Supplementary Material

for

"A New Class of Temporarily Phenotypic Enhancers Identified by CRISPR/Cas9 Mediated Genetic Screening "

Yarui Diao, Bin Li, Zhipeng Meng, Inkyung Jung, Ah Young Lee, Jesse Dixon, Lenka Maliskova, Kun-liang Guan, Yin Shen, and Bing Ren

Table S3. Data analysis of CRISPR/Cas9 screening experiments. We first identified sgRNA candidates that are enriched in eGFP- population by meeting the criteria of (a) each sgRNA should have a sequencing coverage of 50 reads and (b) each sgRNA should have at least two fold enrichment in eGFP- population. We then generated the high

confident elements that control *POU5F1* expression if they are at least supported by the enrichment two sgRNAs targeting the same elements in three different experiments.

1. Supplementary Methods

Array oligo synthesis and pooled library cloning

The oligo pool were amplified with primers with 15 bp overlap with BbsI cutting site in LentiCRISPR plasmid:

F: AAAGGACGAAACACC

R: TTCTAGCTCTAAAAC

The PCR product was size selected and gel-purified with NucleoSpin Gel and PCR Clean-Up Kit (Clontech, Cat# 740609), followed by In-fusion cloning (Clontech, Cat# 638909). The cloning product was electro-transformed into 5-alpha Electrocompetent E. coli (NEB, Cat#C2989K) and amplified on Agar plates. About 1 million of independent bacterial colonies growing on Agar plates were collected and subjected to plasmids extraction with EndoFree Plasmid Mega Kit (Qiagen, Cat#12381).

Lentiviral particle production, concentration, and titration.

The lentiCRIPSR library was prepared as previously described (Meng et al. 2013) with minor modifications. Briefly, 5ug of lentiCRISPR plasmid library was co-transfected with 4 ug PsPAX2 and 1 ug pMD2.G (Addgene #12260 and #12259) into a 10-cm dish of HEK293T cells in DMEM (Life Technologies) containing 10% FBS (Life Technologies) by PolyJet transfection reagents (Signagen, Cat# SL100688). Medium was changed 6 hours after transfection. The supernatant of cell culture media was harvested at 24 hours and 48 hours after transfection, and filtered by Millex-HV 0.45 μm PVDF filters (Millipore, Cat# SLHV033RS). The viruses were further concentrated with 100, 000 NMWL Amicon Ultra-15 Centrifugal Filter Units (Amicon, Cat#UFC910008).

For viral titration, 0.5 million hESC POU5F1-eGFP cells were seeded per well on 6-well plate. 12 hours later, different amount (1ul, 2ul, 4ul, 8ul) of concentrated viral-containing media were added to the cell culture media to infect the hESC following the same

protocol described in the lentiviral screening section. The same amount of non-infected cells were seeded and not treated with puromycin as the control. 24 hours post-infection, the viral infected cells were treated with 250ng/ml Puromycin (Life Technologies, Cat#A1113802) for another 72 hours. We counted the number of Puromycin resistant cells and the control cells to calculate the ration of infected cells, and then viral titer. In the screening, 12 million POU5F1-eGFP hESCs were used in each independent screening and infected with viral particles at low MOI (0.02) to make sure each infected cell get one viral particle. The library coverage in each screen is more than 120 times.

Lentiviral CRISPR/Cas9 screening

To perform sgRNA screening, 12 million H1-eGFP cells were seeded on matrigel-coated petri dish overnight. Low titer of lentiviral particles (MOI 1-2 for experiment 1 and 2, and MOI 0.02 for experiment 3 and 4) were added into the media, supplemented with 8μg/ml polybrene. Immediately after viral infection, the plates was centrifuged at 2,000 rpm for 2 h at 37°C. The viral containing media was then replaced with fresh E8 media without polybrene. 24 hours after viral transduction, the cells were cultured in E8 media with 250ng/ml puromycin for 7 days, followed by another 3-day-culture in regular E8 media without puromycin before FACS sorting. 4 million of non-sorted cells and 1 million of FACS sorted GFP-low cells were collected for the further analysis. The genomic DNA was extracted, followed by two rounds of PCR amplification to prepare DNA library ready for deep sequencing. The sgRNA integrated into the genomic DNA was amplified with F1/R1 primers for 25 cycles, and F2/R2 (R2 index varies) primers for 10 cycles in the 1st step and the 2nd step PCR, respectively. The final PCR products were purified and sequenced.

F1: AATGGACTATCATATGCTTACCGTAACTTGAAAGTATTTCG
R1: CTTTAGTTTGTATGTCTGTTGCTATTATGTCTACTATTCTTTCC
F2:AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTT
CCGATCTtcttgtggaaaggacgaaacaccg

R2_index1:

CAAGCAGAAGACGGCATACGAGAT**CGTGAT**GTGACTGGAGTTCAGACGTGT

GCTCTTCCGATCtctactattctttcccctgcactgt

R2_index2:

CAAGCAGAAGACGGCATACGAGAT**ACATCG**GTGACTGGAGTTCAGACGTGT

GCTCTTCCGATCtctactattctttcccctgcactgt

R2_index3:

CAAGCAGAAGACGGCATACGAGAT**GCCTAA**GTGACTGGAGTTCAGACGTGT

GCTCTTCCGATCtctactattctttcccctgcactgt

R2_index4:

CAAGCAGAAGACGGCATACGAGAT**TGGTCA**GTGACTGGAGTTCAGACGTGT

GCTCTTCCGATCtctactattctttcccctgcactgt

R2_index5:

CAAGCAGAAGACGGCATACGAGAT**CACTGT**GTGACTGGAGTTCAGACGTGT

GCTCTTCCGATCtctactattctttcccctgcactgt

R2_index6:

