGUGUAG-3'
12^{th}-24 Mut3	5'-CGCGGCUUGCCCUAUUAUAGUGUAG-3'
12^{th}-24 Mut4	5'-CGCGGCUUGGGGUAUCGCGGUGUAG-3'
12^{th}-24 Mut5	5'-CGCGGCUUGAAAUAUCGCGGUGUAG-3'
12^{th}-24 Mut6	5'-UAUAGCUUGAAAUAUUAUAGUGUAG-3'
RL	5'-CACACACACACACACACACACACAC-3'
Dnmt3b RNA	5'-CAGCUUUUCCCUGUAGUCAUGAAUG-3'
Dicer1 RNA	5'-CCUGCAGCUCAUGACCCCUGCUUCC-3'
Sox2/Prox1 DNA	5'-CTAGCATTGTCTGGG-3' 3'-GATCGTAACAGACCC-5'

Substituted bases were underlined.

Supplementary Table 4. Primers used for plasmid constructions.

	Sequence
For protein expression plasmid construction	
Sox2- forward	GGAATTCCATATGTATAACATGATGGAGACGGAGC
Sox2-reverse	CCGCTCGAGTCACATGTGCGACAGGGGCA
Sox2(1-180) -reverse	CCGCTCGAGTCACAGCTGCTCCTGCATCATGCTGT
Sox2(1-120)-reverse	CCGCTCGAG TCA CTTGGTTTTCCGCCGCGGC
Sox2-HMG- forward	GGAATTCCATATGGACCGCGTCAAGAGGCCCATGAA CGCC
Sox2-HMG-reverse	CCGCTCGAGTCACTTGGTTTTCCGCCGCGGCC
Sox2-ΔHMG(1-40)-reverse	GTACTTATCCTTCTTCATGAGCGTCGGGCTGTTCTTCTGGTTGC
Sox2-ΔHMG(119-319)- forward	GCAACCAGAAGAACAGCCCGACGCTCATGAAGAAGGATAAGTAC
Sox11- forward	GGAATTCCATATGGTGCAGCAGGCCGAG
Sox11-reverse	CCGCTCGAGTCAATACGTGAACACCAGGTC
Sox11N2C- forward	GCCGCGCAAAAAGCCCAAGACGCTCATGAAGAAGGATAAGTACA
Sox11N2C-reverse	TGTACTTATCCTTCTTCATGAGCGTCTTGGGCTTTTTGCGCGGC
For retroviral plasmid construction	
V-Sox2- forward	CGCGGATCCTATAACATGATGGAGACGGAGC
V-Sox2-reverse	CCGCTCGAGTCACATGTGCGACAGGGGCA
V-Sox2(1-180) -reverse	CCGCTCGAGTCACAGCTGCTCCTGCATCATGCTGT
V-Sox2(1-120)-reverse	CCGCTCGAGTCACTTGGTTTTCCGCCGCGGC
V-Sox2-HMG-forward	CGCGGATCCGACCGCGTCAAGAGGCCCATGAA CGCC
V-Sox2-HMG-reverse	CCGCTCGAGTCACTTGGTTTTCCGCCGCGGCC
V-Sox11- forward	CGCGGATCCGTGCAGCAGGCCGAG
V-Sox11-reverse	CCGCTCGAGTCAATACGTGAACACCAGGTC
V-Sox2-FLAG-reverse	CCGCTCGAG**TCACTTGTCATCGTCGTCCTTGTAGTC**GGAGAATTCGTATAC ACTAGTCATGTGCGACAGGGGCA

The underlined are restriction sites. Sequence for FLAG tag is highlighted in bold face.

Supplementary Table 5. Oligos used for RNA SELEX library construction.

Name	Sequence
DNA library	5’-TGGGCACTATTTATATCAAC$(N)_{25}$AATGTCGTTGGTGGCCC-3’
Forward primer	5’-CCCGACACCCGCGG ATCCATGGGCACTATTTATATCAAC-3’
Reverse primer	5’-CGCGGATCCTAATACGACTCACTATAGGGGCCACCAACGACATT-3’

The T7 promoter sequence is underlined.

Supplementary Table 6. Primers used for ChIP-qPCR.

	Sequence
ChIP-Sox2-forward	AGTTCCCAGTCCAAGCTAGG
ChIP-Sox2-reverse	CCGTCATTTGGGTCTTTATTCAA
ChIP-Rest-forward	CCATTGGCCTATTTGCATCAC
ChIP-Rest-reverse	AGAGGGATCACAGCCTAGTC
ChIP-Nanog-forward	CAAGGCTAGCGATTCACACC
ChIP-Nanog-reverse	AATAGGGAGGAGGGCGTCTA
ChIP-Fbxo15-forward	TTTGACTTATTGCACGGCCC
ChIP-Fbxo15-reverse	GCAGCCAGGGATTCTTGTCA
ChIP-Atg7-forward	GAGCCCTGTTCTTACTCTAGTCA
ChIP-Atg7-reverse	AAATACAGGGATGCACACCG

Supplementary Table 7. Primers used for RT-PCR and qRT-PCR.

