Supplemental data for:

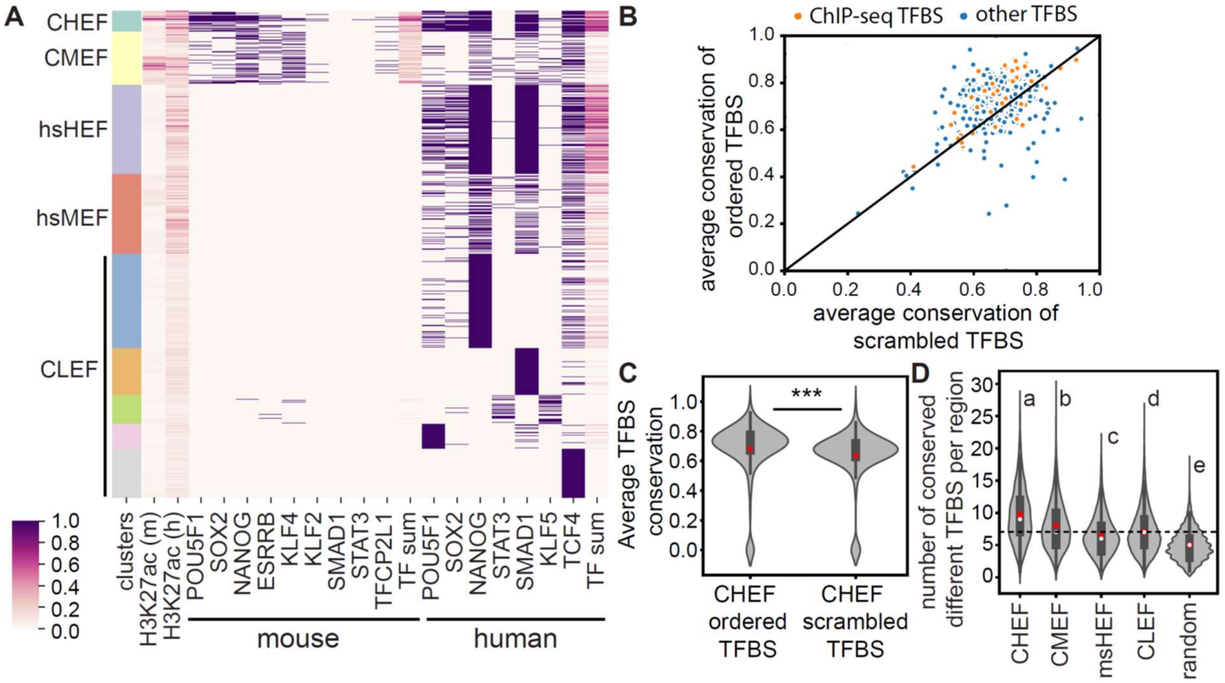
A flexible repertoire of transcription factor binding sites and diversity threshold determines enhancer activity in embryonic stem cells

Gurdeep Singh[1], Shanelle Mullany[1], Sakthi D Moorthy[1], Richard Zhang[1], Tahmid Mehdi[2], Ruxiao Tian[1], Andrew G Duncan[1], Alan M Moses[1,2,3], Jennifer A Mitchell[1]*

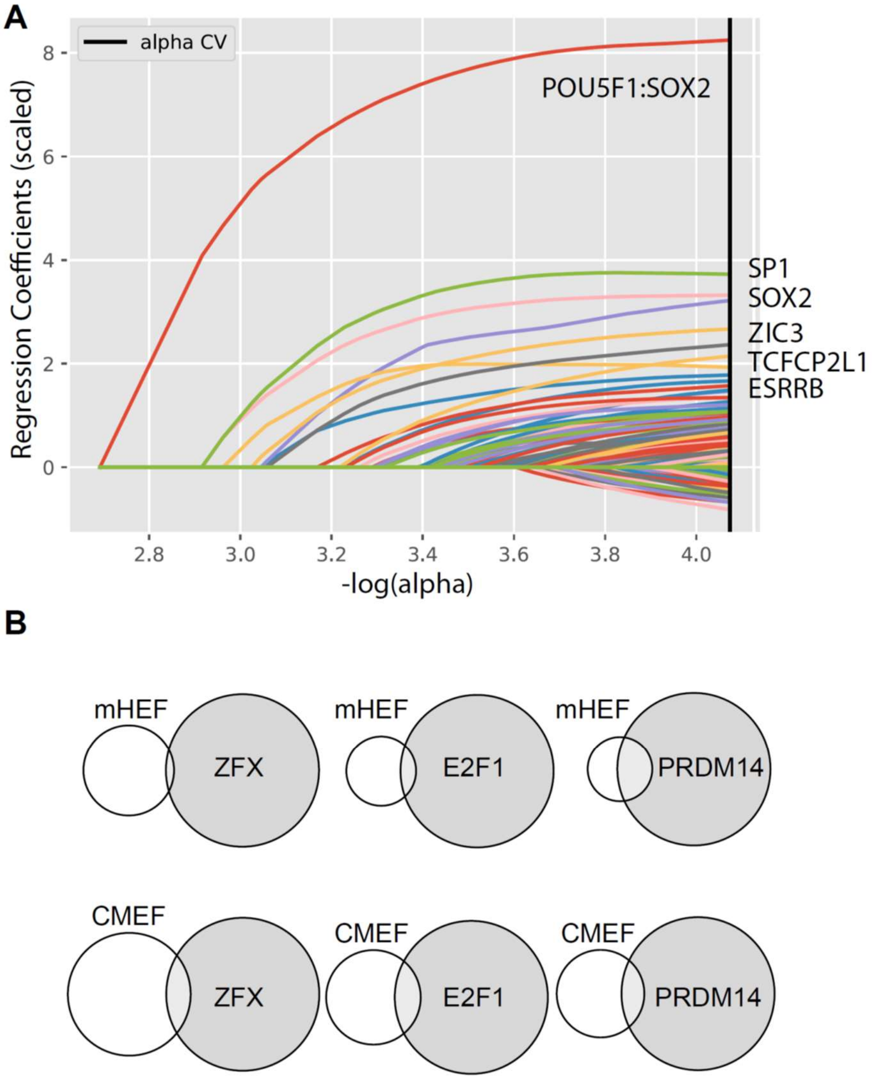
1) Department of Cell and Systems Biology, University of Toronto, Toronto, ON, M5S 3G5, Canada.