CAAGCAGAAGACGGCATACGAGAT**ATTGGC**GTGACTGGAGTTCAGACGTGT

GCTCTTCCGATCtctactattctttcccctgcactgt

Transfection and luciferase reporter assay

Luciferase assays were conducted as previously described (Heintzman et al. 2007). Detailed information can be found be Supplemental Materials. Briefly, the DHS_37, DHS_65, DHS_108, and DHS_115 elements were PCR amplified from H1 hESC genomic DNA, and cloned into pGL3-promoter vector (Promega) by In-fusion cloning. The sequences of the primers are:

DHS_37 _F GCATGAGCCACAGGAGGTAG

DHS_37 _R CGCTTTCTCTCCCTCAACC

DHS_65 _F GAGGCAGCATCTAACCTTGC

DHS_65 _R TCCTTACCATGTGGCATTTG

DHS_108 _F GAATTCCGAAGGAGGGGTAG
DHS_108 _R CGTGCAATACGAACACATCA

DHS_115 _F GATGCTAGGGAATTCGATCCCCT
DHS_115 _R ATCCGAGCTCTGCAGGATTTGCT

CTCF_38_F CCAAGCCTCCTGCTTCTGG
CTCF_38_R TCTTACTTCCCACACCCTGAGC

To test the enhancer activity of these elements with native *POU5F1* promoter and the proximal enhancer (DHS_115), the 360bp *POU5F1* minimal promoter (Chia et al. 2010) (hg18 Chr 6:31,246,377-31,246,736) was synthesized (IDT gblock) to replace the SV-40 promoter in pGL3-promoter vector.

After validation of sequence by Sanger sequencing, the constructs were co-transfected with pRL-SV40 Renilla reporter vector in H1 hESCs with Fugene HD (Roche) at a 4:1 reagent to DNA ratio. Transfected cells were cultured for an additional 2 days prior to harvest for reporter assay. The Dual-Luciferase Reporter Assay kit (Promega Cat#:E1960) was used according to manufacturer's protocol. The adjusted firefly luciferase activity of each sample was normalized to the average of active of 3 negative control regions.

RT-qPCR

Total RNA was isolated using TRIzol (Thermo Fisher Scientific, Cat# 15596026) as per manufacturer's instruction, and 1ug of RNA was reverse transcribed into cDNA by SuperScript® III First-Strand Synthesis System (Thermo Fisher Scientific, Cat# 18080051). The gene expression levels was quantified by real time PCR by KAPA SYBR® FAST qPCR Kits (KAPA biosystems, Cat# KK4600) on LightCycler 480 System (Roche). The following primers were used.

POU5F1_F	CGAAAGAGAAAGCGAACCAGTATCGAGAAC
POU5F1_R	CGTTGTGCATAGTCGCTGCTTGATCGC
SOX2_F	ATGCACCGCTACGACGTGA
SOX2_R	CTTTTGCACCCCTCCCATTT
NanoG_F	TGATTTGTGGGCCTGAAGAAA
NanoG_R	GAGGCATCTCAGCAGAAGACA
AFP_F	TGGGACCCGAACTTTCCA
AFP_R	GGCCACATCCAGGACTAGTTTC
T brachyury transcription factor_F	TGCTTCCCTGAGACCCAGTT
T brachyury transcription factor_R	GATCACTTCTTTCCTTTGCATCAAG
Sox1_F	GTACAGCCCCATCTCCAACTC
Sox1_R	GGCTCCGACTTCACCAGAGAG
18sRNA F	GGCCCTGTAATTGGAATGAGTC
18sRNA R	CCAAGATCCAACTACGAGCTT

Embryoid body formation and differentiation

Embryoid bodies (EB) were generated with AggreWell 400 (Stemcell Technologies, Cat# 27845) following the manufacture's protocol. Briefly, hESC was dissociated into single cell and suspended into AggreWell™ Medium (Stemcell technologies, Cat# 05893) supplemented with ROCK inhibitor Y-27632 (Stemcell technologies, Cat# 72304). About 2.4 million cells were plated into each well to let each EB contain about 2000 cells. After 24 hours, the media was switched to STEMdiff™ APEL™ Medium (Stemcell technologies, Cat# 05210) to induce differentiation. The EB was maintained in STEMdiff APEL medium for 14 days and fresh media was changed on a daily basis.

ChIP-seq, 4C-seq experiments and data analysis

The ChIP-seq and 4C-seq experiments and data analysis were performed following the the procedures described earlier (van de Werken et al. 2012). The 4 base cutter used for 4C-seq experiments are DpnII and NlaIII. The following primers including Illumina's TruSeq adaptors are used to amplified the target sequence interacting with bait region (*POU5F1* promoter, hg18 Chr 6:31,247,202-31,248,137)

Pou5f1_4C_F:

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTTGCCTAGGATTCTGGATGGA

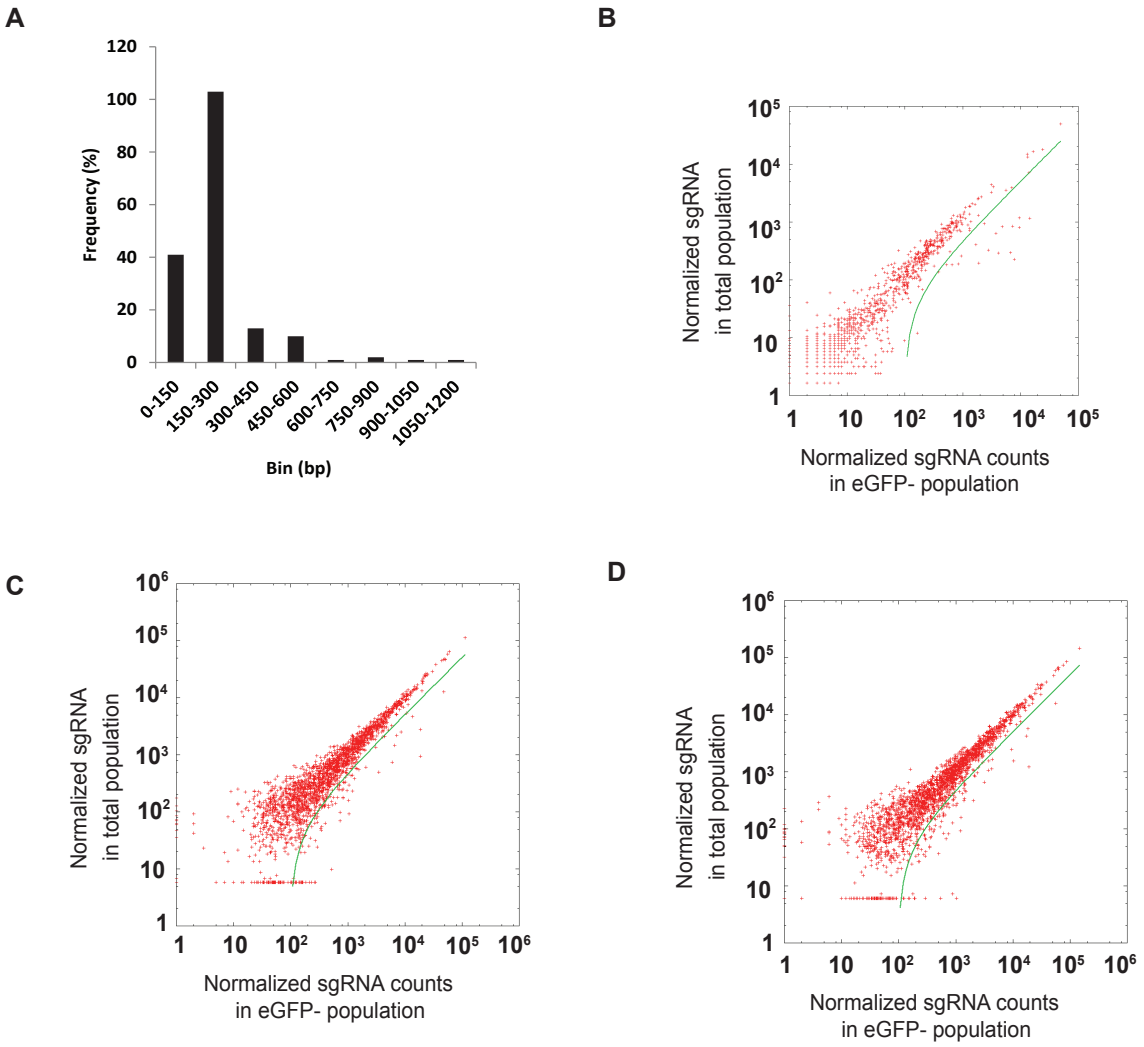
Pou5f1_4C_R21:

CAAGCAGAAGACGGCATACGAGATTCCGAAACGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGCAGAGGAGGTGGAGAGTGA

Pou5f1_4C_R20:

CAAGCAGAAGACGGCATACGAGATAAGGCCACGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGCAGAGGAGGTGGAGAGTGA

Supplemental Fig. S1

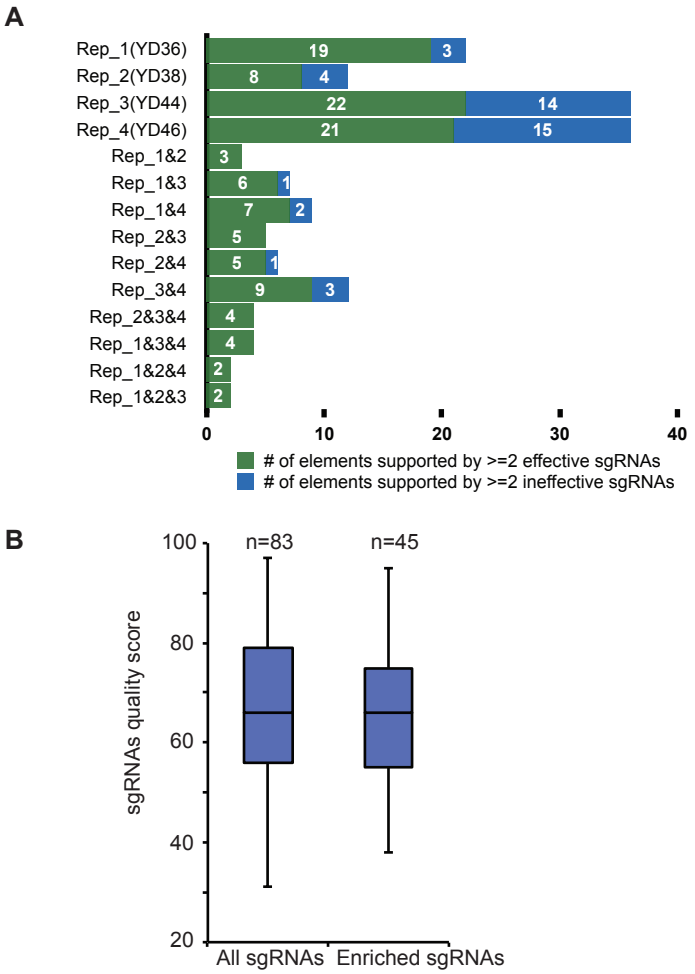


SUPPLEMENTAL Fig. S1. Experimental design

(A) Size Distributions of candidate regulatory regions surveyed in this study.

(B-D) H1 POU5F1-eGFP cells were infected by lentiCRISPR library, selected under puromycin for 7 days, and cultured without puromycin for another 10-days. Genomic DNA was collected from the non-sorted control cells and FACS sorted eGFP negative cells, followed by PCR amplification of sgRNA sequence and deep sequencing. Scatter plots for sgRNA read counts in eGFP- cells compared to the control cells for experiment 2, 3, and 4 after LOESS normalization. Green line: 2-fold cutoff.