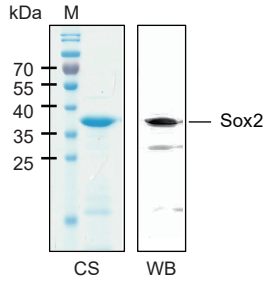
	Sequence
For RT-PCR	
Exo-Sox2-forward	GGGTGGACCATCCTCTAGAC
Exo-Sox2-reverse	GGGCTGTTCTTCTGGTTG
Cy5-Srebf1-forward	ATCAAAGAGGAGCCAGTGCC
Srebf1-reverse	TGTCAGCAGCAGTGAGTCTG
Cy5-Ctbp1-forward	AGCTCCCACTTGCTCAACAA
Ctbp1-reverse	GCTACTGTGGCCACATCCTT
Cy5-Dnmt3b-forward	CCCAGTGATGATCGATGCCA
Dnmt3b-reverse	ATGACCGGTACACTCCAGGA
Cy5-Lef1-forward	ACCTTCTACCCCCTGTCTCC
Lef1-reverse	GGTGCTCCTGTTTGACCTGA
Cy5-Prmt9-forward	CCTGTGATGTAGTGGCTGCA
Prmt9-reverse	ATCTCTGCACACTCGACTGC
Cy5-Dicer1-forward	ACTCTGAAAGAACTTAGAGTCAGCA
Dicer1-reverse	TGAGCCAGTGTTCAAGCACA
Cy5-Tada2a-forward	CCCTCTGTTTGCATCCACCCTG
Tada2a-reverse	CATGTAAGATGTCTGAGTCGTCTTC
Cy5-Gapdh-2-forward	CCAGGAGCGAGACCCCACTA
Gapdh-2-reverse	GCAGTGATGGCATGGACTGTGGT
For RIP qRT-PCR using cell lysate	
Srebf1-forward	GGCTATTCCGTGAACATCTCC
Srebf1-reverse	TCCAAGGGCATCTGAGAACT
Ctbp1-forward	CTCCAGGCGTCGTGAGTG
Ctbp1-reverse	TCGTATGGTCTCTATCCGCC
Prmt9-forward	CTGGTGGCACGGTCCTTG
Prmt9-reverse	CTTCACGTCGTCTTTTAGCTC
Dnmt3b-forward	GGCTTCCCTGCTCACTACA
Dnmt3b-reverse	TGGGTAGAACTATTCACAGGCA
Lef1-forward	GAGGCCTGTACAACAAGGGA
Lef1-reverse	GGTGGAGAAAGGGACCCATT
Dicer1-forward	AGCGGAGAGAACTTATGATGGA
Dicer1-reverse	AGCTGTTAGGAACCTGAGGC
Tada2a-forward	CCCTCTGTTTGCATCCACCCTG
Tada2a-reverse	CATGTAAGATGTCTGAGTCGTCTTC
Gapdh-2-forward	CCAGGAGCGAGACCCCACTA
Gapdh-2-reverse	GCAGTGATGGCATGGACTGTGGT
RMST-forward	AATGTGACTTGGGACTGGCC
RMST-reverse	CCGTTCTCTCCGAGGATTGA
For qRT-PCR and transgene insertion testing	
exo-Oct4-forward	GGGTGGACCATCCTCTAGAC

exo-Oct4-reverse	CCAGGTTCGAGAATCCAC
exo-Sox2-forward	GGGTGGACCATCCTCTAGAC
exo-Sox2-reverse	GGGCTGTTCTTCTGGTTG
exo-Klf4-forward	GGGTGGACCATCCTCTAGAC
exo-Klf4-reverse	GCTGGACGCAGTGTCTTCTC
exo-c-Myc-forward	GGGTGGACCATCCTCTAGAC
exo- c-Myc -reverse	CCTCGTCGCAGATGAAATAG
For qRT-PCR and selected pluripotency markers testing	
Oct4-forward	GGGTGGACCATCCTCTAGAC
Oct4-reverse	CCAGGTTCGAGAATCCAC
Nanog-forward	CTCAAGTCCTGAGGCTGACA
Nanog-reverse	TGAAACCTGTCCTTGAGTGC
Esrrb-forward	CAGAGTGCCTGGATGGAGAT
Esrrb-reverse	TACCAGGCGAGAGTGTTCCT
Sall4-forward	CCCTGGGAACTGCGATGAAG
Sall4-reverse	TCAGAGAGACTAAAGAACTCGGC