2) Department of Computer Science, University of Toronto, Toronto, M5S 2E4, Canada.

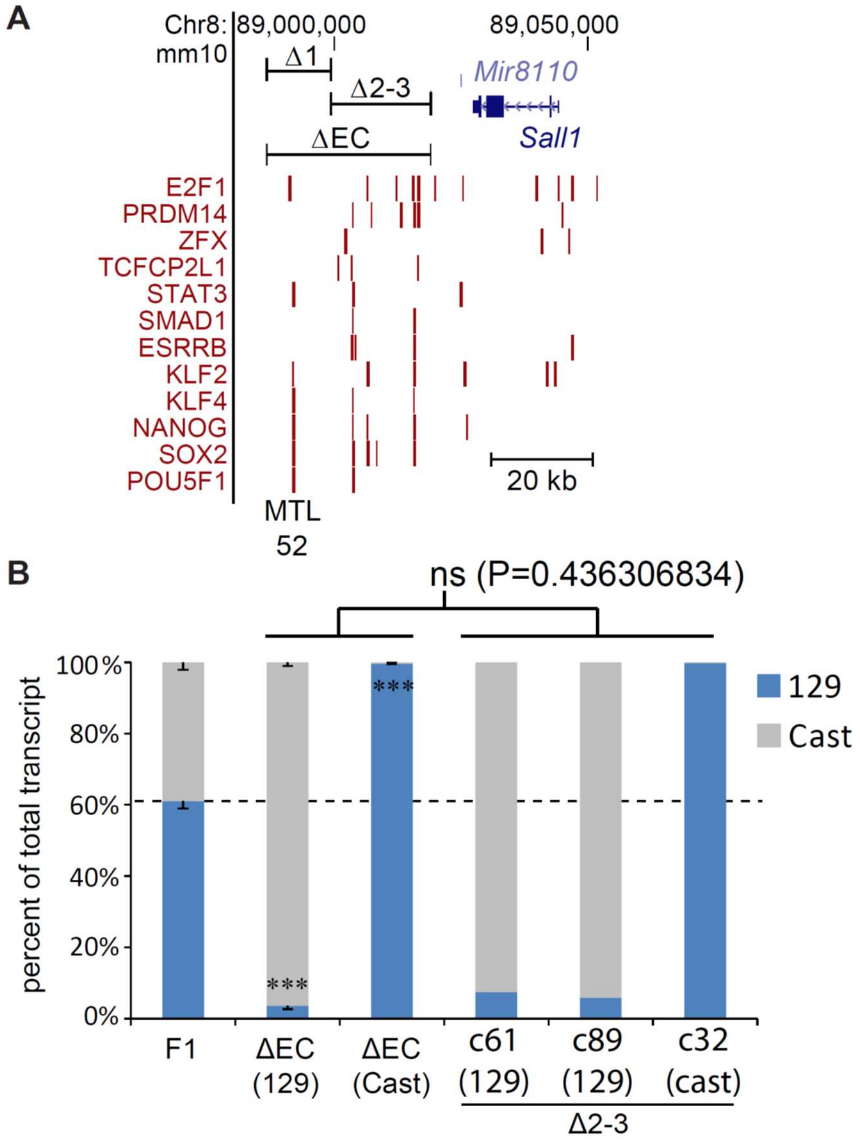
3) Department of Ecology & Evolutionary Biology, University of Toronto, Toronto, M5S 3B3, Canada.