Supplemental Fig. S2

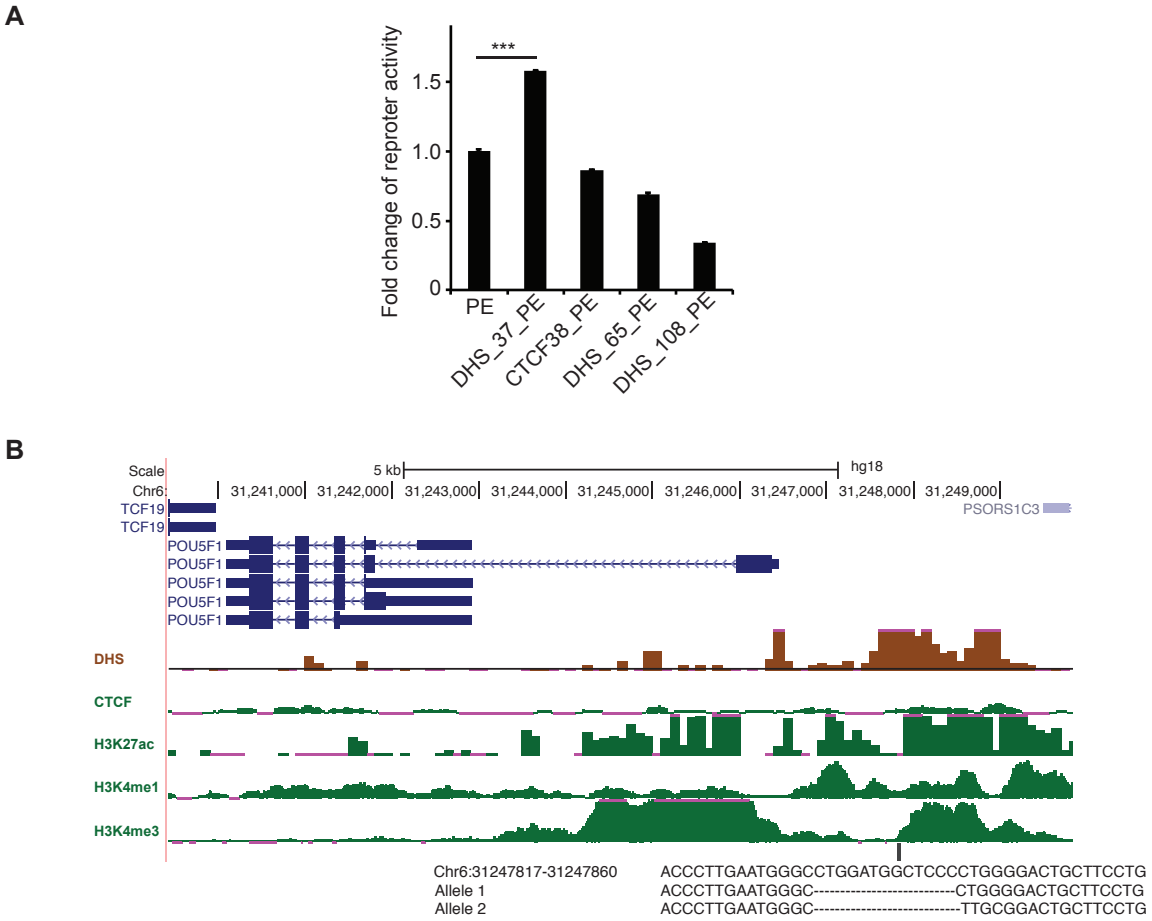


SUPPLEMENTAL Fig. S2. The number and quality score of positive hit called from screenings.

(A) The element was called positive in each replicate when hit by at least two significant sgRNAs. The bar graph shows the number of true (green) and false (blue) elements identified from each experiment and how these positive elements overlapped between replicates.

(B) Box plot shows the distribution of quality scores of all sgRNAs targeting the six identified (left) compared to the ones enriched in the GFP- cells (right). We obtained quality scores of a total of 83 detectable sgRNAs from http://crispr.mit.edu (>50 reads in at least one experiment in either the control or eGFP- enriched sequencing libraries). The sgRNA enriched in the eGFP- cells are plotted on the right.