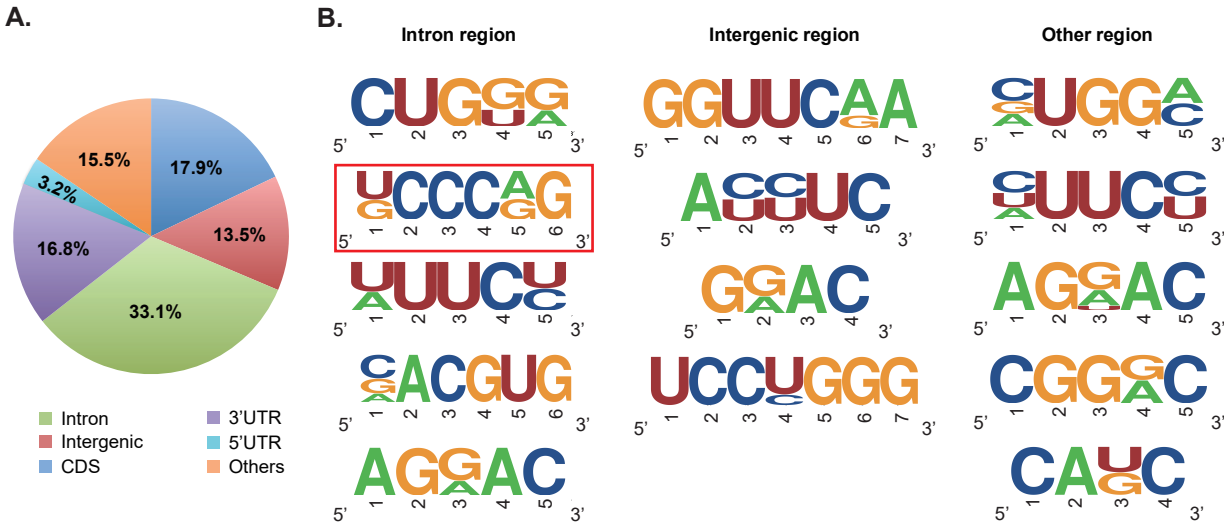
REFERENCES

1. Silva, J., O. Barrandon, J. Nichols, J. Kawaguchi, T.W. Theunissen, and A. Smith. (2008) Promotion of reprogramming to ground state pluripotency by signal inhibition. PLoS Biol, 6(10): p. e253.
2. Robinson, J.T., H. Thorvaldsdottir, W. Winckler, M. Guttman, E.S. Lander, G. Getz, and J.P. Mesirov. (2011) Integrative genomics viewer. Nat Biotechnol, 29(1): p. 24-6.

