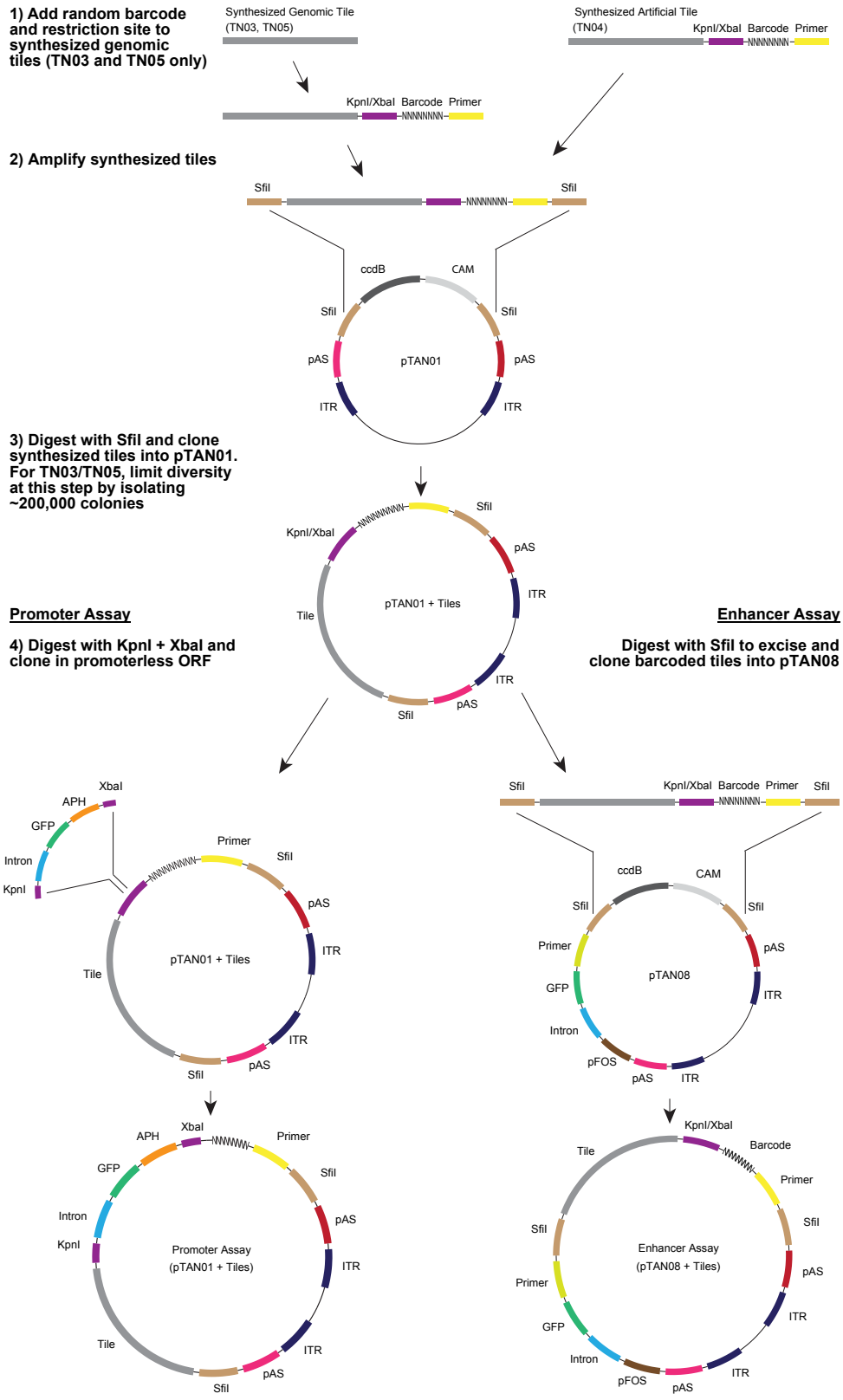
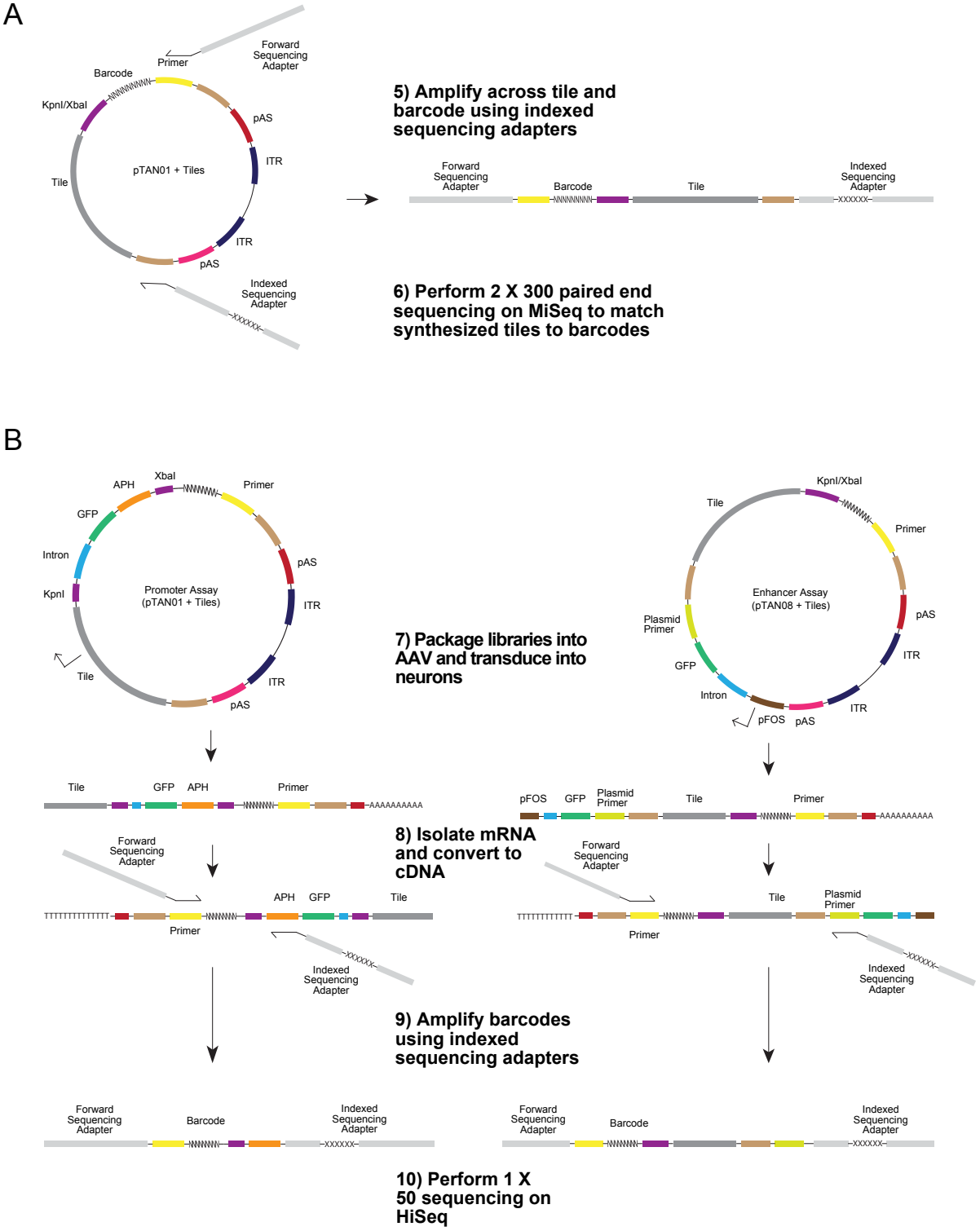
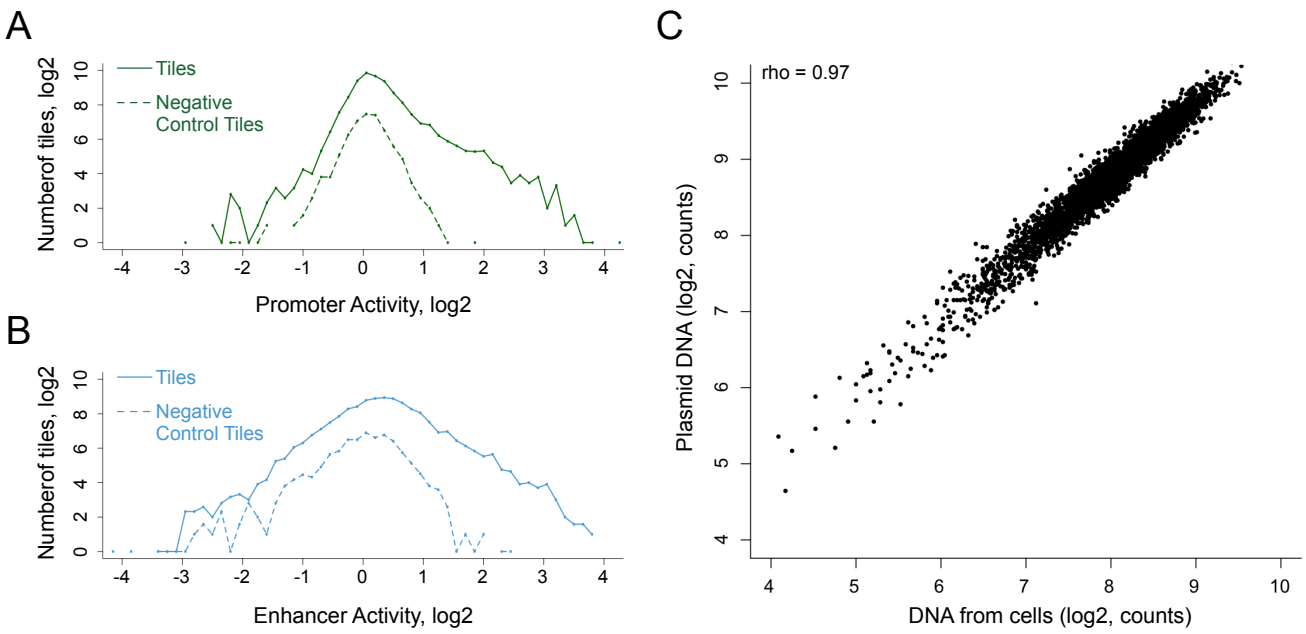