Supplemental Fig. S3

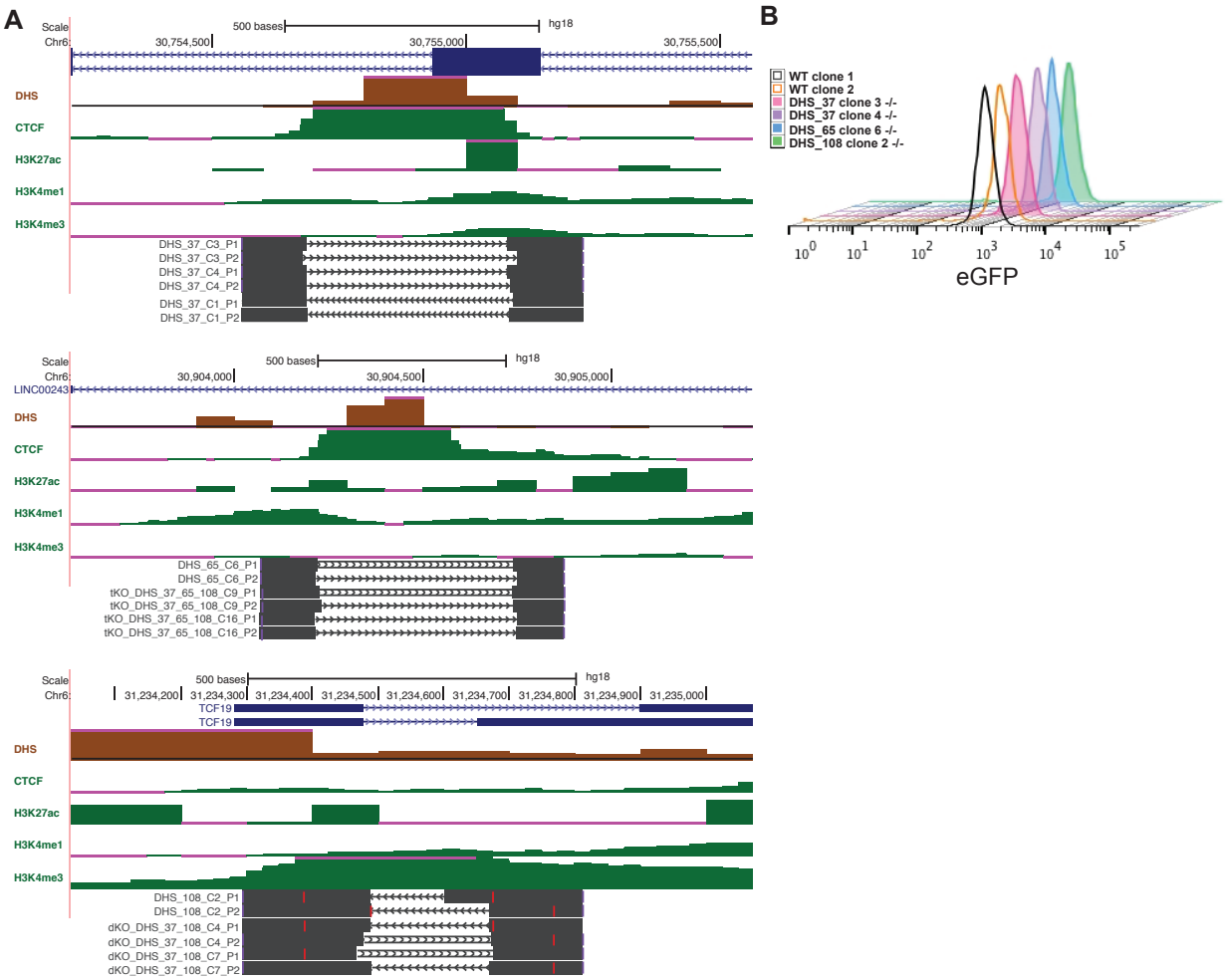


SUPPLEMENTAL Fig. S3. Characterization of regulatory elements identified by lentiCRISPR screening.

(A) DHS_37, CTCF_38, DHS_65 or DHS_108 sequence was cloned into the PE plasmid, a luciferase reporter construct containing the 360bp minimal *POU5F1* promoter and the proximal enhancer DHS_115. H1 POU5F1-eGFP cells were transfected with indicated reporter constructs. 48 hours after transfection, the cells were subjected to luciferase reporter assay to test the effect of given element on PE reporter activity. P-value was calculated by t-test. ** P =0.004.

(B) Genotyping results of an eGFP- clone derived from the lentiCRISPR screening showed bi-allelic mutation of DHS_115 region.