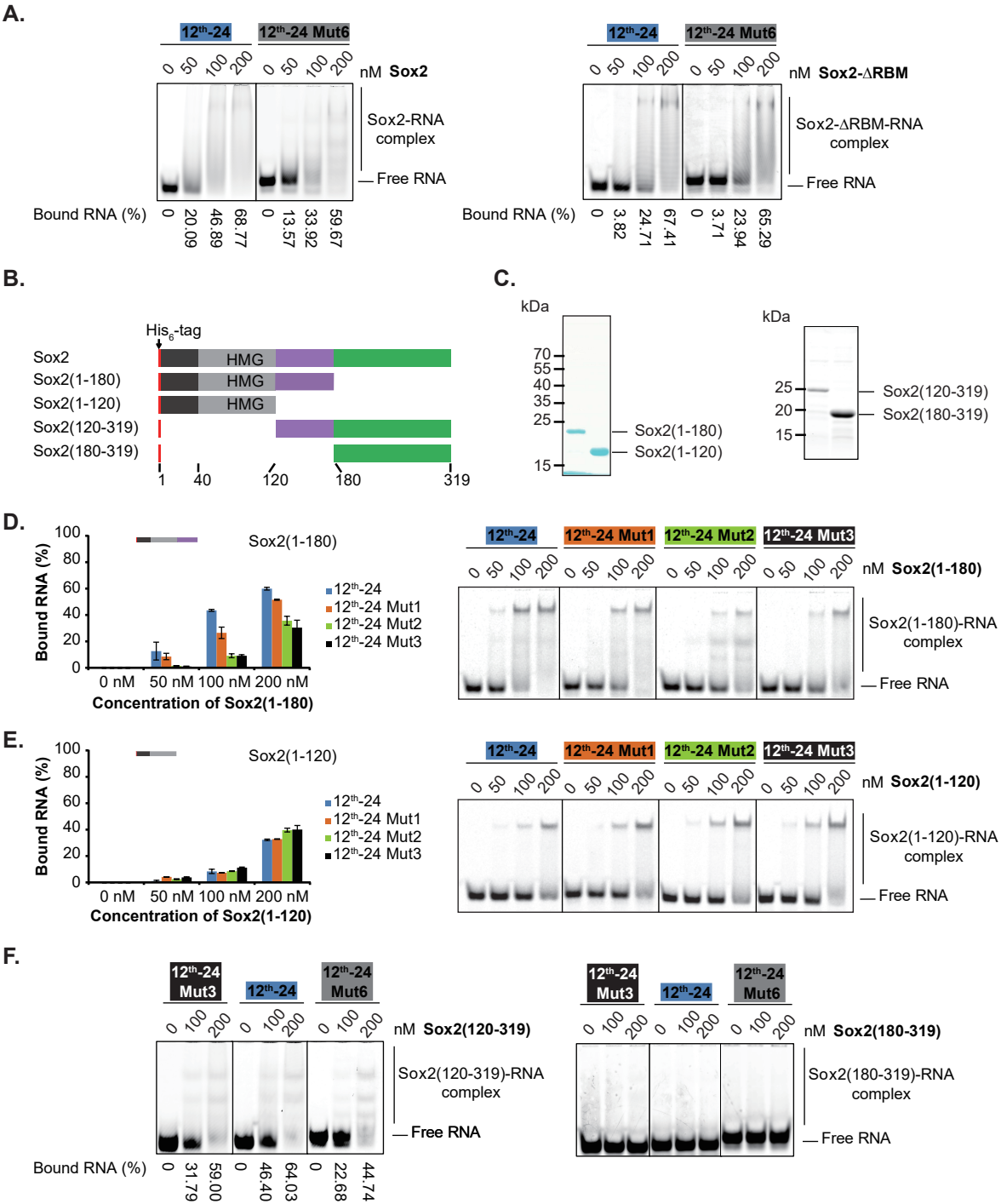
Supplementary Figure 1



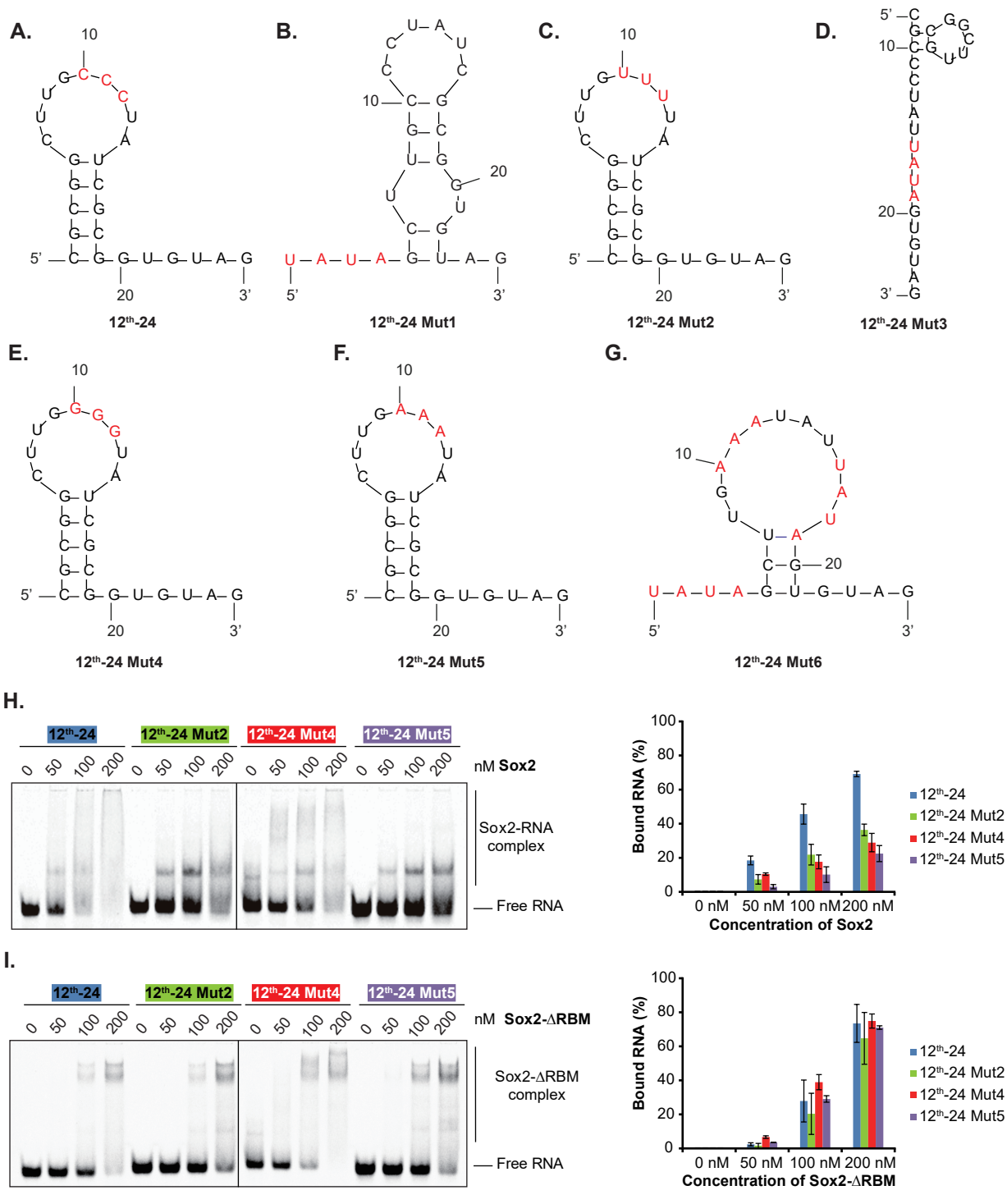
Supplementary Figure 2