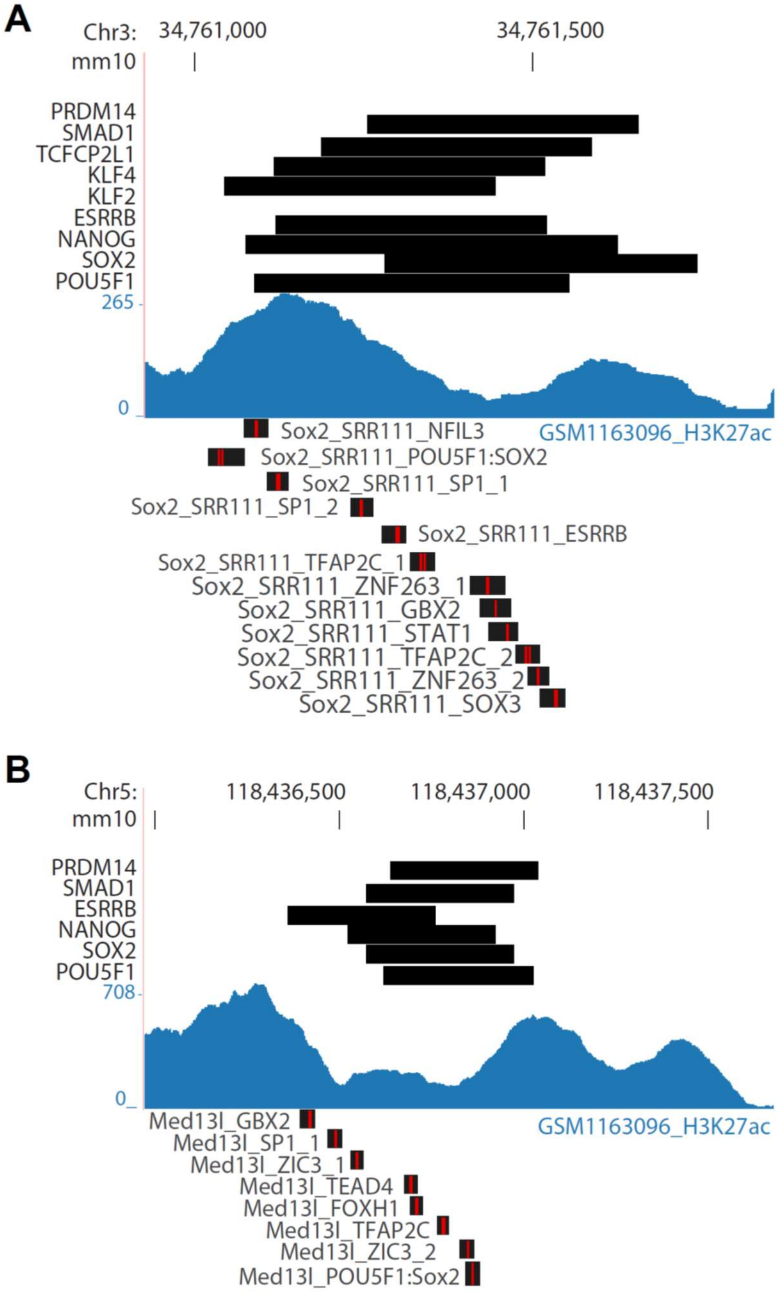
Supplemental Figure S1: Mouse-human conserved high enhancer feature regions (CHEF) contain conserved TFBS. A) Clustering of TF bound regions in human ESCs, using H3K27ac histone modification in human, TF binding and H3K27ac histone modification at associated mouse regions. Clusters are labeled as CHEF, conserved medium enhancer feature (CMEF), human specific high enhancer feature (hsHEF), human specific medium enhancer feature (hsMEF), and conserved low enhancer feature (CLEF) regions. B) Average TFBS conservation for ordered TFBS compared to scrambled TFBS in mouse CHEF regions. Each dot represents a different specific TFBS. The majority of the ordered TFBS motifs are above the diagonal line indicating they are more conserved between mouse-human species than the scrambled TFBS motifs. C) Average mouse-human TFBS sequence conservation in mouse CHEF regions is significantly higher for ordered TFBS motifs compared to scrambled TFBS motifs. D) CHEF cluster regions in mouse contain an increased number of conserved TFBS across six species compared to other clusters and random non-TF bound regions after removal of all TFBS bound by ChIP-seq TFs. Groups determined by one-way ANOVA to be significantly different ($P<0.05$) are labelled with different letters.



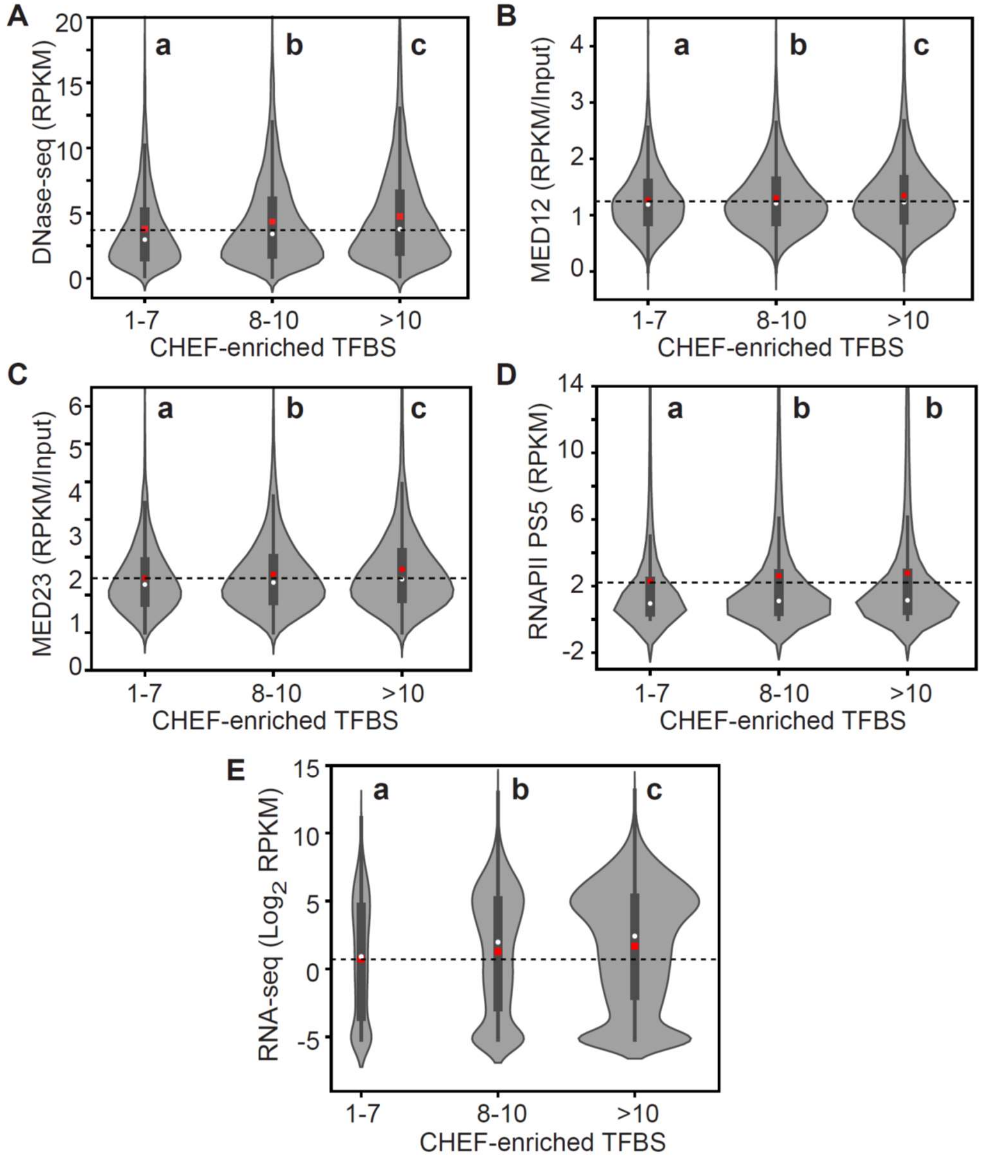
Supplemental Figure S2: TFBS enrichment in CHEF regions and those mutated by site directed mutagenesis (SDM). A) Regression coefficient progression for LASSO paths. B) Overlap of mouse CMEF and msHEF regions with the EP300 bound regions and ChIP-seq for PRDM14, E2F1, or ZFX.