Supplemental Fig. S4

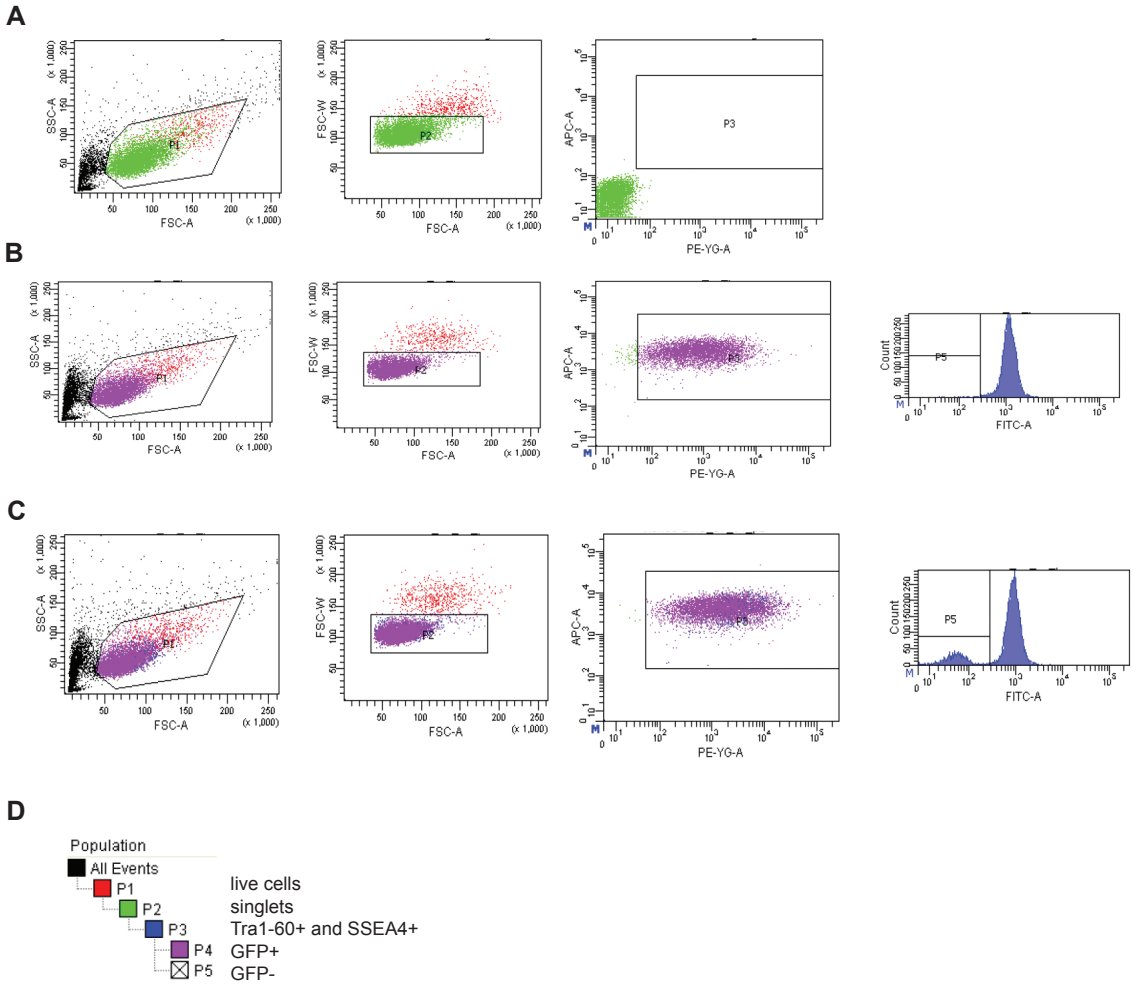


SUPPLEMENTAL Fig. S4. Characterization of bi-allelic mutants.

(A) Genomic DNA was isolated from each mutant clones and the genotypes were confirmed by Sanger sequencing. Genotyping information for all the mutants used in this study with deletions on the loci of DHS_37, DHS_65, and DHS_108 are showing on top panel, middle panel and bottom panel, respectively. DKO: DHS_37 and DHS_108 double knockout clones. DKO clones were derived from DHS_37 bi-allelic mutant clone C1 by transfection of sgRNA pair targeting DHS_108. TKO: DHS_37, DHS_65, and DHS_108 triple knockout clones, derived from DHS_37_108 bi-allelic mutant clone C4 by transfection of sgRNA pair targeting DHS_65.

(B) The bi-allele KO clones were cultured for 14 days and subjected to FACS analyses. All clones exhibit recovered eGFP levels similar to the parental cells after long-term culture.

Supplemental Fig. S5



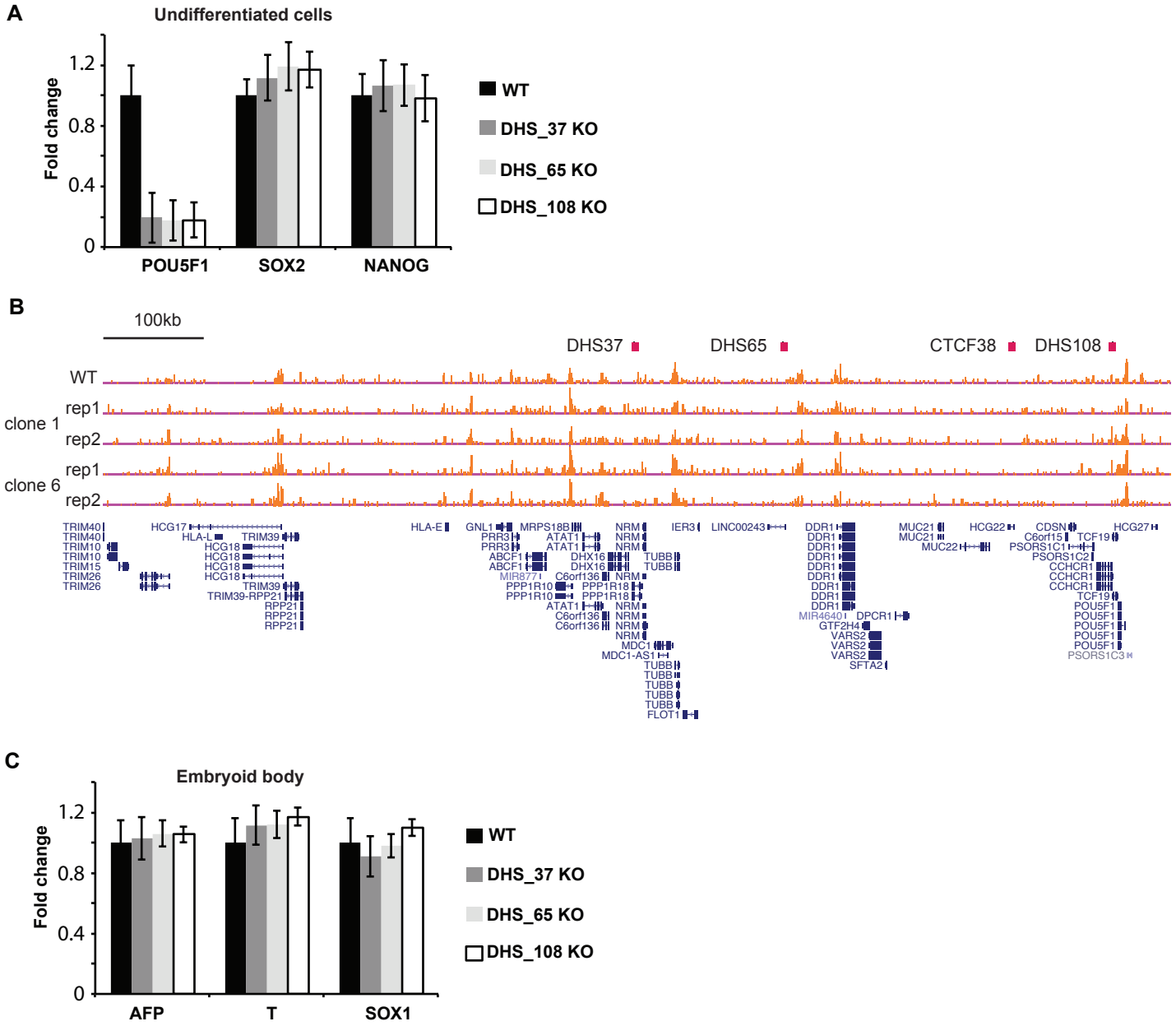
SUPPLEMENTAL Fig. S5. Representative FACS plots for clonal characterization of eGFP levels. WT and temp enhancer KO mutants were dissociated by Accutase and stained with PE-Tra1-60 and APC-SSEA4 antibody (markers for hESCs). The cells were subjected to FACS analysis as indicated, and the un-stained cells were used as negative control for gating.

