Profiling the long noncoding RNA interactome in the regulatory elements of target genes by chromatin in situ reverse transcription sequencing

Shilin Zhang1*, Yichen Wang1*, Lin Jia1*, Xue Wen1*, Zhonghua Du1*, Cong Wang1,2, Yajing Hao3, Dehai Yu1, Lei Zhou1, Naifei Chen1, Jingchen Chen1,2, Huiling Chen4, Hui Zhang7, Ilkay Celik2, Günhan Gülsoy6, Jianjun Luo3, Baoming Qin7, Xueling Cui8, Zhonghui Liu8, Songling Zhang1#, Miguel A. Esteban7#, Ferhat Ay5#, Wei Xu1#, Runsheng Chen3#, Wei Li1#, Andrew R. Hoffman2#, Ji-Fan Hu1,2#, Jiuwei Cui1#

1 Key Laboratory of Organ Regeneration and Transplantation of Ministry of Education, The First Hospital of Jilin University, Changchun, Jilin 130061, P.R. China
2 Stanford University Medical School, VA Palo Alto Health Care System, Palo Alto, CA 94304, USA
3 CAS Key Laboratory of RNA Biology, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China, P.R. China
4 Department of Endocrinology, Xiangya Hospital, Central South University, Changsha, Hunan, P.R. China
5 La Jolla Institute for Allergy and Immunology, La Jolla, California 92037, USA
6 Google Inc., Mountain View, CA 94043, USA
7 Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences, Guangzhou, Guangdong 510530, P.R. China
8 Department of Immunology, College of Basic Medical Sciences, Jilin University, Changchun, Jilin 130021, P.R. China
## Supplemental Tables

### Table S1. Top fifty RNAs that interact with the Sox2 promoter

| gene.id       | gene.name                    | Description                                                                 | locus               | p-value     | Quality |
|---------------|------------------------------|------------------------------------------------------------------------------|---------------------|-------------|---------|
| ENSMUSG00000075014 | Gm10800                        | predicted_gene_10800                                                          | 299666546-96667301  | 0.000       | 0.00    |
| ENSMUSG00000079312 | Gm26870                        | predicted_gene_26870                                                          | 2996662236-96664063 | 0.000       | 0.00    |
| ENSMUSG00000075015 | Gm10801                        | predicted_gene_10801                                                          | 2996662366-96664083 | 0.000       | 0.00    |
| ENSMUSG00000058691 | Gm10717                        | predicted_gene_10717                                                          | 2996662813-96663316 | 0.000       | 0.00    |
| ENSMUSG00000095280 | Gm21738                        | predicted_gene_21738                                                          | 3000002813-96663316 | 0.000       | 0.00    |
| ENSMUSG00000069547 | Gm10719                        | predicted_gene_10719                                                          | 3000002813-96663316 | 0.000       | 0.00    |
| ENSMUSG00000086519 | Gm10721                        | predicted_gene_10721                                                          | 3000002813-96663316 | 0.000       | 0.00    |
| ENSMUSG00000095186 | Gm10718                        | predicted_gene_10718                                                          | 3000002813-96663316 | 0.000       | 0.00    |
| ENSMUSG00000074564 | Gm10720                        | predicted_gene_10720                                                          | 3000002813-96663316 | 0.000       | 0.00    |
| ENSMUSG00000091028 | Gm10722                        | predicted_gene_10722                                                          | 3000002813-96663316 | 0.000       | 0.00    |
| ENSMUSG00000096201 | Gm10715                        | predicted_gene_10715                                                          | 3000002813-96663316 | 0.000       | 0.00    |
| ENSMUSG00000096662 | Gm10714                        | predicted_gene_10714                                                          | 3000002813-96663316 | 0.000       | 0.00    |
| ENSMUSG00000079600 | Gm17604                        | predicted_gene_17604                                                          | 3000002813-96663316 | 0.000       | 0.00    |
| ENSMUSG00000096736 | Gm17535                        | predicted_gene_17535                                                          | 3000002813-96663316 | 0.000       | 0.00    |
| ENSMUSG00000016234 | Gm43535                        | predicted_gene_43535                                                          | 3000002813-96663316 | 0.000       | 0.00    |
| ENSMUSG00000016028 | Gm43520                        | predicted_gene_43520                                                          | 3000002813-96663316 | 0.000       | 0.00    |
| ENSMUSG00000066810 | Gm10101                        | predicted_gene_10101                                                          | 3000002813-96663316 | 0.000       | 0.00    |
| ENSMUSG00000088607 | Pflx21                         | plxipoxeny_associated_transcript_21                                           | 3000002813-96663316 | 0.000       | 0.00    |
| ENSMUSG00000016612 | Gm43078                        | predicted_gene_43078                                                          | 3000002813-96663316 | 0.000       | 0.00    |
| ENSMUSG00000064945 | Rny3                          | RNA-Y3 small cytoplasmic-associated_with_Ro_protein                           | 3000002813-96663316 | 0.000       | 0.00    |
| ENSMUSG00000010826 | Snhg14                        | small_nuclear_RNA_host_gene_14                                               | 3000002813-96663316 | 0.000       | 0.00    |
| ENSMUSG00000010628 | Gm4358                          | predicted_gene_4358                                                          | 3000002813-96663316 | 0.000       | 0.00    |
| ENSMUSG00000031654 | Cbml1                          | cerebellin_1_precurso protein                                                 | 3000002813-96663316 | 0.000       | 0.00    |
| ENSMUSG00000026790 | Od71f                          | outer_dense_fiber_of_sperm_tails_2                                            | 3000002813-96663316 | 0.000       | 0.00    |
| ENSMUSG00000080592 | Gm17571                        | predicted_gene_17571                                                          | 3000002813-96663316 | 0.000       | 0.00    |
| ENSMUSG00000017569 | RP24-47KCI5.2                  | -                                                                            | 3000002813-96663316 | 0.000       | 0.00    |
| ENSMUSG00000013322 | Gm37404                        | predicted_gene_37404                                                          | 3000002813-96663316 | 0.000       | 0.00    |
| ENSMUSG00000085328 | Gm17131                        | predicted_gene_17131                                                          | 3000002813-96663316 | 0.000       | 0.00    |
| ENSMUSG00000094472 | Gm21987                        | predicted_gene_21987                                                          | 3000002813-96663316 | 0.000       | 0.00    |
| ENSMUSG00000055491 | Pprc1                          | peroxisome_proliferative_activsed_receptor_gama_coactivator_related_1         | 3000002813-96663316 | 0.000       | 0.00    |