Supplemental Fig S1. **MPRA library construction.** Amplification and cloning of synthesized genomic (TN03), mutated genomic (TN05), and artificial (TN04) sequences. For each library 12,472 distinct sequences were synthesized. ITR, inverted terminal repeat. pAS, cleavage and polyadenylation signal. pFos, basal *Fos* promoter.



Supplemental Fig S2. MPRA library processing. (A) Amplification and sequencing of MPRA plasmid libraries (TN03 and TN05 only) to generate a barcode key relating genomic or synthetic tiles to corresponding barcodes. For TN04, barcodes were instead specified and synthesized with the desired motif sequences. **(B)** The transcription, amplification, and sequencing of MPRA cDNA barcodes in the promoter (left) and enhancer (right) test assays.

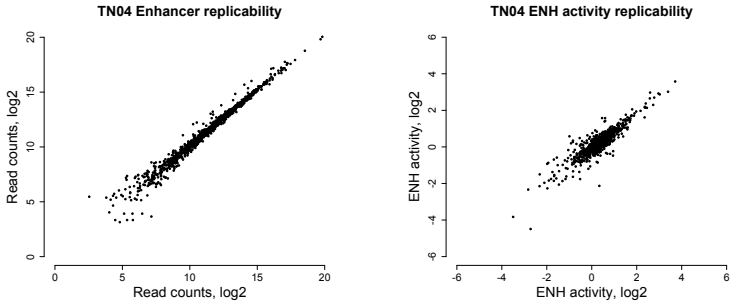


Supplemental Fig S3. Validation of negative control tiles in TN03. (A-B) Number of promoter (A) and enhancer (B) activities for tiles (solid line) as well as 498 reversed (nonsense) control tiles (dashed line). The activities from unstimulated and KCl-depolarized neurons are both included. Control tiles have symmetrical distributions around near-zero median values, suggesting that few or none are transcriptionally active. **(C)** Replicability of DNA read counts between plasmid DNA and DNA extracted from cells.