(A) FACS plots gating for a wild type, unstained clone (PE-YG-A indicates Tra1-60 and APC-A indicates SSEA4).

(B) FACS plots gating for a wild type clone stained with Tra1-60 and SSEA4.

(C) FACS plots gating for a mutant clone stained with Tra1-60 and SSEA4.

(D) Step-wise gating for live, singlets, Tra1-60 and SSEA4+, eGFP- and eGFP+ cells.



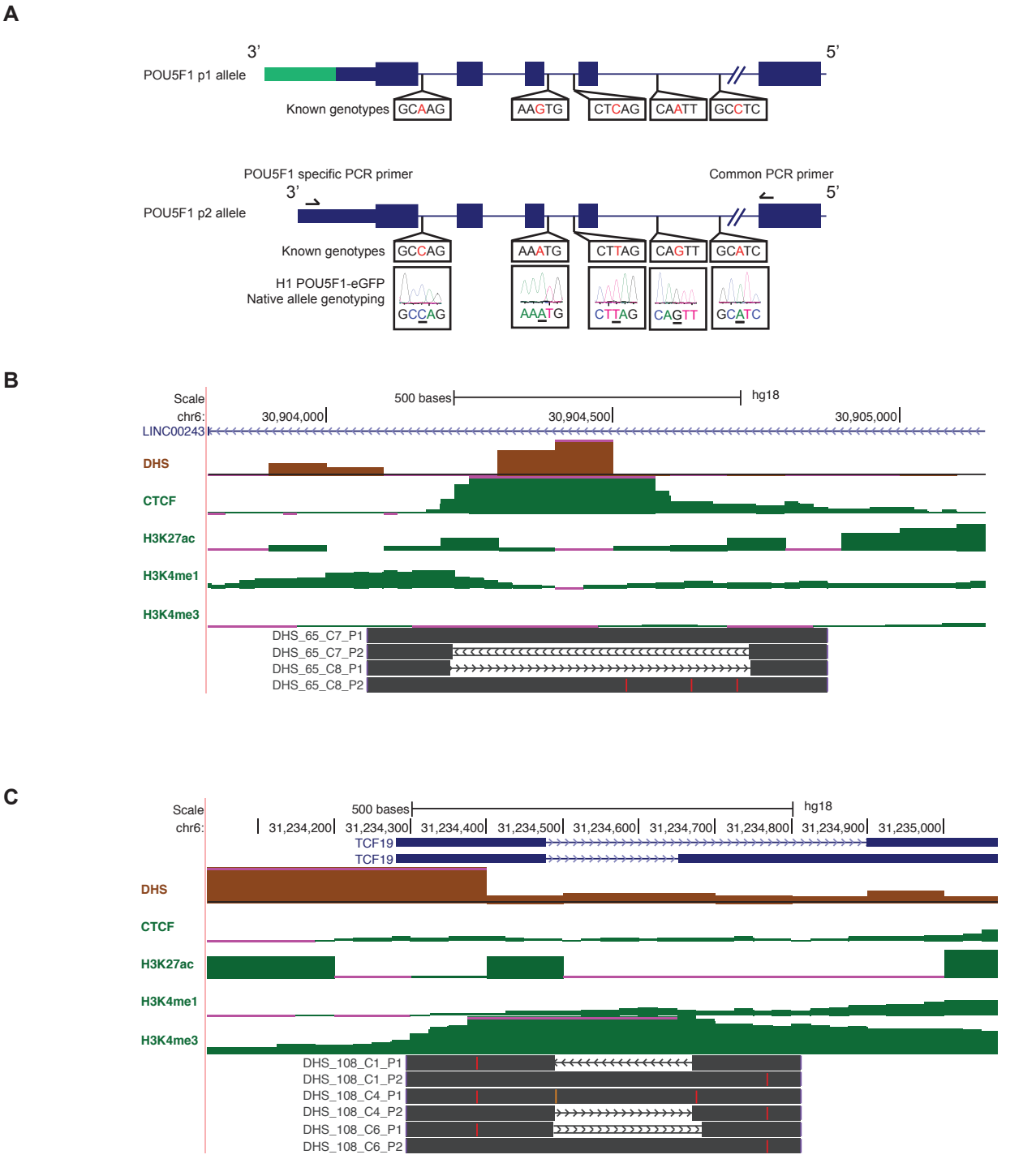
SUPPLEMENTAL Fig. S6. Characterization of pluripotency and differentiation potentials of clones with temp enhancer deletion.

(A) The eGFP negative cells in the indicated mutant lines were FACS sorted, followed by RT-qPCR analysis. The eGFP- single KO cells still maintain normal level of transcription of other stem cell marker including *SOX2* and *NANOG*.

(B) A genome browser snapshot of H3K27ac ChIP-seq signals at *POU5F1* locus in WT and two DHS_37-/- clones.

(A) Embryoid bodies (EB) were generated from FACS sorted eGFP negative cells described in (A). 24 hours after EB formation, the media was switched to differentiation media to induce EB differentiation for 14 days. Total RNA isolated from EBs was analyzed by qPCR. RT-qPCR data showed similar expression levels differentiation marks of the three germ layers in the mutant compared to the embryonic bodies derived from WT cells.