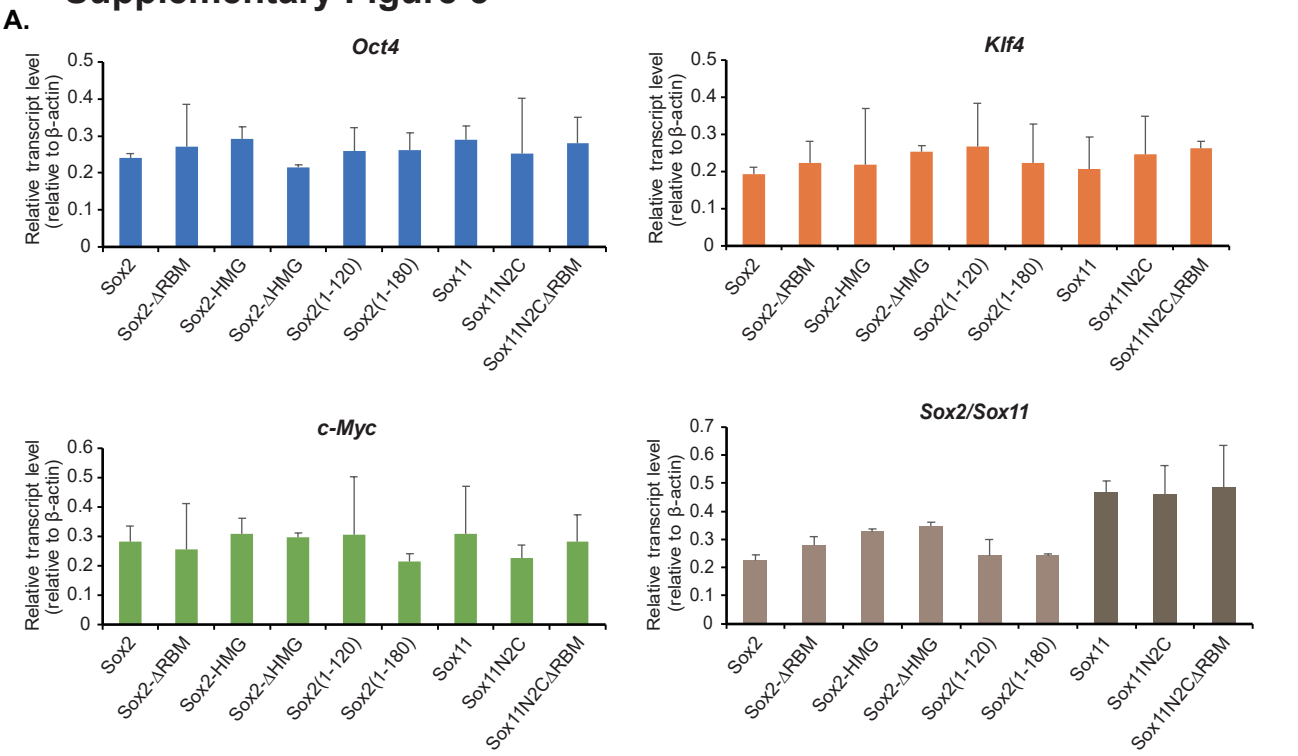
Supplementary Figure 3

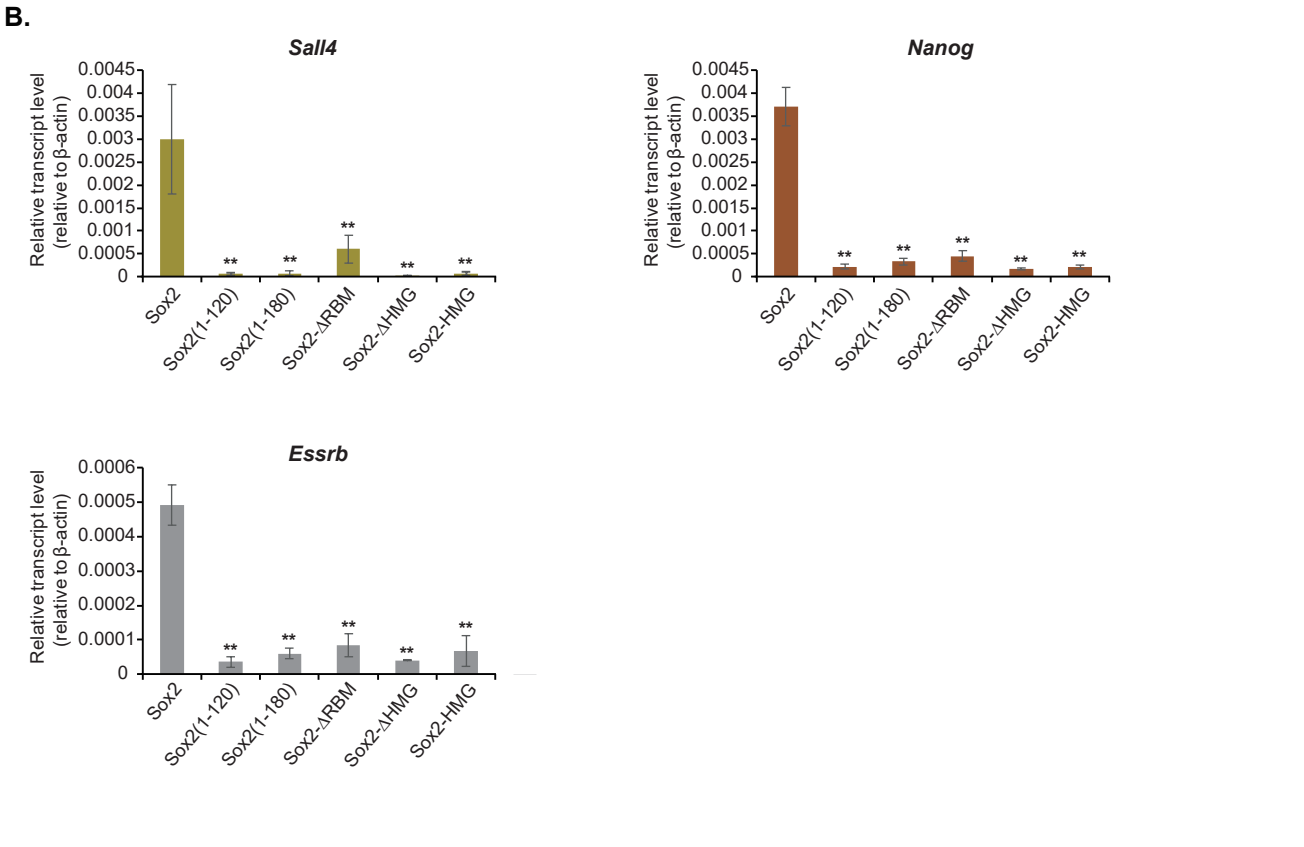


Supplementary Figure 4

