SUPPLEMENTAL METHODS

Table of Contents

Design of mouse and human genomic tiling MPRA libraries (TN03)

Mouse genomic DNA fragments (TN03a)

To select loci, we began with 41k CBP-bound loci (Kim et al. 2010). We limited ourselves to the 26k that align to the human genome (minimum of 30bp in alignment), using a 200bp mouse fragment. Of these only 12k are DNase-accessible in human fetal brain (sample E081-DNase.narrowPeak) (Roadmap Epigenomics Consortium et al. 2015), and have SNPs within the human DNase-accessible region (Thousand Genomes Phase 2, http://www.1000genomes.org/). Most of this reduction was due to a lack of a fetal brain DNase accessibility. We then took the top differentially H3K27ac acetylated loci in mouse neurons upon KCl depolarization (Malik et al. 2014), with the additional requirement that half of the loci be > 1kb from the nearest annotated TSS (distal enhancers). Admittedly, the human fetal DNase accessibility and SNP requirements are not strictly relevant to the questions we are asking here. At the same time, we know of no reason why these criteria should bias the relative promoter and enhancer activities we observe.

For the mouse sequences, we inserted all tiles whose nearest annotated TSS is on the positive strand into our reporter assay in the “sense” orientation. We inserted all mouse tiles whose nearest annotated TSS is on the negative strand into our reporter assay in the “antisense” orientation. Negative control sequences are described in the human section for TN03.

Human genomic DNA fragments (TN03b)

We mapped the 200bp fragments from mouse to human using multiple sequence alignments built from placental mammals (http://hgdownload.soe.ucsc.edu/goldenPath/mm10/multiz60way/) (Blanchette et al. 2004) using a procedure (Gjoneska et al. 2015) that requires regions to be > 30bp and <=400 bp to avoid including large insertions. For each mouse locus, we also included the corresponding human locus.

For human sequences, all tiles whose nearest annotated TSS was on the positive genomic strand was inserted into our reporter assay in the sense orientation. All tiles whose

nearest annotated TSS was on the negative strand were reversed without complementation, creating nonsense sequences that serve as controls and also enabled testing of the effects of CpG and G+C content.

Design of MPRA library TN04, artificial sequences designed to contain specific motifs

We identified the top 40 motifs from (Xie et al. 2007; 2005) enriched at CBP-bound distal enhancers and gene promoters (Kim et al. 2010): DGTTGCCATGGMRACH, TGACGTCA, TGACKCAC, RYGATGTCAY, CTATTTWTAG, TAAWWATAG, RTGASTMA, GGTTGCYA, CYTAGCAAC, KTTGRCGC, CCATYKGCT, CTGCCAAA, NMRMMCAKMTGKYMYS, WCGCGTY, STGTMATTA, CTAWWWATA, NYRMDRTAATTARY, TGACRTCACAR, TGCSRAAA, TATGYAAATKMK, MTGACGT, CAGMTGKSAAA, YTCCRKMTGT, KWTAATTAG, MCAGATGK, RNCAGRKGGCA, TGCCAAR, TAATTRSY, RTGACWMAGCA, STAATTA, YTATTTNTGG, GCGYGGGYG, MAGATGGY, YNATGGAARC, GACGCAMM, YGACYAAT, CTTWGTCAY, TGCWRATT, MTTAGTMAKMR.

We selected and instantiated these motifs with the goal of maximizing enrichment of the instantiated versions, as well as including CpG-containing and CpG-lacking versions of motifs where possible. We also in some cases performed multiple instantiations of a motif or included instantiations of related motifs to test specific mini-hypotheses. For each instantiated motif, we included it as described in the relevant figures.

Design of MPRA library TN05, interrogation of mutations in RFX motifs

To investigate the function of RFX motifs on the promoter activity of distal enhancers, we synthesized a 138 bp MPRA library called TN05. The library was synthesized to contain 175 unmutated tiles. Of the unmutated tiles, 44 were inducible in TN03, 41 were active in TN03, 40 were inactive negative controls from TN03, and 50 were chosen from mouse genomic DNA based on the presence of RFX motifs, the binding of CBP, and the binding of H3K27ac.

We mutated the RFX motif shown in Figure 7B that matched at PWM threshold 0.75 (RFX5) at:

~38 RFX sites in 28 inducible tiles
~36 RFX sites in 27 active tiles
~36 scrambled control RFX sites in 36 inducible tiles
~27 scrambled control RFX sites in 33 active tiles

For each of the sites indicated above, we created targeted point mutations:

For mutant1, we switched the base with #1 and #4 highest info content from A->T, C->G, G->C, or T->A. For mutant2, we did the same for the bases with the 2nd and 3rd highest information content. For the “destroyed” motifs, we swapped every nucleotide A->T, C->G, G->C, or T->A. For the “perfected” motifs, we made any changes necessary to match the motif maximally.