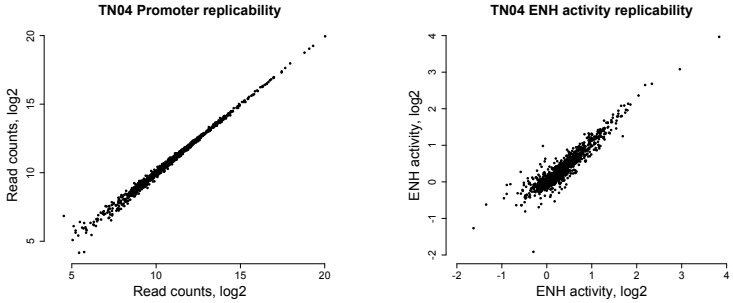
A

Library	Replicates
TN03	2
TN04	3
TN05	3

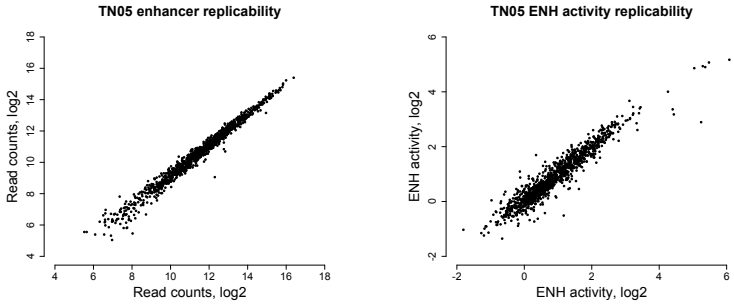
B



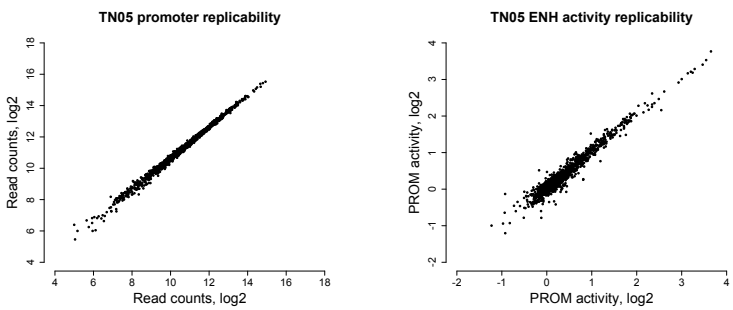
C



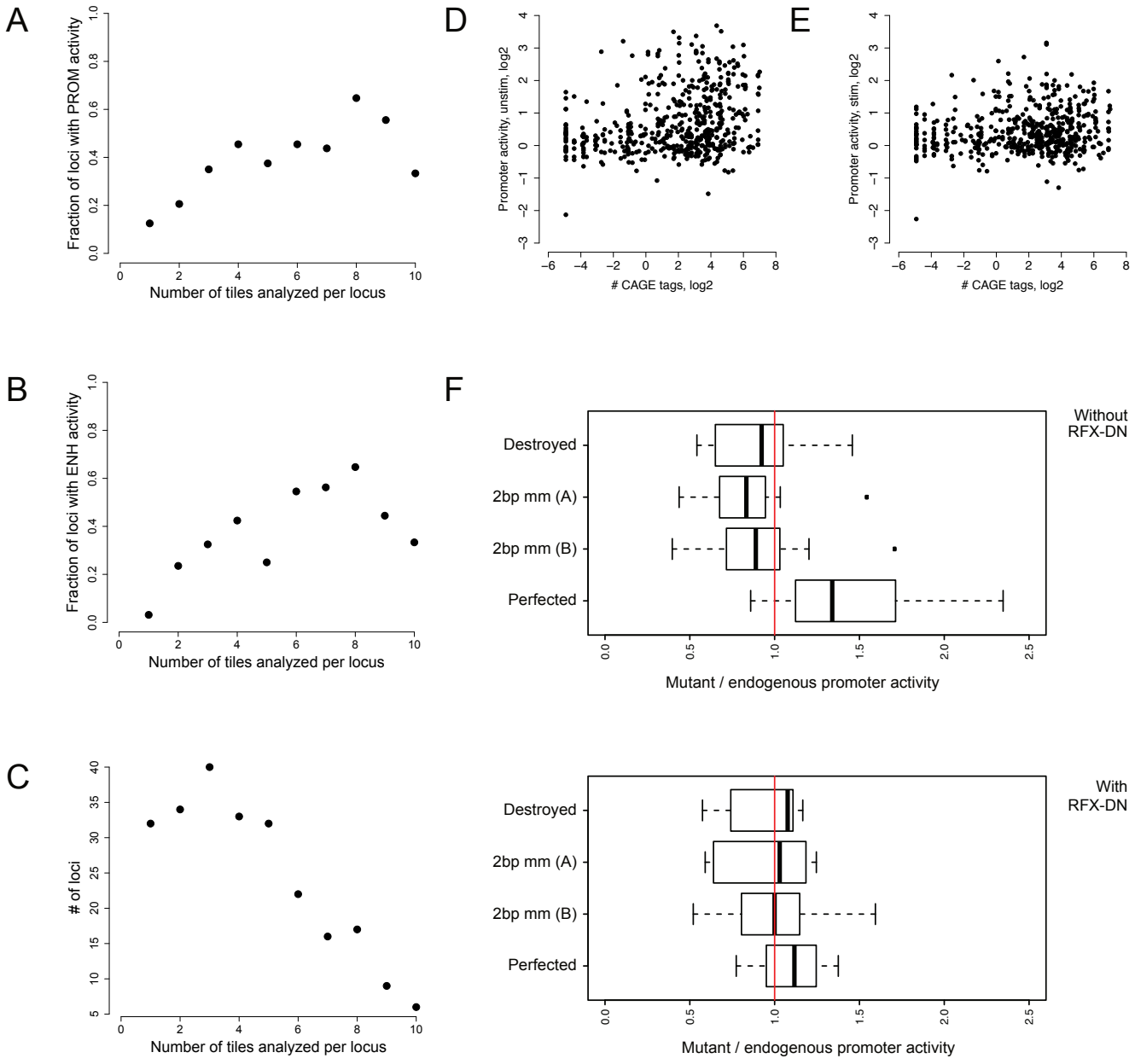
D



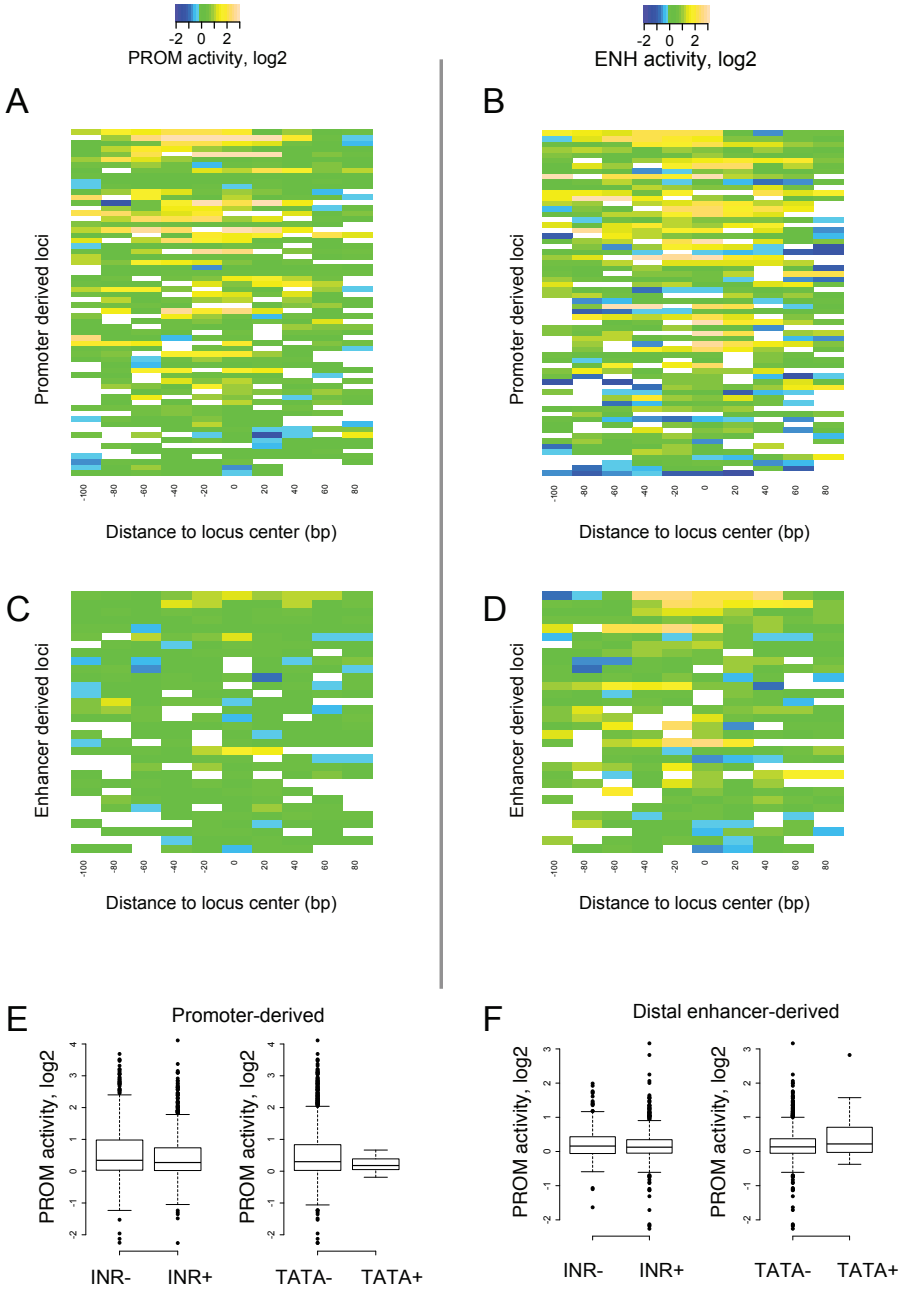
E



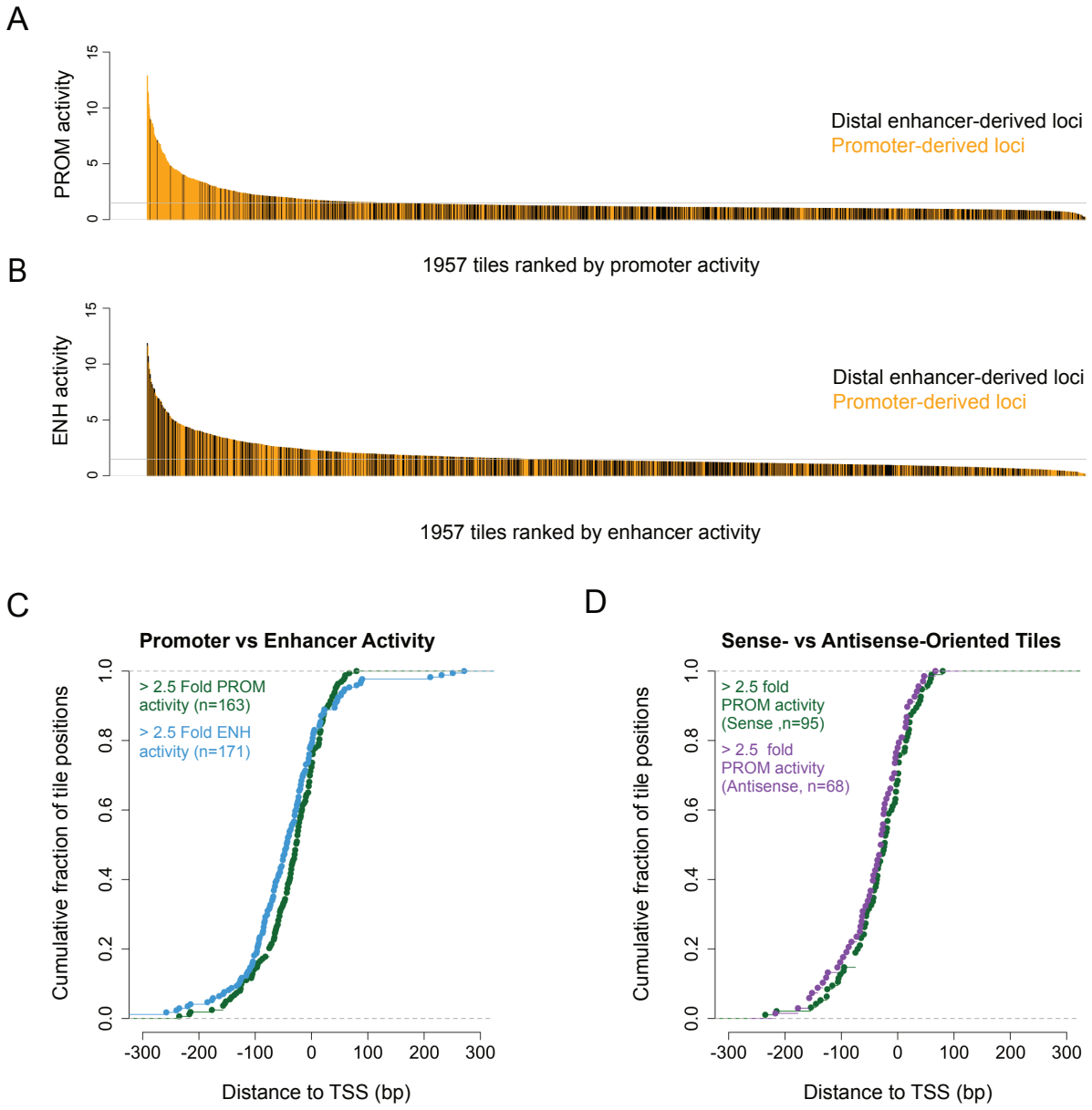
Supplemental Fig S4. Replicates and replicability of MPRA results. (A) The number of replicates for each MPRA experiment. **(B-E)** Replicability of read counts (left panels) and enhancer /promoter activities (right panels) for unstimulated conditions. TN03 is a genomic array with mouse and human tiles. TN04 is a synthetic array with motif sequences. TN05 has both genomic and RFX-mutated genomic sequences. Negative controls used for normalization to obtain enhancer/promoter activity are: TN03, 498 nonsense sequences obtained by reversing human genomic sequences; TN04, 27 sequences in which a synthetic sequence with 4 CREB sites was fully mutated at each CREB site; TN05, 38 genomic regions that lack detectable TF binding.