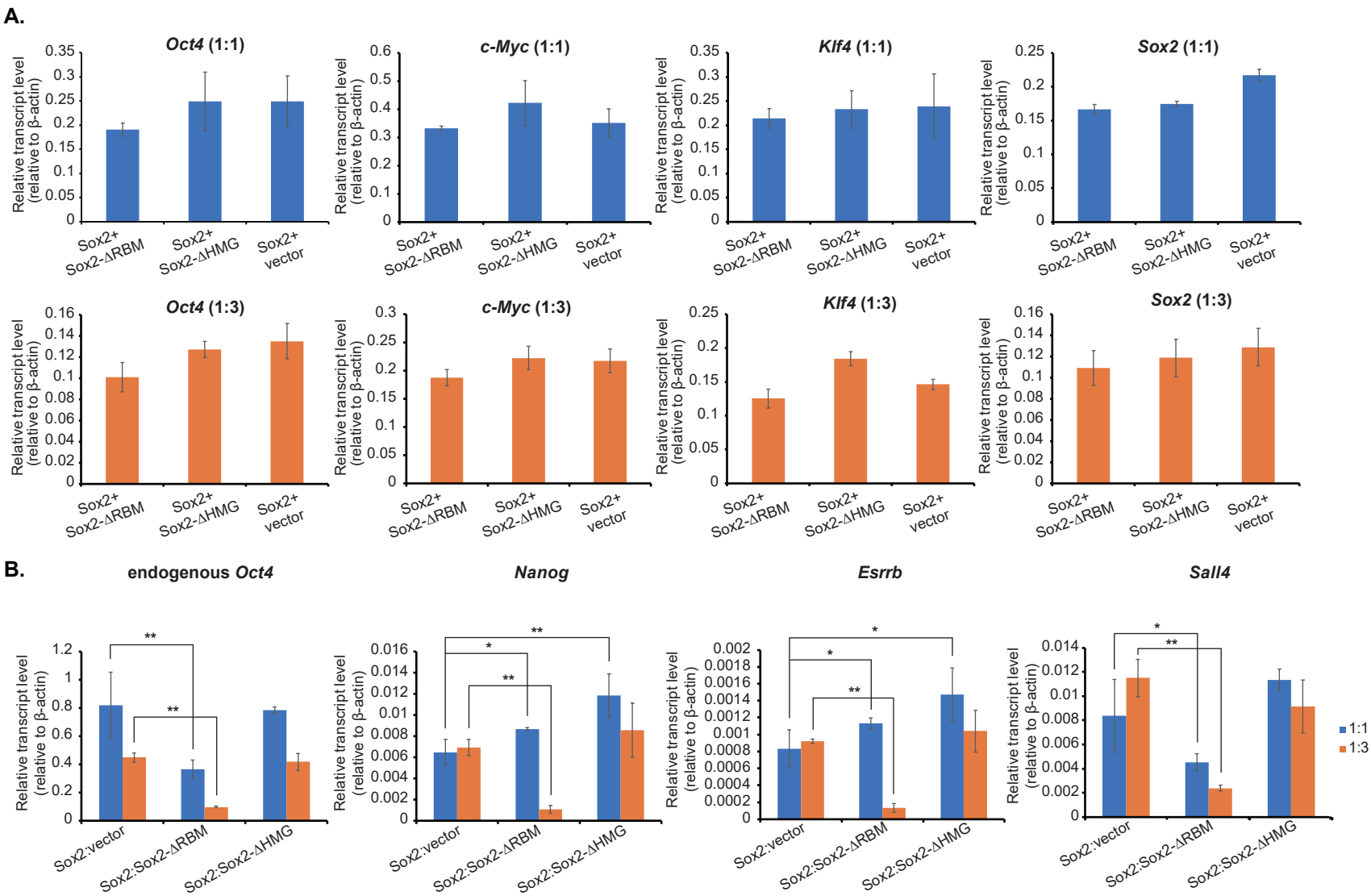


Supplementary Figure 5

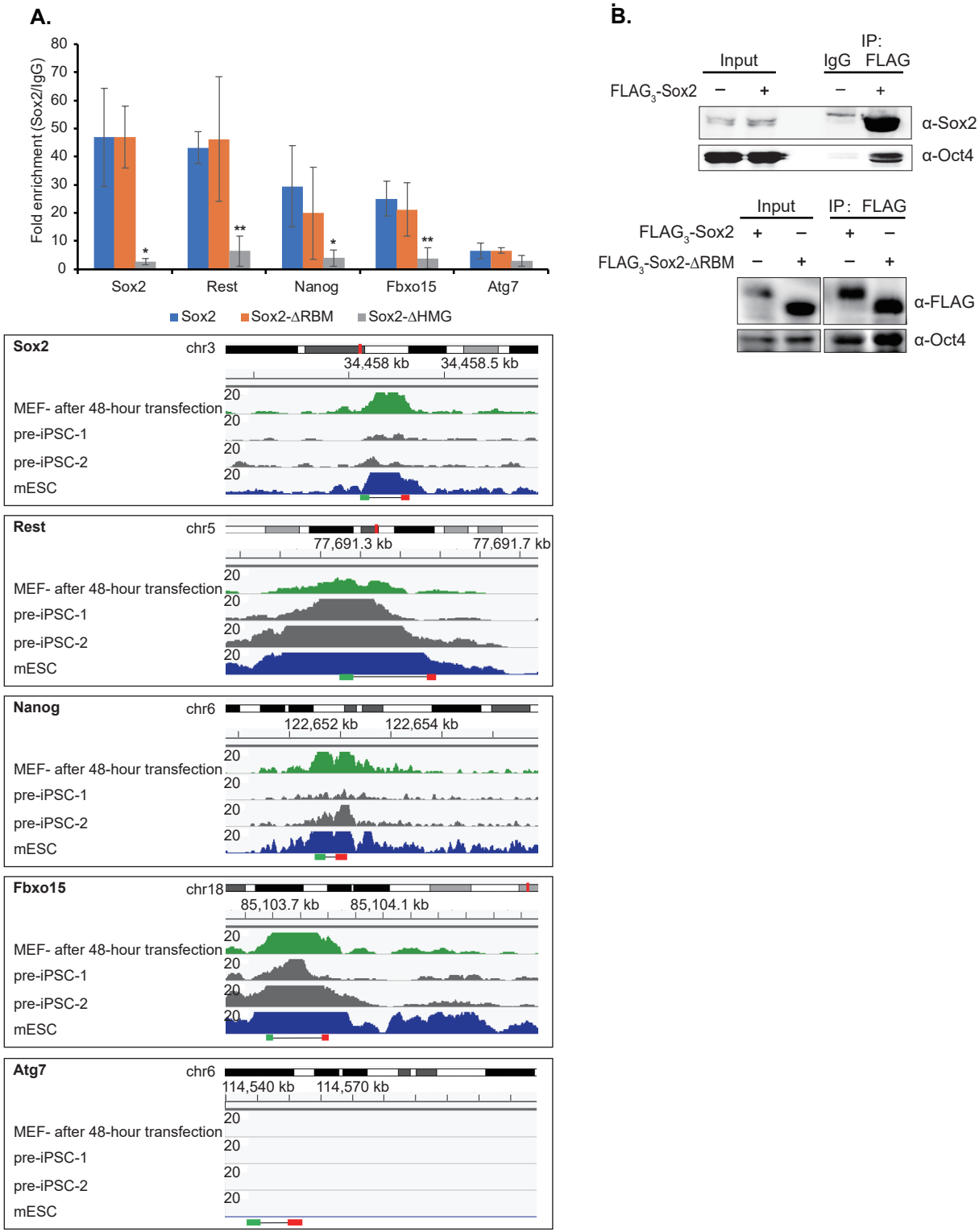