2
Table S2. Top fifty pathways of the Sox2 promoter-interacting RNAs

| pathway ID | PATHWAY_DESC | target_gene_in_this_pathway | target_gene_in_all_pathway | rich_factor | P_Value | q_value |
|------------|--------------|-----------------------------|---------------------------|-------------|---------|---------|
| mmu0110    | Metabolic pathways | 143 | 121 | 1286 | 7510 | 0.7449539 | 1 | 1 |
| mmu0200    | Pathways in cancer | 56 | 121 | 397 | 7510 | 0.9450021 | 1 | 1 |
| mmu0415    | PDK-Akt signaling pathway | 53 | 121 | 345 | 7510 | 1.0281794 | 0.7723307 | 1 |
| mmu0460    | Neuroactive ligand-receptor interaction | 47 | 121 | 285 | 7510 | 1.1048033 | 0.4516955 | 1 |
| mmu0492    | ligand-receptor interaction | 45 | 121 | 198 | 7510 | 1.5225854 | 0.0086541 | 0.6585235 |
| mmu0414    | Endocytosis | 45 | 121 | 282 | 7510 | 1.0690493 | 0.5969877 | 1 |
| mmu0415    | Rap1 signaling pathway | 40 | 121 | 215 | 7510 | 1.2463955 | 0.4532003 | 1 |
| mmu0410    | MAPK signaling pathway | 40 | 121 | 258 | 7510 | 1.0591898 | 0.6498145 | 1 |
| mmu0420    | Calcium signaling pathway | 38 | 121 | 180 | 7510 | 1.6143126 | 0.0305302 | 1 |
| mmu0505    | Proteoglycans in cancer | 38 | 121 | 203 | 7510 | 1.2540793 | 0.1516755 | 1 |
| mmu0414    | RAS signaling pathway | 37 | 121 | 229 | 7510 | 1.0824319 | 0.5657807 | 1 |
| mmu0430    | Axon guidance | 36 | 121 | 175 | 7510 | 1.3781573 | 0.0522714 | 1 |
| mmu0480    | Regulation of actin cytoskeleton | 36 | 121 | 217 | 7510 | 1.1115171 | 0.4730308 | 1 |
| mmu0472    | Phosphatidylinositol 3-kinase pathway | 35 | 121 | 143 | 7510 | 1.6397073 | 0.0052294 | 0.7425711 |
| mmu0422    | cGMP-PKG signaling pathway | 35 | 121 | 172 | 7510 | 1.3623246 | 0.0882948 | 1 |
| mmu0450    | Focal adhesion | 35 | 121 | 203 | 7510 | 1.1556648 | 0.3512603 | 1 |
| mmu0470    | Olfactory transduction | 30 | 121 | 1082 | 7510 | 0.1867498 | 1 | 1 |
| mmu0474    | Glioma signaling pathway | 28 | 121 | 115 | 7510 | 1.6311523 | 0.0121716 | 0.4901462 |
| mmu0426    | Adrenergic signaling in cardiomyocytes | 28 | 121 | 140 | 7510 | 1.2594341 | 0.1978534 | 1 |
| mmu0430    | Hippo signaling pathway | 28 | 121 | 154 | 7510 | 1.2180583 | 0.2860805 | 1 |
| mmu0450    | Trifluridine signaling pathway | 27 | 121 | 140 | 7510 | 1.2920224 | 0.1633235 | 1 |
| mmu0421    | Cytokine-cytokine receptor interaction | 27 | 121 | 265 | 7510 | 0.6825779 | 1 | 1 |
| mmu0515    | Phosphatidylinositol 3-kinase pathway | 26 | 121 | 176 | 7510 | 0.9155512 | 1 | 1 |
| mmu0450    | Signaling pathways regulating pluripotency of stem cells | 25 | 121 | 140 | 7510 | 1.1006117 | 0.5638137 | 1 |
| mmu0452    | Chemokine signaling pathway | 23 | 121 | 197 | 7510 | 0.7821806 | 1 | 1 |
| mmu0496    | FoxO signaling pathway | 22 | 121 | 134 | 7510 | 1.0999875 | 0.5544964 | 1 |
| mmu0430    | Jak-STAT signaling pathway | 22 | 121 | 161 | 7510 | 0.9154435 | 1 | 1 |
| mmu0515    | Tuberculosis | 22 | 121 | 176 | 7510 | 0.8374219 | 1 | 1 |
| mmu0530    | Urease metabolism | 22 | 121 | 178 | 7510 | 0.9280127 | 1 | 1 |
| mmu0411    | Insulin secretion | 21 | 121 | 80 | 7510 | 1.6356834 | 0.0256943 | 1 |
| mmu0473    | Glioblastoma multiforme | 21 | 121 | 98 | 7510 | 1.4264906 | 0.0894218 | 1 |
| mmu0470    | Inflammatory mediator regulation of TRP channels | 21 | 121 | 126 | 7510 | 1.1166825 | 0.5246433 | 1 |