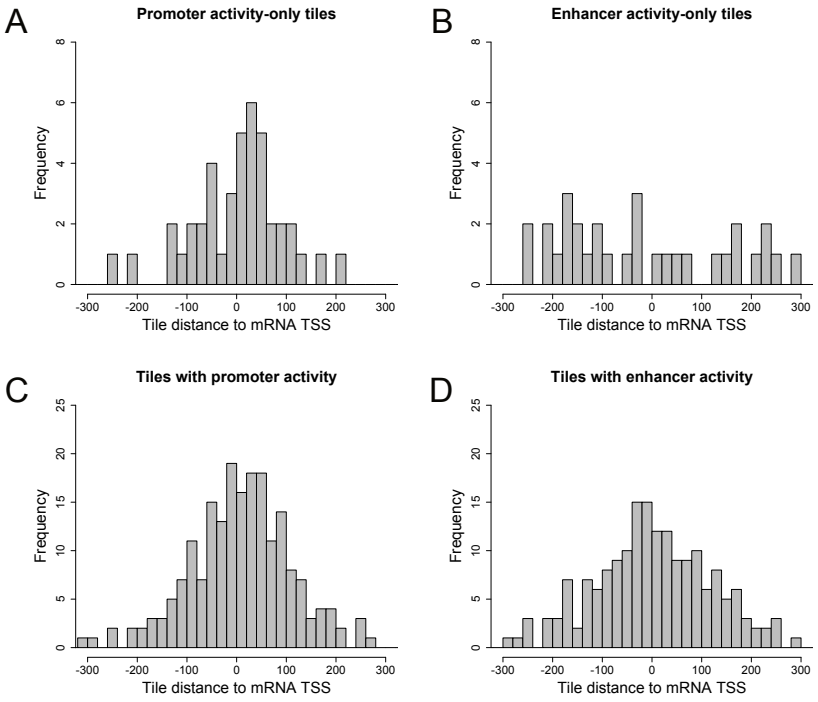
Supplemental Fig S5. Active loci versus number of tiles analyzed per locus. (A) Number of loci with promoter activity (FDR < 0.1) versus coverage (# tiles analyzed per locus). **(B)** The number of loci with enhancer activity (FDR < 0.1) versus coverage. **(C)** Coverage (# of loci with 1-10 tiles analyzed). **(D)** Number of CAGE tages versus promoter activity in the unstimulated condition. **(E)** KCl-stimulated condition. **(F)** Boxplots of ratios of mutant to control promoter activities (TN05), for distal enhancer-derived tiles (n=11), with (top) and without (bottom) the RFX dominant negative.



Supplemental Fig S6. Heatmaps of enhancer and promoter activities by locus and tile. (A) Promoter activity of promoter-derived loci. **(B)** Enhancer activity of promoter-derived loci. **(C)** Promoter activity of enhancer-derived loci. **(D)** Enhancer activity of enhancer-derived loci. Only loci with at least 7 tiles analyzed are shown. White boxes indicate tiles that were not recovered from synthesis. **(E-F)** Distributions of promoter activities for tiles that contain versus do not contain INR (YYRRWYY) or TATA (TATAWAWR) motifs (p > 0.08 for each – vs + pairwise comparison, by t-test). Promoters have 1039 INR-lacking and 391 INR-containing tiles and 1418 TATA-lacking and 12 TATA-containing tiles. Distal enhancers have 301 INR-lacking and 940 INR-containing tiles and 1212 TATA-lacking and 29 TATA-containing tiles.



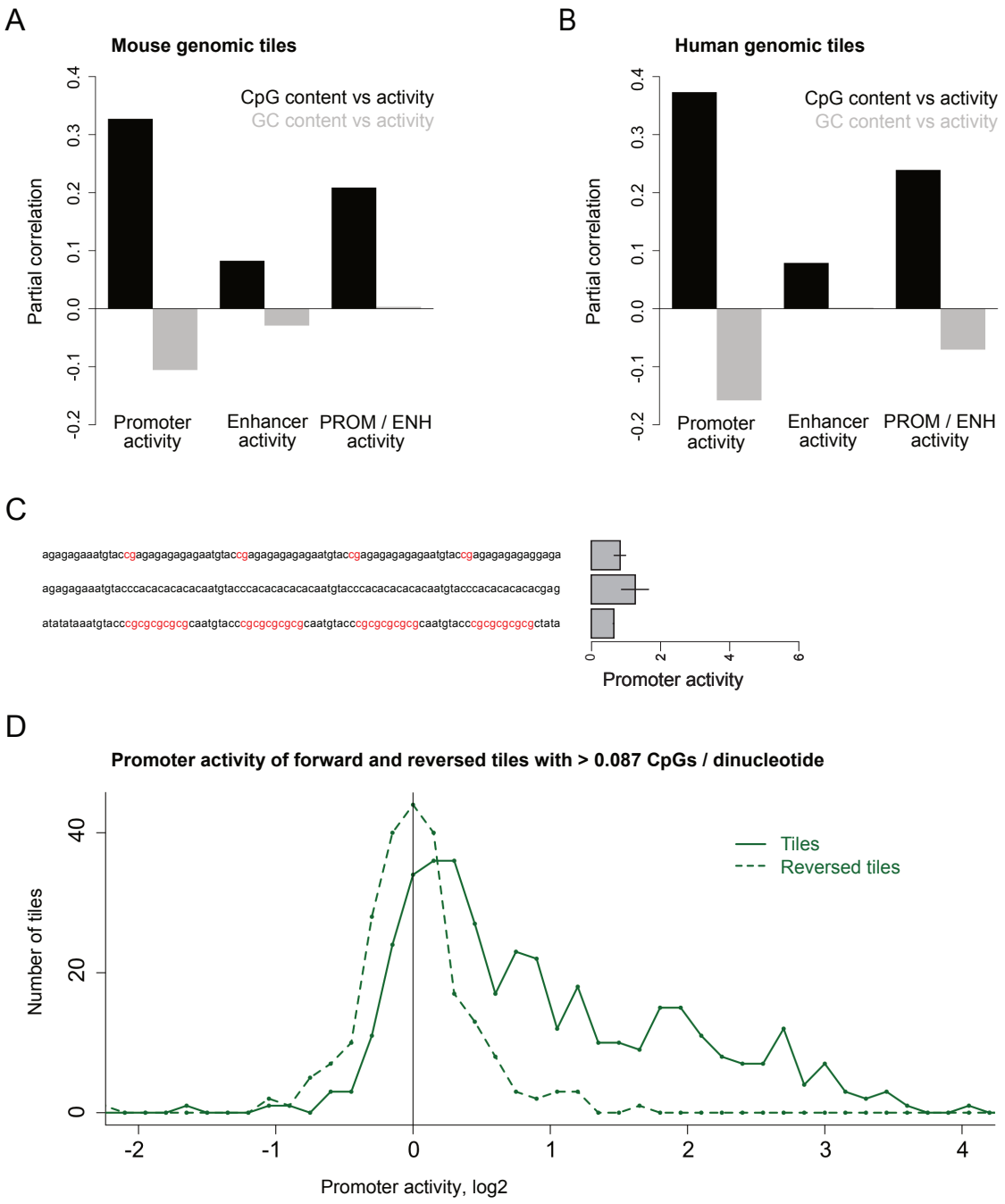
Supplemental Fig S7. Enhancer and promoter activities of distal enhancers and gene promoters. (A-B) Promoter (A) and enhancer (B) activities of tiles derived from promoter and distal enhancer loci. **(C)** Within 300 bp of mRNA TSSs, the locations of the tiles with the greatest promoter and enhancer activities are indistinguishable ($p > 0.05$, KS test). PROM, promoter; ENH, enhancer. **(D)** Within 300 bp of mRNA TSSs, the locations of the tiles with the greatest sense and antisense promoter activities are indistinguishable ($p > 0.05$, KS test). Sense and antisense promoter activities are distinguished based on the orientation in which tiles were inserted into reporter plasmids.