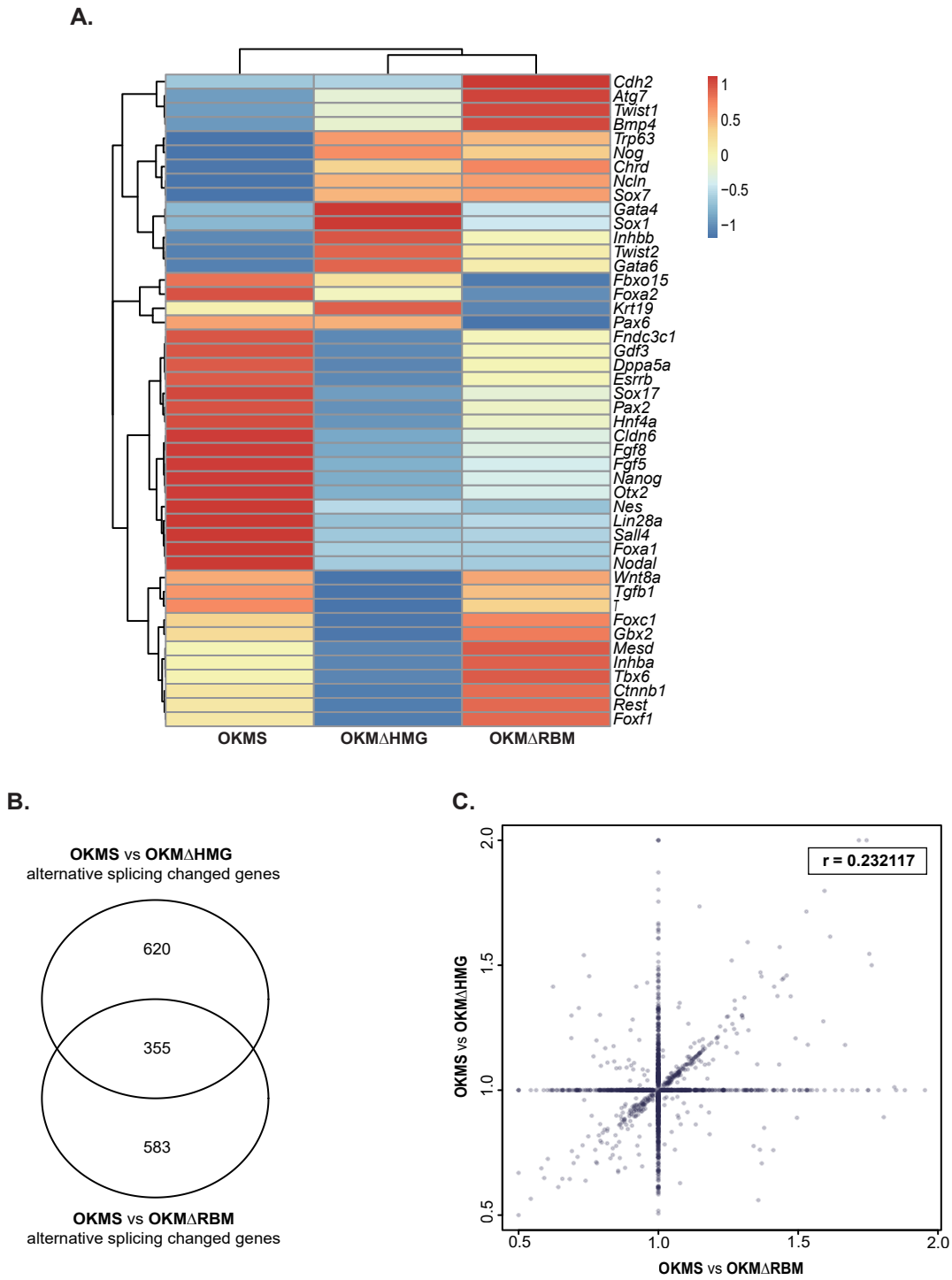
Supplementary Figure 6



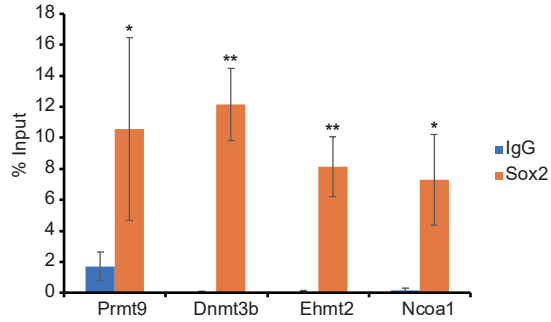
Supplementary Figure 7