| mmu0491    | Insulin signaling pathway | 21 | 121 | 142 | 7510 | 0.9807527 | 1 | 1 |
| mmu0454    | Cell adhesion molecules (CAMs) | 21 | 121 | 164 | 7510 | 0.8574849 | 1 | 1 |
| mmu0515    | Influenza A | 21 | 121 | 171 | 7510 | 0.927303 | 1 | 1 |
| mmu0515    | Epstein-Barr virus infection | 21 | 121 | 217 | 7510 | 0.6463657 | 1 | 1 |
| mmu0420    | Adipogenesis | 20 | 121 | 74 | 7510 | 1.810642 | 0.010298 | 0.7313571 |
| mmu0520    | Morphine addiction | 20 | 121 | 83 | 7510 | 1.4487259 | 0.0359613 | 1 |
| mmu0470    | Phosphatidylinositol signaling system | 20 | 121 | 96 | 7510 | 1.3570322 | 0.1794444 | 1 |
| mmu0472    | Pancreatic secretion | 20 | 121 | 101 | 7510 | 1.329655 | 0.1756367 | 1 |
| mmu0470    | Retinoic acid signaling pathway | 20 | 121 | 102 | 7510 | 1.3186031 | 0.1096189 | 1 |
| mmu0460    | HIF-1 signaling pathway | 20 | 121 | 110 | 7510 | 1.2389085 | 0.3109193 | 1 |
| mmu0491    | Insulin resistance | 20 | 121 | 111 | 7510 | 1.2678047 | 0.330999 | 1 |
| mmu0515    | Glioblastoma multiforme | 20 | 121 | 112 | 7510 | 1.1563171 | 0.3561545 | 1 |
| mmu0472    | Glioblastoma multiforme | 20 | 121 | 120 | 7510 | 1.0150969 | 0.832695 | 1 |
| ID | Oligo Name | Oligo sequence | Product size |
|----|------------|----------------|--------------|
|    |            |                |              |
| **RT-PCR** |          |                |              |
| Platr10  | SJ229     | CACTGCTGTTTTGGAGCTCCAT | 121bp         |
|         | SJ230     | TGGGACAGTCTCTGGATGGCCT |             |
| Pou5f1   | J648      | CAGAGGATGCTGAGTGGGCTGTA | 142bp         |
|         | J649      | TCAACCCTCAAGGTCCTCCAC |             |
| Sox2     | J652      | GAAGTTTTGAGCCCGAGGCTTAAG | 164bp         |
|         | J653      | TGCACGCGCCCTGCGGAGATCTG |             |
| Nanog    | J650      | GCTTAGGGGGCATTCTCTGATCTA | 133bp        |
|         | J651      | ACCTGAAACCTCCACTAGAGT |             |
| Peln4    | JH4902    | TGGCAGCCTCTGGATGGGCA | 122bp        |
|         | JH4093    | CTTCCTCTGACTTGGTAACATC |             |
| Actb     | J880      | CAGGTCATACCATGGCAATGAGC | 135bp        |
|         | J881      | CGGATGTCCACGTCAACATGAGT |             |
| U2       | JH1055    | ATCTGTTCCTTACGTTTAATATCTG | 151bp      |
|         | JH1056    | GGGTGCACCCTCTGGAGGTA |             |
|         |            |                |              |
| **shRNAs** |          |                |              |
| Platr10  |            | TTCTGTGTATCTGTGGAGCCAGGCA |         |
|         | 1         | CCTGCTGCTCTGTAATCAATCCAAATGTA |         |
|         | 2         | CTGCCAGCATCTGACTAAGATAGAT |         |
|         | 3         | TAGATGCTCTCTCCAGGAGGTAG |         |
|         | 4         |             |              |
| Snhg14   |            | TTCAAAACAATTTGTGAAGAATGTAA |         |
|         | 1         | TCTCAGTATGTCAATAGATGAAG |         |
|         | 2         |             |              |
| Control (shCT) |  | GCAGCAACCTGGACACGTGATCTTAA |     |
|         |            |                |              |
| **CRIST Cas9 gRNA** |          |                |              |
| Sox2     | pSox2-1   | GGGGGTGAGGACACGTGCTG |             |
|         | pSox2-2   | GAGGCAATTCGCCCTCATCAGCAT |         |
|         | 5'-gCT    | GAGAGGTACAATGGTCACTC |             |
| Pou5f1   | pPou5f1-1 | GAACATTCAATGGATGTTC |             |
### pPou5f1-2