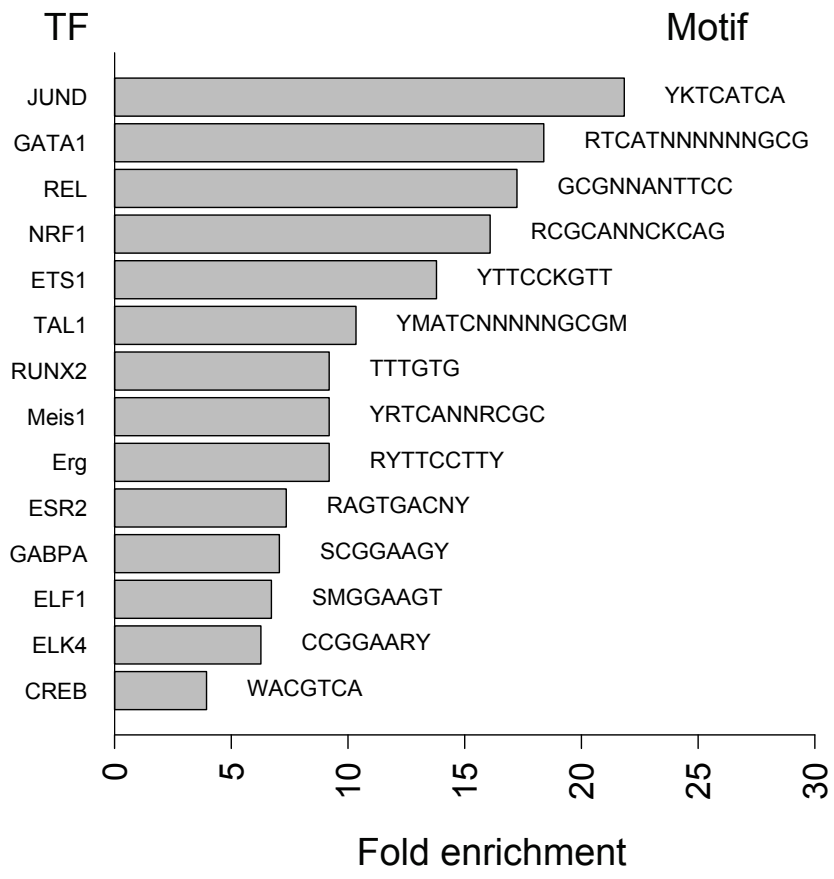
E

TF	Motif	# Prom. Tiles	# Enh. Tiles	Ratio	P-value
Jun	RTGASTMA	0	31	INF	1.13E-06
Lhx3	NYMATTAN	1	21	3.544362	7.94E-04
Fos	TGANTCA	1	20	3.473917	1.18E-03
Lhx3	NNYMATTANN	1	20	3.473917	1.18E-03
HNF1B	NYNATTAN	2	27	2.906891	3.88E-04
Atoh1	CAGNTGG	2	22	2.611408	2.47E-03
KLF5	GGGNGGG	28	18	-1.486004	1.48E-04
EGR1	GGGCGG	29	18	-1.535332	8.17E-05
EGR1	GGGCGGR	26	16	-1.547932	1.55E-04
SP2	RGGCGGGNY	16	16	-2.04972	2.17E-04
NPAS4	TCGTG	17	7	-2.12658	1.02E-04
KLF5	GGGNGGGG	23	8	-2.373327	2.61E-06
GABPA	SCGGAAGY	24	2	-4.442222	1.13E-10

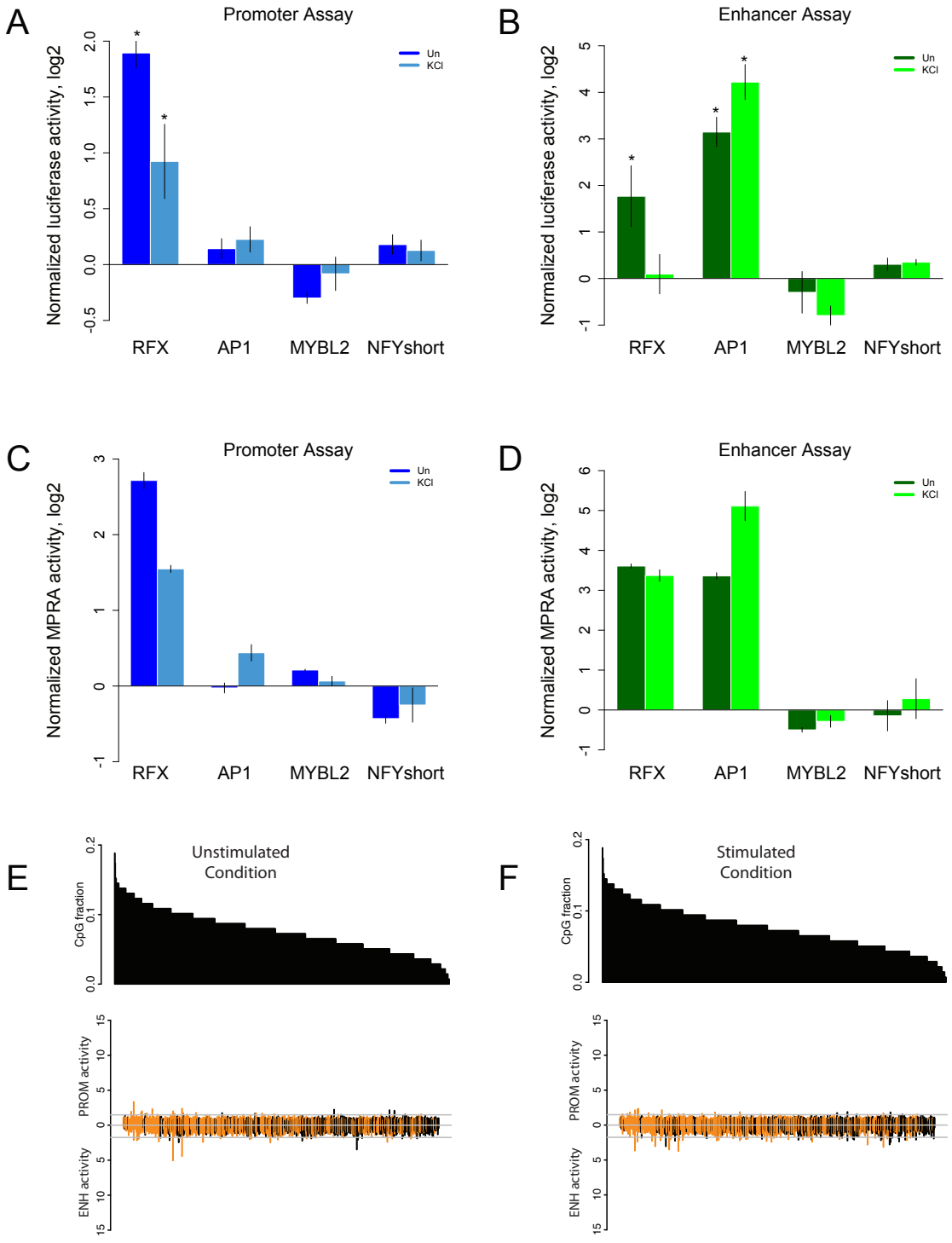
Supplemental Fig S8. Distribution of active tiles across promoter loci. (A) Positions of promoter activity-only tiles relative to mRNA TSSs **(B)** Positions of enhancer activity-only tiles. The promoter activity-only tiles are also distributed differently across promoters than enhancer activity-only tiles, as they are more likely to be situated between 0 and +100 bp relative to mRNA TSSs (compare **A** and **B**, $p < 0.003$, *Chi*-squared test). Promoter activity-only is FDR < 0.1 for promoter activity and enhancer activity < 1.5-fold above controls. Enhancer activity-only is FDR < 0.1 for enhancer activity and promoter activity < 1.5-fold above controls. **(C)** Positions of all tiles with promoter activity. **(D)** Positions of all tiles with enhancer activity. **(E)** Table of motifs enriched in enhancer activity-only or promoter activity-only tiles, including the number of occurrences in 45 promoter activity-only and 81 enhancer activity-only tiles, the ratio of counts per tile in each group, and a *p*-value based on a binomial test.