Supplemental Figure S3: The multiple transcription factor bound region 52 kb downstream of *Sall1* (MTL52) does not drive transcription of *Sall1* in ESCs. A) Schematic representation of the *Sall1* locus in the mouse genome. Transcription factor bound regions from ChIP-seq (red bars) are displayed on the mm10 assembly of the University of California at Santa Cruz (UCSC) Genome Browser. CRISPR deleted regions (ΔEC, Δ1, Δ2-3) are displayed. The location of MTL52 in the Δ1 region is indicated. B) *Sall1* expression in wild type F1 clones (F1) compared to clones with the indicated deletion. Allele-specific primers detect 129 or Cast RNA in RT-qPCR. Expression for each allele is shown relative to the total. Error bars represent SEM. Significant differences from the F1 values are indicated by *** $P < 0.001$ for the EC deletion replicates. Deletion of the entire EC dramatically reduces expression of the linked *Sall1* allele. Deletion of the Δ2-3 region similarly reduces expression of the linked *Sall1* allele in 3 separate clones (c61, c89, c32). Comparison of the percent reduced transcription for ΔEC (n=8) compared to Δ2-3 (n=3) revealed no significant difference (ns), indicating MTL52 does not significantly contribute to *Sall1* transcription in ESs.



Supplemental Figure S4: Mutations in the Sox2 or Med13l enhancers. Schematic representation of the *Sox2* (A) or *Med13l* (B) enhancer in the mouse genome. Transcription factor bound regions from ChIP-seq (black bars at top) are displayed on the mm10 assembly of the University of California at Santa Cruz (UCSC) Genome Browser. Mutations introduced by SDM are indicated by red lines in black bars below. H3K27ac ChIP-seq data from ESCs is displayed in blue.



Supplemental Figure S5: TF bound regions containing an increased number of CHEF-enriched TFBS are enriched in enhancer-associated chromatin features and associated with increased gene expression. In all panels, groups determined by one-way ANOVA to be significantly different ($P < 0.01$) are labelled with different letters. Dashed line indicates the average for the 1-7 TFBS group. Transcription factor bound regions in mouse ESC with >10 TFBS have significantly increased (A) DNase I sensitivity (DNase-seq), (B/C) MED12 and MED23 compared to transcription factor bound regions with 8-10 and 1-7 TFBS in the 700bp sequence window. D) RNA polymerase II phosphor S5 (RNAPII PS5) is significantly enriched at regions containing 8-10, or >10 CHEF-enriched TFBS compared to regions with 1-7. E) Genes were separated into 3 groups; those with a >10 TFBS region within 200kb, an 8-10 TFBS or 1-7 TFBS region within 200kb.

Supplemental Table S1: Clustering of transcription factor bound regions in mouse and human ESCs. Data provided in a separate file.

Supplemental Table S2: LASSO coefficients for TFBS enriched or depleted in CHEF compared to the NANOG bound cluster.
Data provided in a separate file.

Supplemental Table S3: Orientation preferences in CHEF sequences.
Data provided in a separate file.

Supplemental Table S4: Synthetic enhancer sequences.
Vector sequence added for assembly is underlined.