| 5'-gCT | GTGTGAGGGGATTTGGGCTC |
|--------|----------------------|
|        | GAAGTGAGGATGATCCTCTGA |

### FLI1

| pFL11-1 | GATGAGTGGGTAGCCGCTC |
|---------|---------------------|
| pFL11-2 | GTGGACCGCGTATTGCAAG |
| 5'-gCT  | GTATGACTGGTGTCCTTTATA |

### IGF2

| pIGF2-1 | GCCTTGGCTTTCCCCAAAATT |
|---------|-----------------------|
| pIGF2-2 | GTCGCCCGGCTTCCAGATAAG |
| 5'-gCT  | GTTCTACGGGTGTATGTCAA |

### Control (gCT)

| GAAGTGAGGATGATCCTCTGA |

### LncRNA PCR

| LncRNA | Accession | Sequence                                                                 | Length |
|--------|-----------|--------------------------------------------------------------------------|--------|
| **Spilr32** | JH4117 | GAGTAGTGCATTAACTAACGGA | 112bp |
|        | JH4118 | CCACCTCTAGTTTTCAGAATC | |
| **Spilr33** | JH4119 | GAGATGTTGCTAAACCAGG | 102bp |
|        | JH4120 | GAGGCACCTGGAGGACGATG | |
| **Spilr22** | JH4096 | CTCAGGCTTAAGCTCCAAGGC | 126bp |
|        | JH4097 | CTTGATGCTGCCCTTCCAAGC | |
| **Spilr5** | JH4016 | CCGATTGCTGCTGTTTACTT | 118bp |
|        | JH4017 | CCAGGCTCAGGTTAGCTCCAAC | |
| **Spilr7** | JH4020 | TGGGCCTTCAGACTCTATCCA | 113bp |
|        | JH4021 | GCCTCAGGTCGCCATCCAAC | |
| **Spilr8** | JH4082 | CTCGGAGAGGCTCAGCCACATG | 119bp |
|        | JH4083 | GAGCTGTCAGCAACATGACATGTC | |
| **Spilr9** | JH4027 | CACTGCTGGTTTGAGGCTCCCATG | 122bp |
|        | JH4028 | GTGGGACAGTCTGGATGCGCT | |
| **Snhg14** | JH5188 | CTGCACTGCAAGAGTAAGTGC | 159bp |
|        | JH5189 | TGGCCTGTGAACTCAGGCTTA | |
| **Snurf** | JH5190 | TCAATTGCACTCTGCAAGGCT | 208bp |
|        | JH5191 | GGCACGAGCAATGCTGTA | |
| **Actb** | J880 | CAGGTCTACACCAGTAGGATGAAC | 135bp |
|        | J881 | CGGATGTCACGTCAGCTCATGA | |