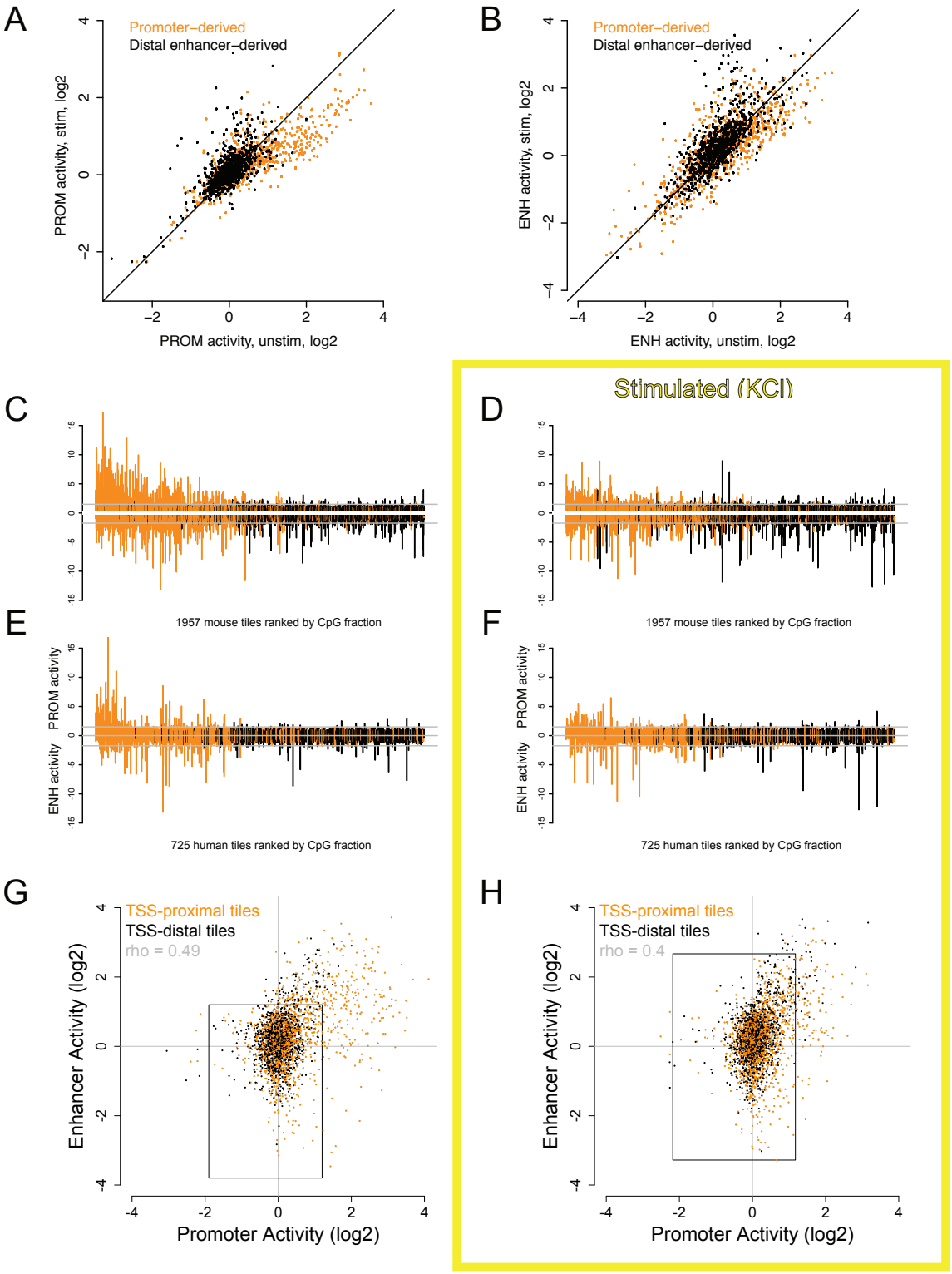
Supplemental Fig S9. CpG content and promoter activity. (A-B) Partial correlation coefficients for CpG and G+C content versus promoter activity, enhancer activity, and the ratio of these activities for mouse (A) and human (B) genomic tiles, respectively. Partial correlation is the correlation of either CpG or G+C with activity while controlling for G+C or CpG respectively. **(C)** Promoter activities for 3 synthetic tiles designed with ApG (GpA), CpA (ApC), and CpG (GpC) dinucleotide repeats. **(D)** Distribution of promoter activity for tiles (solid line) and reversed control tiles (dashed line) with > 0.087 CpGs per dinucleotide.