14OS	POU5F1:SOX2, POU5F1:SOX2, POU5F1:SOX2, POU5F1:SOX2, POU5F1:SOX2, POU5F1:SOX2, POU5F1:SOX2, POU5F1:SOX2, POU5F1:SOX2, POU5F1:SOX2, POU5F1:SOX2, POU5F1:SOX2, POU5F1:SOX2, POU5F1:SOX2	
GGTCTGACAGCGGCCGCACTTGTGCCTGAACACCATATCCCctttgttatgcaaatCCctttgttatgcaaatCCctttgttatgcaaatCCctttgttatgcaaatCCctttgttatgcaaatCCctttgttatgcaaatCCctttgttatgcaaatCCctttgttatgcaaatCCctttgttatgcaaatCCctttgttatgcaaatCCctttgttatgcaaatCCctttgttatgcaaatCCctttgttatgcaaatCCctttgttatgcaaatCCGTTTTCTTGGCAATCGACTCTCGCGGCCGCAAATGCTAA		GC: 40.57%
ksOE(4TFBS) (longer)	*klf4, sox2*, POU5F1, Esrrb	
GGTCTGACAGCGGCCGCACTTGTGCCTGAACACCATATCAGCTACCGGCCCCGCCCCGTAGCTACGAAACAATGAGCGTAGCTACGGGATGCTAATCGTAGCTACGTTCAAGGTCACGTGTTTTCTTGGCAATCGACTCTCGCGGCCGCAAATGCTAA		GC: 55.70%
sOKE(4TFBS) (longer)	*sox2*, POU5F1, Klf4, Esrrb	
GGTCTGACAGCGGCCGCACTTGTGCCTGAACACCATATCAGCTACGAAACAATGAGCGTAGCTACGGGATGCTAATCGTAGCTACGGGGCGGGGCCGGTAGCTACGTTCAAGGTCACGTGTTTTCTTGGCAATCGACTCTCGCGGCCGCAAATGCTAA		GC: 55.70%
ksOE(4TFBS) (CC)	*klf4, sox2*, POU5F1, Esrrb	
GGTCTGACAGCGGCCGCACTTGTGCCTGAACACCATATCCCCGGCCCCGCCCCCGAAACAATGAGCCCGGGATGCTAATCCCGTTCAAGGTCACCCGTTTTCTTGGCAATCGACTCTCGCGGCCGCAAATGCTAA		GC: 60.29%
sOKE(4TFBS) (CC)	*sox2*, POU5F1, Klf4, Esrrb	
GGTCTGACAGCGGCCGCACTTGTGCCTGAACACCATATCCCGAAACAATGAGCCCGGGATGCTAATCCCGGGGCGGGGCCGCCGTTCAAGGTCACCCGTTTTCTTGGCAATCGACTCTCGCGGCCGCAAATGCTAA		GC: 60.29%
14dTFBS_a	Zic3, POU5F1:SOX2, Tfcp2, Smad3, Sp1, ZNF263, KLf4, Ahr::Arnt, Foxh1, Foxj3, Esrrb, STAT1, TFAP2C, TEAD4	
GGTCTGACAGCGGCCGCACTTGTGCCTGAACACCATATCCCggcccccgctgtgcCCctttgttatgcaaatCCAAACCGGTTTCCgtctagacaCCccccgcccccCCggaggaggaggaggaggaggaCCtaaaggaaggCCcgcagacaagcagcggCCtccaatccacaCCataaagtaaacaaacacCCagctcaaggtcaCCagaaaatgaaactgCCtgccctagggcaCCcacattccatCCGTTTTCTTGGCAATCGACTCTCGCGGCCGCAAATGCTAA		GC: 58.08%
14dTFBS_b	FOXP1, STAT1:STAT2, ZFP263, SMAD3, ZIC3, LIN54, ESRRB, TFCP2, PRDM1, MLXIP, E2F3, RREB1, TEAD3, TCF7L2	
GGTCTGACAGCGGCCGCACTTGTGCCTGAACACCATATCCCataaagtaaacaaacacCCgtagtttcatttcccCCggaggaggaggaggaggaggaCCgtctagacaCCggcccccgctgtgcCCatttgaattCCagctcaggtcaCCAAACCGGTTTCCagaaagtgaaagttaCCgcacgtgtCCctcccgcccccgcccCCacccaaaccacccccacacaCCcacattccatCCgaagttcaaaggaaCCGTTTTCTTGGCAATCGACTCTCGCGGCCGCAAATGCTAA		GC: 56.86%

14dTFBS_c	TFCP2, TFAP2C, SMAD3, ESRRB, LIN54, ZIC3, ZFX, ZBTB7A, FOXH1, PRDM1, MLXIP, TCF7L2, FOXP1, ZFP263	
GGTCTGACAGCGGCCGCACTTGTGCCTGAACACCATATCCCAAACCGGTTTCCtgccctagggcaCCcgtctagacaCCagctcaaggtcaCCatttgaattCCggccccccgctgtgcCCcccgccgcggcctgCCggcgaccacagaCCtccaatccacaCCagaaagtgaaagttaCCgcacgtgtCCgaagttcaaaggaaCCataaagtaaacaaacacCCggaggaggaggaggaggaggaCCGTTTTCTTGGCAATCGACTCTCGCGGCCGCAAATGCTAA		GC: 58.33%
14dTFBS_d	HIC2, FOXD1, Myc, MAFK, Hic1, TFEB, TP53, ZNF410, HINFP, HLF, TFAP2C(var.2), Bhlhe40, JUNB, Stat4	
GGTCTGACAGCGGCCGCACTTGTGCCTGAACACCATATCCCatgcccaccCCgtaaacatCCccatgtgcttCCaagactcagcaatttCCatgccaaccCCatcacgtgacCCctggacatgcctgtgcctgtCCtgcatcccataataactCCctacgtccgcCCggttacataattCCagcctcaggcaCCctcacgtgcacCCggatgactcatCCtttccaggaaataaCCGTTTTCTTGGCAATCGACTCTCGCGGCCGCAAATGCTAA		GC: 55.27%
14dTFBS_e	JUNB, FOSL1, Hic1, Stat4, Ddit3::Cebpa, TFAP2C(var.2), HLF, TFEB, TP53, GLI2, Bhlhe40, PBX1, Klf12, MAX	
GGTCTGACAGCGGCCGCACTTGTGCCTGAACACCATATCCCggatgactcatCCggtgactcatgCCatgccaacccCCtttccaggaaataaCCggatgcaatcccCCagcctcaggcaCCggttacataattCCatcacgtgacCCacatgccctggcatgCCgcgaccacactgCCctcacgtgcacCCccatcaatcaaaCCgaccacgccctttctCCaagcacatggCCGTTTTCTTGGCAATCGACTCTCGCGGCCGCAAATGCTAA		GC: 57.51%
14dTFBS_f	SREBF2, MEF2D, Bhlhe40, Stat4, JUNB, TFEB, ETV5, REST, HIC2, SMAD2::SMAD3::SMAD4a, MTF1, RARA, MAFK, Klf12	
GGTCTGACAGCGGCCGCACTTGTGCCTGAACACCATATCCCatggggtgatCCactataaatagcCCctcacgtgcacCCtttccaggaaataaCCggatgactcatCCatcacgtgacCCaccggaagtgCCgcgctgtccctggttctgaCCatgcccaccCCctgtctgtcacctCCtttgcacacggcacCCaaggtcagaagttcaaggCCaagactcagcaatttCCgaccacgccctttctCCGTTTTCTTGGCAATCGACTCTCGCGGCCGCAAATGCTAA		GC: 56.40%
14dTFBS_g	Ddit3::Cebpa, EN1, GLI2, MAX, TCF4, SMAD2::SMAD3::SMAD4a, SREBF2, PBX1, Hic1, Gmeb1, HIC2, FOXD1, Myc, MTF1	
GGTCTGACAGCGGCCGCACTTGTGCCTGAACACCATATCCCggatgcaatcccCCctaattagCCgcgaccacactgCCaagcacatggCCcgcacctgctCCctgtctgtcacctCCatggggtgatCCccatcaatcaaaCCatgccaacccCCcagtttacgtaagaagaCCatgcccaccCCgtaaacatCCccatgtgcttCCtttgcacacggcacCCGTTTTCTTGGCAATCGACTCTCGCGGCCGCAAATGCTAA		GC: 57.25%
14dTFBS	Zic3, POU5F1:SOX2, Tfcp2, Smad3, Sp1, ZNF263, KLf4, Ahr::Arnt, Foxh1, Foxj3, Esrrb, STAT1, TFAP2C, TEAD4	
GGTCTGACAGCGGCCGCACTTGTGCCTGAACACCATATCCCggccccccgctgtgcCCctttgttatgcaaatCCAAACCGGTTTCCgtctagacaCCccccgcccccCCggaggaggaggaggaggaggaCCtaaaggaaggCCcgcagacaagcagcggCCtccaatccacaCCataaagtaaacaaacacCCagctcaaggtcaCCagaaaatgaaactgCCtgccctagggcaCCcacattccatCCGCGGCCGCAAATGCTAA		GC: 59.11%
12dTFBS	Zic3, POU5F1:SOX2, Tfcp2, Smad3, Sp1, ZNF263, KLf4, Ahr::Arnt, Foxh1, Foxj3, Esrrb, STAT1	
GGTCTGACAGCGGCCGCACTTGTGCCTGAACACCATATCCCggccccccgctgtgcCCctttgttatgcaaatCCAAACCGGTTTCCgtctagacaCCccccgcccccCCggaggaggaggaggaggaggaCCtaaaggaaggCCcgcagacaagcagcggCCtccaatccacaCCataaagtaaacaaacacCCagctcaaggtcaCCagaaaatgaaactgCCGCGGCCGCAAATGCTAA		GC: 58.85%