### Sox2

| 5-Ct | JH4405 | CATAGTAGCTCCCCAATCAAAATGG | 135bp |
| JH4406 | AATGCAAGGGATTTGAGCTGCACG |
| pSox2 | JH4373 | GAGGCAATCCCGTACGATG | 136bp |
| JH4374 | CGCTGGGAAATCTTTGTTCATC |
| Off target | JH5877 | AGCCACCTGCTGCTCCTCGCTG | 128bp |
|        | JH5878 | CTGCACGAGATGCAGGGAT | |
A. CRIST-seq

Fixation of Cas9-gRNA cells

Cas9-promoter IncRNA complex

Nuclear RNA in situ reverse transcription

Cas9-FLAG immunoprecipitation

Biotinylated cDNA bead purification

Quantitative PCR

Fold enrichment

cDNA library sequencing

Promoter IncRNA interactome

Figure S1. CRIST-seq assay to map the promoter IncRNA interactome.

Chromatin-RNA in situ reverse transcription trap sequencing (CRIST-Seq) assay. Cells were transfected with CRISPR dCas9 gRNA to target the gene promoter. The Cas9-gRNA expressing cells were crosslinked by formaldehyde to fix the promoter RNA chromatin structure. After cell membrane lysis, the nuclei were isolated and the promoter-interacting RNAs were in situ reverse transcribed into cDNAs with biotin-dCTP. The promoter biotin-cDNA chromatin complex was first immunoprecipitated by a Cas9-FLAG antibody, and the promoter-interacting biotin-cDNAs were separated from genomic DNAs by streptavidin beads. The CRIST-captured chromatin cDNAs were aliquoted for Illumina library sequencing or quantitative PCR to measure the enrichment of the identified IncRNAs to the target gene promoter.
**A. Sox2 CRIST targeting vector**

Sox2 Cas9-gRNA vector used in CRIST assay. Cas9: CRISPR Cas9; gRNA1, 2: two Cas9 guiding RNAs that target the Sox2 promoter (sequences under the diagram); pEF1: the human EEF1A1 promoter; pU6: U6 promoter; pH1: human H1 promoter; T5: the TTTTT termination signal of RNA polymerase III.

**B. CRIST targeting gRNAs**

Location of the two Cas9 gRNAs in the Sox2 promoter. The Sox2 exon 1 mRNA is shown in blue and the coding region with ATG in green. Two Cas9 gRNAs are highlighted in yellow and PAM sequences in red in the Sox2 promoter region. TS+1: transcription initiation site.

**Figure S2. Location of Sox2 Cas9-gRNAs in the CRIST assay.**

A. The CRISPR Cas9 Sox2-gRNA vector used in CRIST assay. Cas9: CRISPR Cas9; gRNA1, 2: two Cas9 guiding RNAs that target the Sox2 promoter (sequences under the diagram); pEF1: the human EEF1A1 promoter; pU6: U6 promoter; pH1: human H1 promoter; T5: the TTTTT termination signal of RNA polymerase III.

B. Location of the two Cas9 gRNAs in the Sox2 promoter. The Sox2 exon 1 mRNA is shown in blue and the coding region with ATG in green. Two Cas9 gRNAs are highlighted in yellow and PAM sequences in red in the Sox2 promoter region. TS+1: transcription initiation site.

C.
A. *Pou5f1* CRIST targeting vector

B. Specific CRIST targeting of *Pou5f1*

---

**Figure S3. Specific targeting of the mouse *Pou5f1* promoter in the CRIST assay.**

A. The CRISPR Cas9 Pou5f1-gRNA vector. gRNA1, 2: two Cas9 guiding RNAs that target the *Pou5f1* promoter (sequences under the diagram).

B. Specific CRIST targeting of the mouse *Pou5f1* promoter. *pPou5f1*: the targeting site in the *Pou5f1* promoter where the Cas9 gRNAs are designed; 5′-Ct: the *Pou5f1* control site that is 13.9 kb away from the *pPou5f1* target site. Cas9 Vector: cells that were treated with the Cas9 control vector that lacks the gRNAs; Cas9-gRNA: cells that were targeted by both Cas9 and *Pou5f1* gRNAs; Cas9-gCT: cells that were treated with the random control gRNA vector. Off-target: a CRIST control site that is 33.8 kb upstream of the housekeeping gene GAPDH. The chromatin complex was immunoprecipitated with a Cas9-FLAG antibody and an IgG control antibody. CRIST signals were quantitated by real-time PCR using specific primers derived from the pOct4 targeting site, 5′-Ct control site and off-target site. All data shown are mean ±SEM from three independent experiments by normalization over the IgG control. **p<0.01 as compared with Vector and gCT controls.**
A. *IGF2* CRIST targeting vector

![Diagram of CRIST targeting vector]

B. Specific CRIST targeting of *IGF2*

![Graph showing IGF2 binding]

**Figure S4. Specific CRIST targeting of the human IGF2 imprinting promoters.**

A. The CRISPR Cas9 *IGF2*-gRNA vector. gRNA1, 2: two Cas9 guiding RNAs that target the IGF2 imprinting promoters (P2-P4, sequences under the diagram).