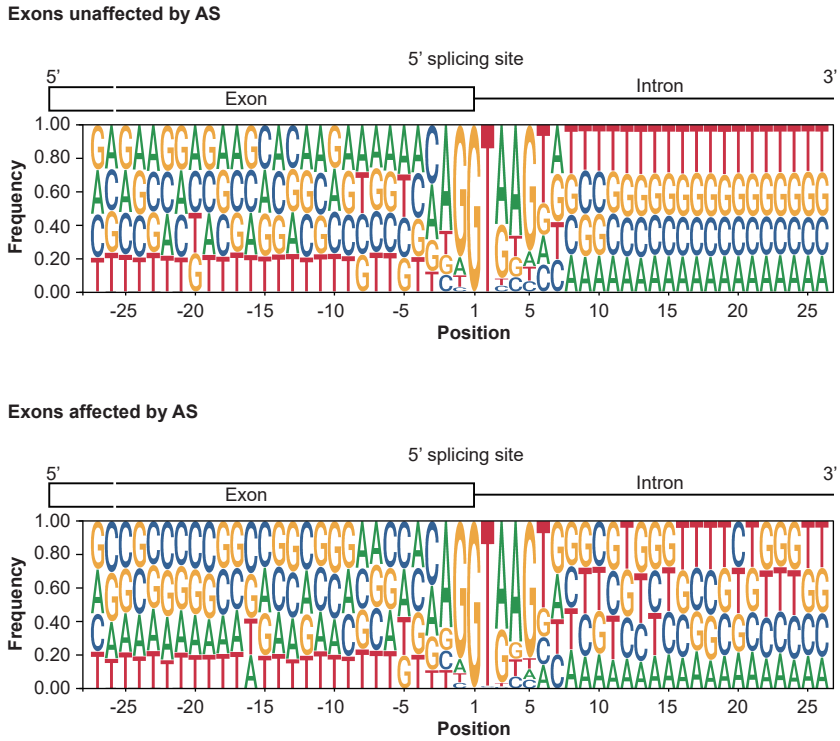
Supplementary Figure 8



Supplementary Figure 9



Supplementary Figure 10



Supplementary Figure 11