Supplemental Fig. S7



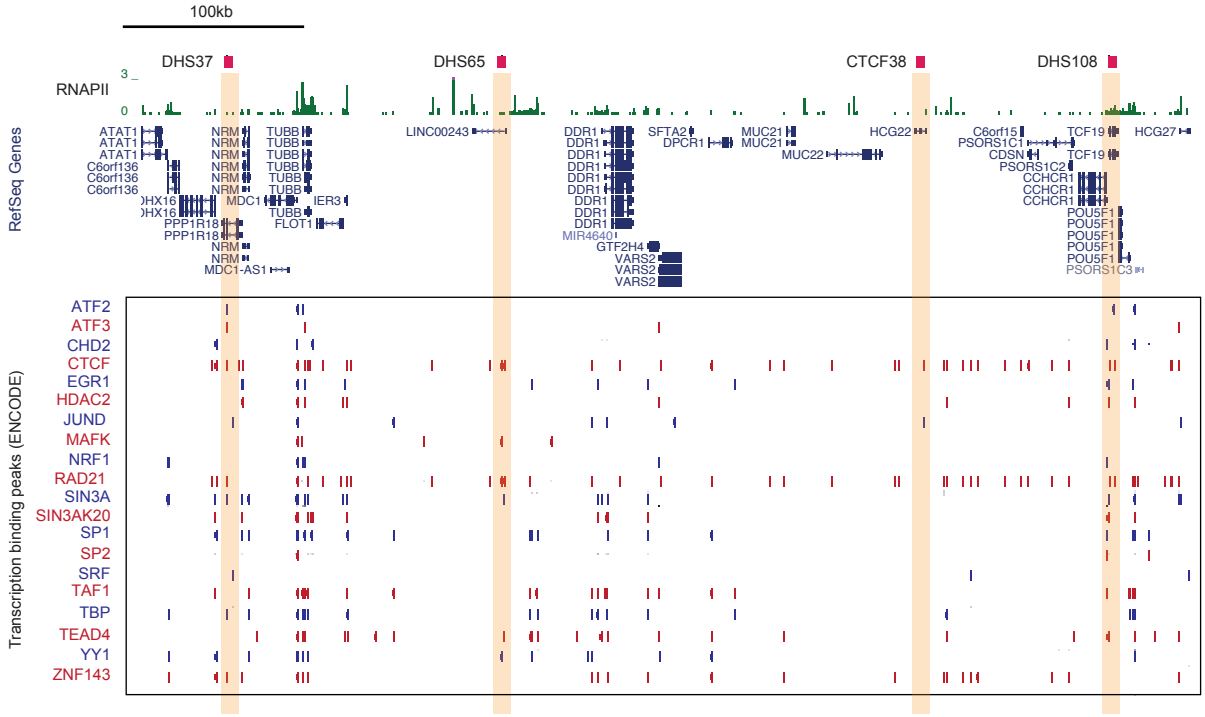
SUPPLEMENTAL Fig. S7. Genotype information for mono-allelic mutant clones

(A) Schemtic of phasing eGFP (P1) and non-eGFP (P2) alleles of H1 POU5F1-eGFP line. We performed PCR from genomic DNA samples of individual clones of a region in the 3' UTR between primer pairs (indicated by black arrows) that would be broken by the inserted transgene, so the only allele that can be amplified is the native one. We then infer what the SNPs on the non-targeted allele are to deduce whether P1 or P2 is the targeted vs. non-targeted allele.

(B) Genotyping information for DHS_65 mono-allelic KO clones.

(C) Genotyping information for DHS_108 mono-allelic KO clones.

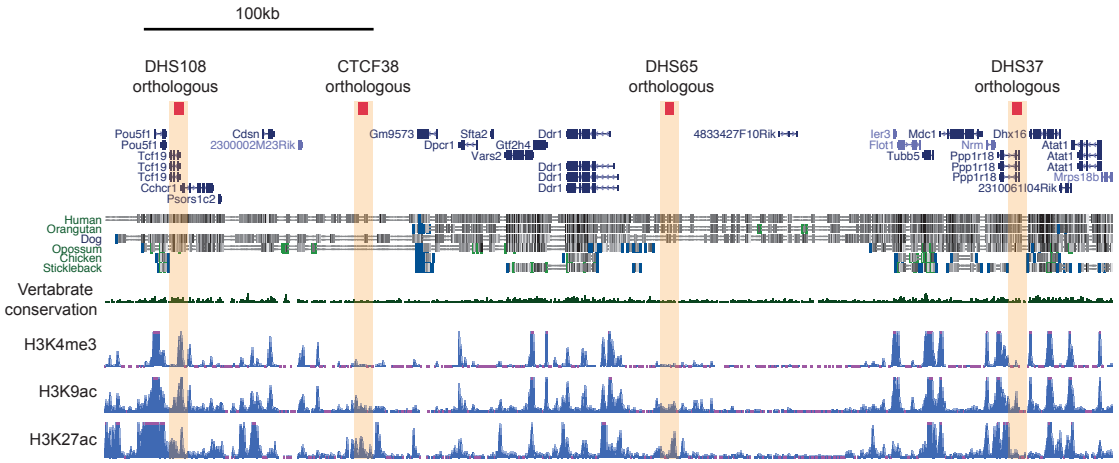
Supplemental Fig.S8



SUPPLEMENTAL Fig. S8. Temp enhancers are bound by multiple TFs.

A genome browser snapshot is shown for PolII binding (top green track) and multiple TF peak patterns at four non-canonical elements together with RefSeq gene annotation. TF peaked regions are obtained from ENCODE dataset. TF peaks that are associated with the novel elements are highlighted with yellow boxes.

Supplemental Fig. S9



SUPPLEMENTAL Fig. S9. Conservation study of identified regulatory elements, and characterization of the pluripotent and differentiation properties of KO clones.

Genome browser snapshot of mouse *Pou5f1* locus. Orthologous of four elements are shown (highlighted in orange) with vertebrate alignments and conservation scores. Three active histone modification marks are shown together.