B. Specific CRIST targeting of the growth factor IGF2 promoter. The human IGF2 has four promoters and nine exons. The promoter P1 is not imprinted and it drives the biallelic expression of the growth factor from exons 1, 2, 3, 7, 8, and 9. In contrast, its promoters P2-P4 are imprinted and are expressed exclusively from the paternal allele. In human tumors, however, this imprinting mechanism is dysregulated and causes biallelic expression of the mitogen that promotes tumor growth. We used the CRIST assay to examine the imprinted P2 promoter. pIGF2: the site in the IGF2 promoter where the Cas9 gRNAs are designed; 5'-Ct: IGF2 control site that is 106 kb away from the pIGF2 target site. Vector: cells that were treated with the Cas9 control vector that lacks the gRNAs; Cas9-gRNA: cells that were targeted by both Cas9 and IGF2 gRNAs; Cas9-gCT: cells that were treated with the random control gRNA vector. Off-target: a CRIST control site that is 33.8 kb upstream of the housekeeping gene GAPDH. The chromatin complex was immunoprecipitated with a Cas9-FLAG antibody and an IgG control antibody. CRIST signals were quantitated by real-time PCR using specific primers derived from the pIGF2 targeting site, 5'-Ct control site and off-target site. All data shown are mean±SEM from three independent experiments by normalization over the IgG control. ** p<0.01 as compared with Vector and gCT controls.
A. **FLI1** CRIST targeting vector

![Diagram of CRISPR Cas9 FLI1-gRNA vector](image)

B. Specific CRIST targeting of **FLI1**

![Graph showing FLI1 binding](image)

**Figure S5. Specific CRIST targeting of the human **FLI1** promoter.**

A. The CRISPR Cas9 FLI1-gRNA vector. gRNA1, 2: two Cas9 guiding RNAs that target the FLI1 promoter (sequences under the diagram).

B. Specific CRIST targeting of the oncogenic FLI1 promoter. pFLI1: the site in the FLI1 promoter where the Cas9 gRNAs are designed; 5'-Ct: FLI1 control site that is 18.5 kb away from the pFLI1 targeting site. Vector: cells that were treated with the Cas9 control vector that lacks the gRNAs; Cas9-gRNA: cells that were targeted by both Cas9 and FLI1 gRNAs; Cas9-gCT: cells that were treated with the random control gRNA vector. Off-target: a CRIST control site that is 33.8 kb upstream of the housekeeping gene GAPDH. The chromatin complex was immunoprecipitated with a Cas9-FLAG antibody and an IgG control antibody. CRIST signals were quantitated by real-time PCR using specific primers derived from the pFLI1 targeting site, 5'-Ct control site and off-target site. All data shown are mean±SEM from three independent experiments by normalization over the IgG control. ** p<0.01 as compared with Vector and gCT controls.
**Figure S6. The FECR1 circRNA-FLI1 promoter interaction as a positive CRIST control.**

A. The FECR1-FLI1 CRIST vector. FECR1 is a known circRNA that binds to the FLI1 promoter and regulates its activity in cis. We thus used it as a CRIST positive control. Two Cas9 FLI1 gRNAs are transcribed by U6 and H1 promoters, respectively, and guides the dCas9 to the FLI1 promoter. The FLI1 CRIST-seq library was used to quantitate the enrichment of FECR1 circRNA in the FLI1 promoter.

B. Enrichment of FECR1 in the FLI1 promoter. Vector: cells that were treated with the Cas9 control vector that lacks the gRNAs; Cas9-gRNA: cells that were targeted by both Cas9 and IGF2 gRNAs; Cas9-gCT: cells that were treated with the random control gRNA vector; Vector: cells that were treated with the empty dCas9 vector. The enrichment of FECR1 in the FLI1 promoter was quantitated by qPCR. For comparison, the value of the IgG group was set as 1. ** p<0.01 as compared with the IgG control group.