10dTFBS	Zic3, POU5F1:SOX2, Tfcp2, Smad3, Sp1, ZNF263, KLf4, Ahr::Arnt, Foxh1, Foxj3	
GGTCTGACAGCGGCCGCACTTGTGCCTGAACACCATATCCCggcccccgctgtgcCCctttgttatgcaaatCCAAACCGGTTTCCcgtctagacaCCccccgcccccCCggaggaggaggaggaggaggaCCtaaaggaaggCCcgcagacaagcagcggCCtccaatccacaCCataaagtaaacaaacacCCGCGGCCGCAAATGCTAA		GC: 60.56%
7dTFBS	Zic3, POU5F1:SOX2, Tfcp2, Smad3, Sp1, ZNF263, KLf4	
GGTCTGACAGCGGCCGCACTTGTGCCTGAACACCATATCCCggcccccgctgtgcCCctttgttatgcaaatCCAAACCGGTTTCCcgtctagacaCCccccgcccccCCggaggaggaggaggaggaggaCCtaaaggaaggCCGCGGCCGCAAATGCTAA		GC: 63.19%
5dTFBS	Zic3, POU5F1:SOX2, Tfcp2, Smad3, Sp1	
GGTCTGACAGCGGCCGCACTTGTGCCTGAACACCATATCCCggcccccgctgtgcCCctttgttatgcaaatCCAAACCGGTTTCCcgtctagacaCCccccgcccccCCGCGGCCGCAAATGCTAA		GC: 63.28%
13dTFBS_rOri	Zic3, POU5F1:SOX2, Tfcp2, Smad3, Sp1, ZNF263, KLF4, Foxh1, Foxj3, Esrrb, STAT1, TFAP2C, TEAD4	
GGTCTGACAGCGGCCGCACTTGTGCCTGAACACCATATCCCggcccccgctgtgcCCctttgttatgcaaatCCAAACCGGTTTCCcgtctagacaCCccccgcccccCCggaggaggaggaggaggaggaCCtaaaggaaggCCtccaatccacaCCataaagtaaacaaacacCCagctcaaggtcaCCagaaaatgaaactgCCtgccctagggcaCCcacattccatCCGTTTTCTTGGCAATCGACTCTCGCGGCCGCAAATGCTAA		GC: 57.14%
13dTFBS_rOri+NFYA (14TFBS)	Zic3, POU5F1:SOX2, Tfcp2, Smad3, Sp1, ZNF263, KLF4, NFYA, Foxh1, Foxj3, Esrrb, STAT1, TFAP2C, TEAD4	
GGTCTGACAGCGGCCGCACTTGTGCCTGAACACCATATCCCggcccccgctgtgcCCctttgttatgcaaatCCAAACCGGTTTCCcgtctagacaCCccccgcccccCCggaggaggaggaggaggaggaCCtaaaggaaggCCcgcagccaatcagcggCCtccaatccacaCCataaagtaaacaaacacCCagctcaaggtcaCCagaaaatgaaactgCCtgccctagggcaCCcacattccatCCGTTTTCTTGGCAATCGACTCTCGCGGCCGCAAATGCTAA		GC: 58.08%
13dTFBS_pOri	Zic3, POU5F1:SOX2, Tfcp2, Smad3, Sp1, *znf263, klf4*, Foxh1, Foxj3, Esrrb, STAT1, TFAP2C, TEAD4	
GGTCTGACAGCGGCCGCACTTGTGCCTGAACACCATATCCCggcccccgctgtgcCCctttgttatgcaaatCCAAACCGGTTTCCcgtctagacaCCccccgcccccCCtcctcctcctcctcctccCCccttcctttaCCtccaatccacaCCataaagtaaacaaacacCCagctcaaggtcaCCagaaaatgaaactgCCtgccctagggcaCCcacattccatCCGTTTTCTTGGCAATCGACTCTCGCGGCCGCAAATGCTAA		GC: 57.14%