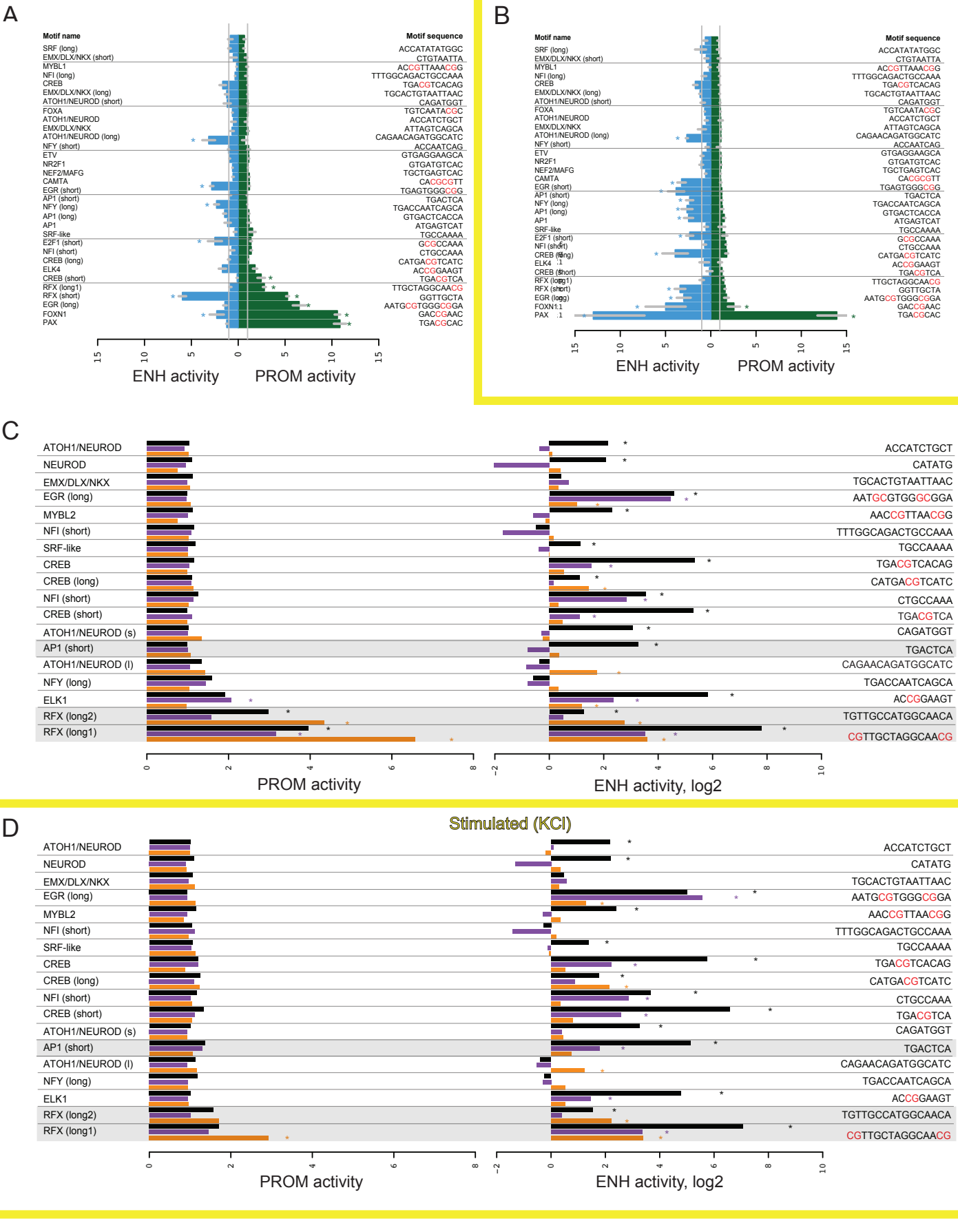
Supplemental Fig S10. Motifs enriched in tiles with promoter activity, compared to tiles lacking promoter activity. Tiles with promoter activity had > 2-fold higher promoter activity compared to negative control tiles. Tiles lacking promoter activity had < 1.2-fold activity compared to negative control tiles.



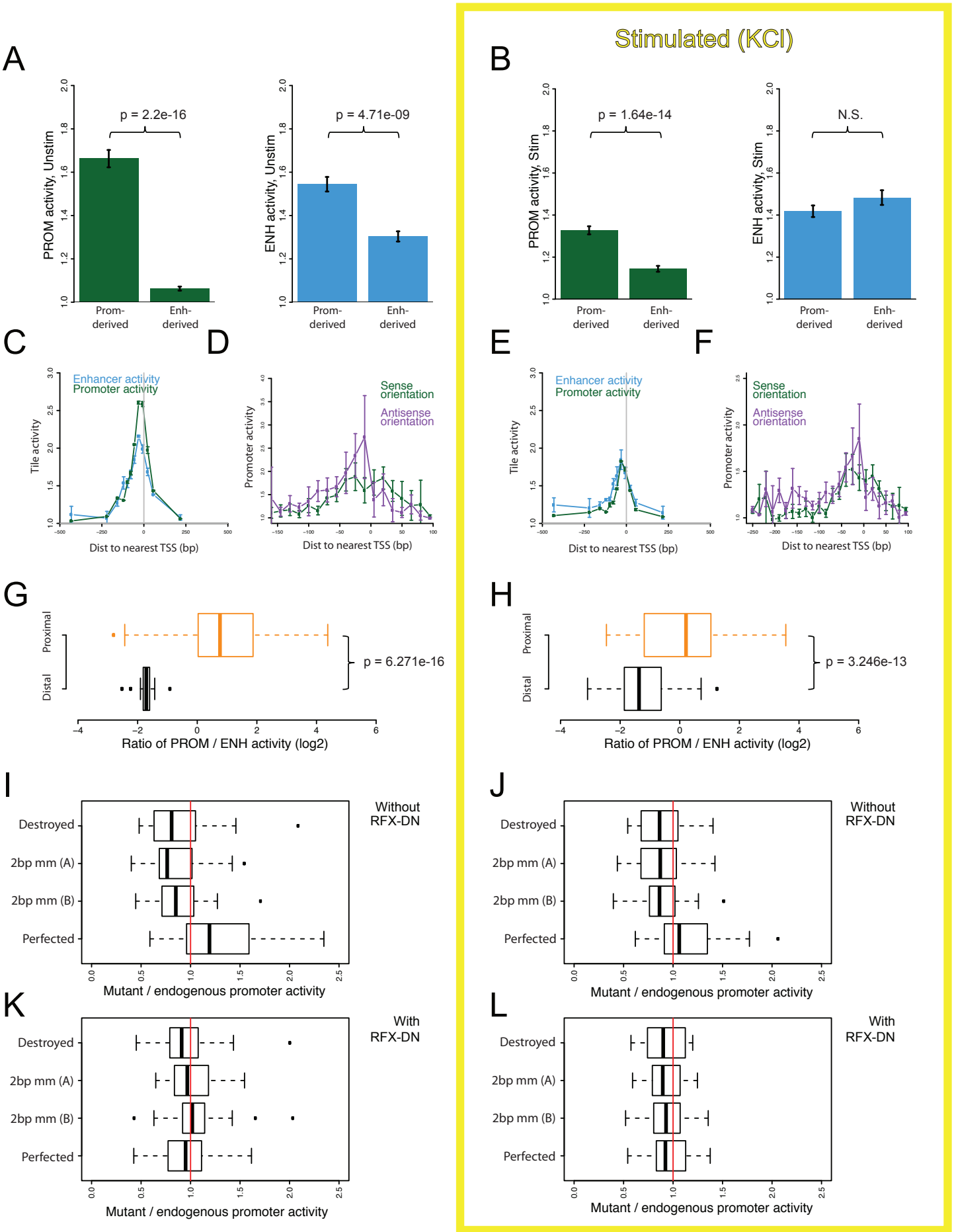
Supplemental Fig S11. Luciferase reporter comparison to MPRA, including KCl depolarization. (A) Promoter activity by luciferase. **(B)** Enhancer activity by luciferase. **(C)** Promoter activity by MPRA. **(D)** Enhancer activity by MPRA. Four artificial tiles from Figure 5B are analyzed in both experiments: RFX with spacer TTATTTTAAGA, AP1 with spacer CCCGCGCTGCC, MYBL2 with spacer TTATTTTAAGA, and NFYshort with spacer CCCGCGCTGCC. **(E)** Same as Fig. 4E but for the unstimulated condition. **(F)** KCl-stimulated condition.



Supplemental Fig S12. Effects of KCl-depolarization on MPRA activities. (A) Changes in promoter activity for promoter-derived and distal enhancer-derived tiles from TN03 with KCl depolarization. **(B)** Same as (A) but for enhancer activity. **(C-F)** Same as Figure 4A,C,D but with unstimulated at left and KCl-depolarized at right, instead of the maximal activity from either condition. **(G-H)** Same as Figure 2D but with unstimulated at left and KCl-depolarized at right, instead of the maximal activity from either condition.