C. The nuclear lncRNA MALAT1 is used as the FLI1 CRIST negative control. The binding of MALAT1 to the FLI1 promoter was quantitated by qPCR and was standardized over the IgG group.
Figure S7. The top 30 GO enriched target genes of the Sox2-interacting RNA.
The Go enrichment was analyzed using the ggplot2 package (http://had.co.nz/ggplot2/). The top 30 enriched target genes are selected based on the enrichment score and q-value.
Figure S8. The top 30 KEGG pathways of the Sox2-interacting RNAs.
The scatterplot of the top 30 enriched KEGG pathways. KEGG: Kyoto Encyclopedia of Genes and Genomes; rich factor: the ratio of the number of DEGs and the number of all the unigenes in the pathways; Q-value: the corrected p-value.
A. Differentially expressed Sox2 CRIST IncRNAs

| Gene_id         | Gene_name  | Locus                  | Fold  | p-value       | q-value       |
|-----------------|------------|------------------------|-------|---------------|---------------|
| ENSMUSG00000073291 | Gm10491    | X:7899356-7908351      | 14300200.0 | 1.22681E-38 | 1.43657E-36   |
| ENSMUSG00000102064 | Gm28625    | 9:8489054-84890749     | 13446000.0 | 1.98227E-36 | 2.15038E-34   |
| ENSMUSG00000065999 | Gm13154    | 4:147553276-147585198  | 1089.6  | 1.19663E-07 | 2.4399E-06    |
| ENSMUSG00000099370 | Platricular | 13:75647445-75655731   | 987.3   | 3.65397E-20 | 2.1904E-18    |
| ENSMUSG00000031535 | S10a3      | 6:12272780-122801640   | 350.0   | 8.5813E-107 | 3.1165E-104   |
| ENSMUSG00000031995 | St14       | 9:31089401-31131853    | 181.8   | 1.00398766 | 0.032151481   |
| ENSMUSG00000106628 | Gm43558    | 13:64357291-134357501  | 108.9   | 0             | 0             |
| ENSMUSG00000084899 | Elav2      | 6:91230762-91400785    | 41.9    | 1.5543E-06  | 2.66539E-05   |
| ENSMUSG00000105270 | Gm42638    | 5:14550406-145507054   | 31.1    | 2.65617E-11 | 8.4575E-10    |
| ENSMUSG00000087306 | A230004M16Rik | 13:134357291-134357501 | 108.9  | 0             | 0             |
| ENSMUSG00000097695 | Gm26905    | 5:14550406-145507054   | 31.1    | 2.65617E-11 | 8.4575E-10    |
| ENSMUSG00000023140 | Cbx7       | 15:79891658-79971119   | 19.7    | 1.57911E-12 | 5.69364E-11   |
| ENSMUSG00000087267 | 4933427J07Rik | 2:128955670-128957861  | 14.2    | 1.11677E-05 | 3.30449E-09   |
| ENSMUSG00000165144 | Zfp229     | 17:21730794-21769342   | 11.8    | 4.14055E-05 | 0.00054497    |
| ENSMUSG00000097695 | Gm26905    | 5:14550406-145507054   | 31.1    | 2.65617E-11 | 8.4575E-10    |
| ENSMUSG00000023140 | Cbx7       | 15:79891658-79971119   | 19.7    | 1.57911E-12 | 5.69364E-11   |
| ENSMUSG00000087267 | 4933427J07Rik | 2:128955670-128957861  | 14.2    | 1.11677E-05 | 3.30449E-09   |

Figure S9. Differentially expressed Sox2 CRIST IncRNAs.
The Sox2 CRIST-seq RNAs were integrated with the RNA-seq >2 fold RNAs data using a VENN program (http://bioinformatics.psb.ugent.be/webtools/Venn). A cut-off threshold of peak enrichment FPKM >50 was arbitrarily set to select CRIST-Seq RNAs. The RNAs are listed in the order of expression fold change of iPSCs over fibroblasts.
Figure S10. Detection of the Sox2 pre-mRNA in the CRIST-seq library.
The abundance of the Sox2 pre-mRNA in the Cas9 Sox2-gRNA CRIST-seq library was detected by two primer sets from the Sox2 coding sequence. The third pair of PCR primers (JH4813/JH4815) from the downstream sequence was used as the negative control. The iPSC cDNA was used as the positive PCR control. Using this more sensitive PCR, we detected the presence of Sox2 pre-mRNA in the CRIST-seq products. As expected, no signals were detected in the Cas9 gCT and the Vector control groups.
A. *Snhg14* RNA-seq data

| Gene_id | Gene_name | Locus          | PSC_FPKM | FIB_FPKM | Fold  | p-value     | q-value         |
|---------|-----------|----------------|----------|----------|-------|-------------|-----------------|
| ENSMUSG00000100826 | Snhg14 | 7:59384596-59411173 | 42.22  | 9.18    | 4.598299 | 2.34342E-05 | 0.000326424 |