Supplemental Table S5: Genome wide ChIP-seq data used

GEO Accession	ChIP	Species	Purpose
GSM288355	Esrrb	Mouse	Clustering
ERR440999	Klf2	Mouse	Clustering
ERR440998	Klf2	Mouse	Clustering
GSM288354	Klf4	Mouse	Clustering
GSM288345	Nanog	Mouse	Clustering
GSM307140	Nanog	Mouse	Clustering
GSM1090230	Nanog	Mouse	Clustering
GSM288346	POU5F1	Mouse	Clustering
GSM566277	POU5F1	Mouse	Clustering
GSM307137	POU5F1	Mouse	Clustering
GSM288348	Smad1	Mouse	Clustering
GSM288347	Sox2	Mouse	Clustering
GSM307138	Sox2	Mouse	Clustering
GSM1050291	Sox2	Mouse	Clustering
GSM288353	Stat3	Mouse	Clustering
GSM288350	Tcfcp2l1	Mouse	Clustering
GSM1505690	Klf5	Human	Clustering
GSM1505745	Smad1	Human	Clustering
GSM1505781	Stat3	Human	Clustering
GSM1505791	Tcf4	Human	Clustering
GSM1124071	Nanog	Human	Clustering
GSM1124070	Nanog	Human	Clustering
GSM1124069	Sox2	Human	Clustering
GSM1124068	Sox2	Human	Clustering
GSM1124067	POU5F1	Human	Clustering
GSE52824	H3K27ac	Human	Clustering
GSE47949	H3K27ac	Mouse	Clustering
GSE24164	EP300	Mouse	MTL Validation
GSM288349	E2F1	Mouse	Enrichment Validation
GSM288352	Zfx	Mouse	Enrichment Validation
GSM415050	Mtf2	Mouse	Enrichment Validation
GSM623989	Prdm14	Mouse	Enrichment Validation
GSM1908902	Med12	Mouse	TFBS Repertoire Validation

GSM1908907	Med23	Mouse	TFBS Repertoire Validation
GSM1163123	Pol2	Mouse	TFBS Repertoire Validation

Supplemental Table S6: Primers for Enhancer cloning and gene expression and SDM primers for mutagenesis.

Enhancer name	**Forward primer**	**Reverse primer**
Sox2 (SRR111.1)*	GTGGGGTACAGGACATTGAAA	CTAGTCACTCATCCCCATGC
Med13l	CTCATCTGTGGTGTCGTTGTTG	GTTCAGGAGGGAGAGTTCTGTG
Sall1_MTL52	TAAGCTGTGATGGCCTTATTGC	GAACCATTCTCATCTGACTCTGC
Sall1_MTL52core	GGTGTGCAGTGGGGATGGGGT	TGCCTGGGTAAGAAGGTCCC
Sall1_MTL40	GAGTCTCTTCAGGACAACACCAT	GTCATCTAACAGCAAGCGAATCC
Sall1_MTL28	GGCACCTCTAGAAAGTAAGACCTG	AGCTACACAAAGGTGGGTATCATT
Sall1_C1	CTGGGTGAAGTCATTTCTGAGAC	ATCAACATCCAGGTGTGTCTTCT
Sall1_C2	AGTGAAAGGAGAACTGTTAGATGACC	CTTAGTAATAGGGGCAGCTTGG
Sall1_129	CGTGGCCTTCTTGTCAATg	CAACAGTACTCTGAACTCCCCAgT
Sall1_Cast	CCGTGGCCTTCTTGTCAATa	CAACAGTACTCTGAACTCCCCAaT
Primer name	**Sequence**	
***Med13l* enhancer mutagenesis primers**		
Med13l_GBX2_F	agggcatctccttacgtaatggttattatgggggtgggaca	
Med13l_GBX2_R	tgtcccacccccataataaccattacgtaaggagatgccct	
Med13l_SP1_1_F	agggtcatctgagaggagttaacctcaggtagaaaacatc	
Med13l_SP1_1_R	gatgttttctacctgaggttaactcctctcagatgaccct	
Med13l_ZIC3_1_F	ggtgacagtcagcatcaaacactgtgcaggttgggt	
Med13l_ZIC3_1_R	acccaacctgcacagtgtttgatgctgactgtcacc	
Med13l_TEAD4_F	gtggactccacctggcagtacggatttacagaatggg	
Med13l_TEAD4_R	cccattctgtaaatccgtactgccaggtggagtccac	
Med13l_FOXH1_F	ccctcctgtgtagtggggcctccacctggcattc	
Med13l_FOXH1_R	gaatgccaggtggaggccccactacacaggaggg	
Med13l_TFAP2C_F	ccacttccctggattcagtcccctaccggtg	
Med13l_TFAP2C_R	caccggtaggggactgaatccagggaagtgg	
Med13l_ZIC3_2_F	aaaatgtaaatctacccaatgcagcccccatacctcttgg	
Med13l_ZIC3_2_R	ccaagaggtatgggggctgcattgggtagatttacatttt	
Med13l_Pou5f1::Sox2_F	cttcctcgggcttttaaaagataaatctacccccctgcagc	
Med13l_Pou5f1::Sox2_R	gctgcaggggggtagatttatcttttaaaagcccgaggaag	
Med13l_SP1_2_F	ccgaggctagagcatactccccacctgcc	
Med13l_SP1_2_R	ggcaggtggggagtatgctctagcctcgg	
***Sox2* enhancer mutagenesis primers**		
Sox2_SRR111_NFIL3_F	gaatccgaggccttagcgccgaaacaggttcgagac	
Sox2_SRR111_NFIL3_R	gtctcgaacctgtttcggcgctaaggcctcggattc	
Sox2_SRR111_SP1_1_F	gcactcagggggcttatgcaggagcatcagg	
Sox2_SRR111_SP1_1_R	cctgatgctcctgcataagccccctgagtgc	
Sox2_SRR111_SP1_2_F	cttccggtaggggtgtatcggagggactgcaac	
Sox2_SRR111_SP1_2_R	gttgcagtccctccgatacacccctaccggaag	
Sox2_SRR111_TFAP2C_1_F	ccaaaccaagcacagcaccaatgtagtcagctaggtct	