A
Motif name
Motif sequence
SRF (long) ACCATATATGGC
EMX/DLX/NKX (short) CTGTAATTA
MYBL1 ACCGTTAAACGG
NFI (long) TTTGGCAGACTGCCAAA
CREB TGACGTCACAG
EMX/DLX/NKX (long) TGCACTGTAATTAAC
ATOH1/NEUROD (short) CAGATGGT
FOXA TGTCAATACGC
ATOH1/NEUROD ACCATCTGCT
EMX/DLX/NKX ATTAGTCAGCA
ATOH1/NEUROD (long) CAGAACAGATGGCATC
NFY (short) ACCAATCAG
ETV GTGAGGAAGCA
NR2F1 GTGATGTCAC
NEF2/MAFG TGCTGAGTCAC
CAMTA CACGCGTT
EGR (short) TGAGTGGGCGG
AP1 (short) TGACTCA
NFY (long) TGACCAATCAGCA
AP1 (long) GTGACTCACCA
AP1 ATGAGTCAT
SRF-like TGCCAAAA
E2F1 (short) GCGCCAAA
NFI (short) CTGCCAAA
CREB (long) CATGACGTCATC
ELK4 ACCGGAAGT
CREB (short) TGACGTCA
RFX (long1) TTGCTAGGCAACG
RFX (short) GGTTGCTA
EGR (long) AATGCGTGGGCGGA
FOXN1 GACCGAAC
PAX TGACGCAC
15 10 5 0 5 10 15
ENH activity
PROM activity
B
Stimulated (KCl)
Motif name
Motif sequence
SRF (long) ACCATATATGGC
EMX/DLX/NKX (short) CTGTAATTA
MYBL1 ACCGTTAAACGG
NFI (long) TTTGGCAGACTGCCAAA
CREB TGACGTCACAG
EMX/DLX/NKX (long) TGCACTGTAATTAAC
ATOH1/NEUROD (short) CAGATGGT
FOXA TGTCAATACGC
ATOH1/NEUROD ACCATCTGCT
EMX/DLX/NKX ATTAGTCAGCA
ATOH1/NEUROD (long) CAGAACAGATGGCATC
NFY (short) ACCAATCAG
ETV GTGAGGAAGCA
NR2F1 GTGATGTCAC
NEF2/MAFG TGCTGAGTCAC
CAMTA CACGCGTT
EGR (short) TGAGTGGGCGG
AP1 (short) TGACTCA
NFY (long) TGACCAATCAGCA
AP1 (long) GTGACTCACCA
AP1 ATGAGTCAT
SRF-like TGCCAAAA
E2F1 (short) GCGCCAAA
NFI (short) CTGCCAAA
CREB (long) CATGACGTCATC
ELK4 ACCGGAAGT
CREB (short) TGACGTCA
RFX (long1) TTGCTAGGCAACG
RFX (short) GGTTGCTA
EGR (long) AATGCGTGGGCGGA
FOXN1 GACCGAAC
PAX TGACGCAC
15 10 5 0 5 10 15
ENH activity
PROM activity
C
ATOH1/NEUROD ACCATCTGCT
NEUROD CATATG
EMX/DLX/NKX TGCACTGTAATTAAC
EGR (long) AATGCGTGGGCGGA
MYBL2 AACCGTTAACGG
NFI (short) TTTGGCAGACTGCCAAA
SRF-like TGCCAAAA
CREB TGACGTCACAG
CREB (long) CATGACGTCATC
NFI (short) CTGCCAAA
CREB (short) TGACGTCA
ATOH1/NEUROD (s) CAGATGGT
AP1 (short) TGACTCA
ATOH1/NEUROD (l) CAGAACAGATGGCATC
NFY (long) TGACCAATCAGCA
ELK1 ACCGGAAGT
RFX (long2) TGTTGCCATGGCAACA
RFX (long1) CGTTGCTAGGCAACG
0 2 4 6 8
PROM activity
−2 0 2 4 6 8 10
ENH activity, log2
D
Stimulated (KCl)
ATOH1/NEUROD ACCATCTGCT
NEUROD CATATG
EMX/DLX/NKX TGCACTGTAATTAAC
EGR (long) AATGCGTGGGCGGA
MYBL2 AACCGTTAACGG
NFI (short) TTTGGCAGACTGCCAAA
SRF-like TGCCAAAA
CREB TGACGTCACAG
CREB (long) CATGACGTCATC
NFI (short) CTGCCAAA
CREB (short) TGACGTCA
ATOH1/NEUROD (s) CAGATGGT
AP1 (short) TGACTCA
ATOH1/NEUROD (l) CAGAACAGATGGCATC
NFY (long) TGACCAATCAGCA
ELK1 ACCGGAAGT
RFX (long2) TGTTGCCATGGCAACA
RFX (long1) CGTTGCTAGGCAACG
0 2 4 6 8
PROM activity
−2 0 2 4 6 8 10
ENH activity, log2

Supplemental Fig S13. Same as Figure 5, except separate presentation of unstimulated and KCl-depolarized activities. (A) Unstimulated. **(B)** Depolarized. **(C)** Unstimulated. **(D)** Depolarized. * indicates significantly activity by FDR < 0.1 and mean activity > 2. *P* values from one-tailed *t*-tests: PAX promoter activity decrease with KCl, p=0.1. PAX enhancer activity increase with KCl, p=0.03; EGR_long promoter activity decrease, p=0.002, EGR_long enhancer activity increase, p=0.02; FOXN1 promoter activity decrease, p=0.0004; FOXN1 enhancer activity increase, p=0.04.


Stimulated (KCl)
A
B
p = 2.2e-16
p = 4.71e-09
p = 1.64e-14
N.S.
PROM activity, Unstim
ENH activity, Unstim
PROM activity, Stim
ENH activity, Stim
Prom-derived
Enh-derived
C
D
E
F
Enhancer activity
Promoter activity
Tile activity
Promoter activity
Sense orientation
Antisense orientation
Dist to nearest TSS (bp)
G
H
Proximal
Distal
p = 6.271e-16
p = 3.246e-13
Ratio of PROM / ENH activity (log2)
I
J
K
L
Without RFX-DN
With RFX-DN
Destroyed
2bp mm (A)
2bp mm (B)
Perfected
Mutant / endogenous promoter activity

Supplemental Fig S14. Separate plots for unstimulated and KCl-stimulated MPRA activities as observed with TN03 and TN05 libraries. (A-B) Mean promoter and enhancer activities of promoter-derived and distal enhancer-derived tiles (TN03) in unstimulated (A) and KCl-stimulated (B) conditions (p-value from two-tailed Student's *t*-test). **(C, E)** Mean promoter and enhancer activities as a function of the distance from the center of the MPRA tile (TN03) to the nearest RefSeq TSS. Unstimulated condition is in (C), KCl-stimulated in (E). Negative distances correspond to tile locations upstream of the TSS. **(D, F)** Mean unstimulated or KCl-stimulated promoter activity for tiles (TN03) cloned into the MPRA promoter array in the sense versus antisense orientation relative to their endogenous mRNA TSS. In both conditions, the distributions were found to be indistinguishable ($p > 0.05$, KS test). **(G, H)** The ratio of unstimulated and KCl-stimulated promoter-to-enhancer activities of tiles derived from promoter and distal enhancer loci (TN03, p-value from two-tailed *t*-test). **(I-J)** Ratios of unstimulated and stimulated mutant to control promoter activities (TN05) using MPRA, with median and 25^{th}-75^{th} percentiles shown as bars and boxes. Mutants are defined in Fig. 7B. **(K-L)** As in I-J but from RFX dominant negative-transduced neurons. Error bars in (A-C) and (I-L) indicate SEM for n=2 and n=3 independent biological replications, respectively. Promoter and enhancer activities are defined as in Fig. 2B.

SUPPLEMENTAL TABLE LEGENDS

Supplemental Table S1. Promoter and enhancer activities for genomic MPRA sequences (TN03 library). The first tab contains all MPRA data for mouse and human genomic tiles in Figures 1-3. Rows are individual MPRA tiles, with data aggregated over all corresponding barcodes. Columns are descriptors of each tile, including read counts for each DNA and cDNA sample. The second tab is a key describing all column headers in the data table.

Supplemental Table S2. Motif enrichments for tiles with > 0.087 CpGs per dinucleotide. Corresponds to data in Figure 3E.

Supplemental Table S3. All MPRA promoter and enhancer activities for Figure 4 (TN04 library). Rows are individual MPRA genomic tiles, with data aggregated over all corresponding barcodes. Columns are descriptors of each tile, including read counts for each DNA and cDNA sample. The first tab is a key describing all column headers in the data table. Subsequent tabs contain data corresponding to Figures 4A, 4B, and S5C.

Supplemental Table S4. Motif enrichments within annotated promoters and distal CBP enhancers compared to flanking sequences. Corresponds to Figure 5A.

Supplemental Table S5. All MPRA promoter and enhancer activities for Figure 6 (TN05 library). Rows are individual MPRA tiles, with data aggregated over all corresponding barcodes. Columns are descriptors of each tile, including read counts for

each DNA and cDNA sample. The first tab is a key describing all column headers in the data table. The second tab contains data corresponding to Figure 6.

Supplemental Table S6. Description of raw and processed sequencing data. Raw and processed sequencing data are available through the NCBI Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) under accession number GSE77213.