B. *Snhg14* IGV

C. *Snhg14* sequence (NR_015456)

```
1  atttgtgatt   tcaatattata   ttctggttta   caaccttataa   agatagtatt   tcaatttataa
61  ttccagtttta   ttctgcaacct   caatgt tatagttq   actaaatccaa   atttggtaatt   aaaacagcc
121  ctgtatatttt   ttctaaatgta   taaatagttt   ttaaatagttt   ttaaatagttt   ttaaatagttt   aaaacagcc
181  tatttttctta   ttaatttttt   taattttttta   taattttttta   taattttttta   taattttttta   aaaacagcc
241  ttgttctacttta   ttaaaacagta   ttaaaatagttt   ttaaatagttt   ttaaatagttt   ttaaatagttt   aaaacagcc
301  tattttttttctta   ttaattttttta   taatttttttt   taatttttttt   taatttttttt   taatttttttt   aaaacagcc
361  ggcccatatatttt    ttaaaacagta   ttaaaatagttt   ttaaatagttt   ttaaatagttt   ttaaatagttt   aaaacagcc
421  aaccccaaaatttt   taattttttta   taattttttta   taattttttta   taattttttta   taattttttta   aaaacagcc
481  tattttttttttt   tttaatttttt  
```

**Figure S11. Differential expression of Snhg14 lncRNA in reprogramming.**

A. RNA-seq expression of Snhg14 between iPSCs and fibroblasts collected in reprogramming.

B. IGV Sashimi plot of Snhg14 lncRNA between iPSCs and fibroblasts.

C. Snhg14 lncRNA sequence.
A. *Snhg14* RNA-DNA FISH

Snhg14 RNA-DNA FISH. The isolated iPSC clones were subjected to RNA FISH (green) using biotin-14-dCTP-labeled single strand DNA probes for Snhg14. RNA probes are designed to cover the intron region for specific staining of the mature RNAs. Slides were subsequently re-fixed and processed for DNA FISH (red) using digoxigenin-dUTP labeled BAC probe. The nuclear DNA was stained with DAPI (blue). Arrows mark the co-localization of the Snhg14 lncRNA signal with the Pou5f1 DNA signal.

B. *Palr35* RNA-DNA FISH

Palr35 RNA-DNA FISH. Although lncRNA *Palr35* was also differentially expressed in reprogramming, it was not in the list of the Sox2 CRIST-seq. We thus used it as the RNA FISH control. As the control, RNA-FISH did not detect the co-localization of Palr35 with Sox2.

Figure S12. Interaction of lncRNAs with the *Sox2* locus by RNA-DNA FISH.

A. *Snhg14* RNA-DNA FISH. The isolated iPSC clones were subjected to RNA FISH (green) using biotin-14-dCTP-labeled single strand DNA probes for Snhg14. RNA probes are designed to cover the intron region for specific staining of the mature RNAs. Slides were subsequently re-fixed and processed for DNA FISH (red) using digoxigenin-dUTP labeled BAC probe. The nuclear DNA was stained with DAPI (blue). Arrows mark the co-localization of the Snhg14 lncRNA signal with the Pou5f1 DNA signal.

B. *Palr35* RNA-DNA FISH. Although lncRNA *Palr35* was also differentially expressed in reprogramming, it was not in the list of the Sox2 CRIST-seq. We thus used it as the RNA FISH control. As the control, RNA-FISH did not detect the co-localization of Palr35 with Sox2.
A. RAT assay

iPSCs → Cell fixation →Cell membrane lysis → Genomic DNA → Strand-specific IncRNA labeling with biotin-14-dCTP → Biotin DNA pulldown by beads → Sox2 binding → q-PCR

B. Location of q-PCR primers

5'-Ct → Sox2 promoter → Sox2 → 3'-Ct

A B C D E F

C. Snhg14-Sox2 interaction

Enrichment

Snhg14

IncCT

Figure S13. Validation of the Snhg14-Sox2 interaction by the RAT assay.

A. The RAT assay. iPSCs were fixed and Snhg14 IncRNA was labelled by biotin-dCTP in the nucleus using four Snhg14-specific oligonucleotides. The random oligonucleotides were used as the RAT control. After sonication, the biotin-Snhg14-chromatin complex was pulled down by streptavidin beads. The chromatin DNA was purified for qPCR.

B. Location of qPCR primers. Primers B and C are designed from the Sox2 promoter.

C. The Snhg14-Sox2 interaction by qPCR. The Snhg14-pulled down chromatin DNAs were mapped by a series of primers in the Sox2 locus. Note the enriched signal of the Sox2 promoter (B, C sites) in the Snhg14-pulled down chromatin complex.