Sox2_SRR111_TFAP2C_1_R	agacctagctgactacattggtgctgtgcttggtttgg
Sox2_SRR111_GBX2_F	caggttctttttttaaaccctacgtgtcctccaccttccatttgag
Sox2_SRR111_GBX2_R	ctcaaatggaaggtggaggacacgtagggtttaaaaaaagaacctg
Sox2_SRR111_ZNF263_1_F	ttttaaaccctaattgtcctccacgtaccatttgagtcattcttaaattct
Sox2_SRR111_ZNF263_1_R	agaatttaagaatgactcaaatggtacgtggaggacaattagggtttaaaa
Sox2_SRR111_STAT1_F	ggcccatcccaggtgatttttttaaaccctaattgtcctccac
Sox2_SRR111_STAT1_R	gtggaggacaattagggtttaaaaaaatcacctgggatgggcc
Sox2_SRR111_TFAP2C_2_F	ccttcctctaggcagatccagtgggtttacaactggc
Sox2_SRR111_TFAP2C_2_R	gccagttgtaaacccactggatctgcctagaggaagg
Sox2_SRR111_ZNF263_2_F	tctcctccagctccctgctctaggcagctcc
Sox2_SRR111_ZNF263_2_R	ggagctgcctagagcagggagctggaggaga
Sox2_SRR111_SOX3_F	gacctcccctgtggtttctaagctctcctccagct
Sox2_SRR111_ SOX3_R	agctggaggagagcttagaaaccacaggggaggt
Sox2_SRR111_POU5F1:SOX2_F	gagtggggtacaggacactgaagttgagagagggtttcttgtaa
Sox2_SRR111_POU5F1:SOX2_R	ttacaagaaaccctctctcaacttcagtgtcctgtaccccactc
Sox2_SRR111_ESRRB_F	caccaggttatctctggcttcatgatgaccctgact
Sox2_SRR111_ESRRB_R	ccagtcagggtcatcatgaagccagagataacctggtg
Sall1 **MTL 52 mutagenesis primers**	
+OS_F	CCCATTTGCATAACAAAGCAGGAGATTTGTTAC
+OS_R	CCTGCTTTGTTATGCAAATGGGGGGAGGGTG
+ETS_F	CAATGAACGTGTTGCTTTGGGTGGTGCTACTATGACCTTGAGGTCCCCACAGG
+ETS_R	ACCCAAAGCAACACGTTCATTGTTTCTGCCTGCCCCGTCTAGACACCCAATGCAAACATGACTGGTTTGAACTGGGCTTTCGGGTTGGG
+ETS_-FOXP1_F	caaaattaatacgaggcaaatttttttttgtggttaaacacaatgtcttctacaccaggag
+ETS_-FOXP1_R	ctcctggtgtagaagacattgtgtttaaccacaaaaaaaaatttgcctcgtattaattttg
+TS_F	cctgtggggacctccattccatagtagcaccacc
+TS_R	ggtggtgctactatggaatggaggtccccacagg
+TS_+E2F1_F	cacacatcagcagtaacaaacctcccgccctgttctgggaatggggggag
+TS_+E2F1_R	Ctccccccattcccagaacagggcgggaggtttgttactgctgatgtgtg

* Only the 5’ half of 111 was used which corresponded to the TF bound region.

Supplemental Table S7: Guide RNAs.

Deletion	**Sequence**
Sall1_ΔEC	5’ gRNA GTCTAGTGGT CTCTATAGCG (tgg) 3’ gRNA GCAGAGCTAC ACCCTCCGGA G (ggg)
Sall1_Δ1	5’ gRNA GTCTAGTGGT CTCTATAGCG (tgg) 3’ gRNA GGGTGAGCGA ACGAGCTTGG (agg)
Sall1_Δ2	5’ gRNA GGGTGAGCGA ACGAGCTTGG (agg) 3’ gRNA CTTTTCTTGG AGACCGGGCG (ggg)
Sall1_Δ3	5’ gRNA CTTTTCTTGG AGACCGGGCG (ggg) 3’ gRNA GCAGAGCTAC ACCCTCCGGA G (ggg)
Sall1_Δ2-3	5’ gRNA GGGTGAGCGA ACGAGCTTGG (agg) 3’ gRNA GCAGAGCTAC ACCCTCCGGA G (ggg)

Supplemental Table S8: STARR-seq data used

GEO Accession	STARR-seq	Species	Purpose
GSM4261631	RNA 2iL rep 1	Mouse	TFBS Repertoire Validation
GSM4261632	RNA 2iL rep 2	Mouse	
GSM4261635	Input DNA 2iL rep 1	Mouse	
GSM4261636	Input DNA 2iL rep 2	Mouse	