Generation of Promoter and Enhancer Activity-Specific AAV MPRA vectors

Construction of promoter activity-specific vectors

We constructed *pTAN01*, an inverted terminal repeat (ITR) containing adeno-associated virus (AAV) screening vector based on existing massively parallel reporter assay (MPRA) technology (Melnikov et al. 2012). We started with pAAV-CAG-ArchT-tdTomato – a gift from Edward Boyden (Addgene plasmid #29778). We excised the CAG-ArchT-tdTomato insert using AatII (NEB R0117S) and replaced it with a synthetic polyA signal, a ccdB suicide cassette flanked by unique SfiI sites (used for directional cloning of synthesized tiles during library generation), followed by the SV40 late polyA signal.

Our promoter assay also requires an *ORF donor plasmid* pLess-GFP-APH-donor. We replaced the CAG-ArchT-tdTomato insert from pAAV-CAG-ArchT-tdTomato with a KpnI-XbaI flanked synthetic intron (pIRESpuro3, Clontech 631619), GFP, and APH (kanamycin resistance to minimize vector background during cloning) insert.

Construction of enhancer activity-specific vector

For our enhancer assay, we constructed *pTAN08*, an ITR containing AAV screening vector based on STARR-Seq (Arnold et al. 2013). We replaced the CAG-ArchT-tdTomato insert from pAAV-CAG-ArchT-tdTomato with a synthetic polyA signal, a 95 bp basal human FOS promoter (pFOS), the synthetic intron from above, GFP, and a ccdB suicide cassette flanked by unique SfiI sites, followed by the SV40 late polyA signal. pFos has sequence GCaCTCaTTCATAAAACGCTTGT**TATA**AAAGCAGTGGCTGCGGCGCCTCGTACTcCaACCG CATCTGCAGCGAGCAACTGAGAAGCCAAGACTGA. Annotated TSSs are indicated in lowercase. A candidate TATA box is bolded. The sequence has 6 CpGs for a CpG density of 0.06. XbaI is immediately upstream of the indicated pFos sequence in pTAN08

Construction of MPRA Libraries

TN03/TN05 Amplification

For libraries TN03 and TN05, genomic sequence tiles were synthesized with adapters 5'-ACTGGCCGCTTCACTG-3' on the 5' end and 5'-CACTGCGGCTCCTCA-3' on the 3' end. The synthesized oligonucleotides were then amplified in emulsion PCRs (emPCRs) using EURx Micellula DNA Emulsion & PCR Kit (EURx 3600) for 18 cycles according to manufacturer's protocol with primers 5'-ACTGGCCGCTTCACTG-3' and TGAGGAGCCGCAGTG. The purified emPCR product was used in a second emPCR reaction using primers 5'-GGCCTAACTGGCCGCTTCACTG-3', 5'-GTTTAAGGCCTCCGAGGCC-3', and 5'-GTTTAAGGCCTCCGAGGCCGACGCTCTTCCGATCTNNNNNNNNNNNNNNNNTCTAGAGGT ACCTGAGGAGCCGCAGTG-3' (where NNNNNNNNNNNNNNN represents degenerate nucleotides for barcoding) at a 10:9:1 ratio. This reaction adds SfiI sites (for directional cloning), KpnI-XbaI sites (for cloning ORF into promoter assay), a 16 bp random identifying sequence barcode, and a priming site for Illumina sequencing (Figure S1). The resulting product was then

digested with SfiI (NEB R0123S) at 50˚C for three hours, run on a 10% TBE gel (Novex EC62752), stained with SYBR Gold (Invitrogen S-11494), extracted, and precipitated.

TN04 Amplification

For library TN04, artificial sequence tiles were synthesized with adapter 5'-ACTGGCCGCTTCACTG-3' on the 5' end, as well as adapter 5'-GGTACCTCTAGANNNNNNNNNNAGATCGGAAGAGCGTCG-3' containing KpnI-XbaI sites, an 11 bp identifying sequence barcode, and a priming site for Illumina sequencing on the 3' end. The synthesized oligonucleotides were then amplified in one round of emPCRs using 5'-GCTAAGGGCCTAACTGGCCGCTTCACTG-3' and 5'-GTTTAAGGCCTCCGAGGCCGACGCTCTTCCGATCT-3' to add SfiI sites (Figure S1). The resulting product was digested with SfiI and purified as with TN03.

Promoter assay cloning, part I

For the promoter test, amplified oligos were cloned into vector pTAN01, which was digested with SfiI for three hours, run on a 1% agarose gel, and extracted and purified using QIAEX II Gel Extraction Kit (Qiagen 20021) according to manufacturer's protocol (Figure S2A). The linearized vector was ligated with precipitated barcoded tiles at a 1:2 molar ratio in five 20 µl reactions with T4 DNA Ligase (NEB M0202T) overnight at 16˚C. The five reactions were pooled and purified using Qiagen MinElute Reaction Cleanup Kit (Qiagen 28204) before electroporation into five 20 µl aliquots of Invitrogen MegaX DH10B Electrocompetent Cells (Invitrogen C6400-03). The electroporated cells were then pooled, recovered for 90 minutes at 30˚C, dilutions were plated on LB AMP plates (for counting colony forming units; CFUs), and incubated overnight at 30˚C. For *TN03*, 200,000 colonies were scraped the next day, transferred to 500 ml LB AMP media, placed in a 30˚C shaking incubator for four hours, and then isolated using a Qiagen CompactPrep Plasmid Maxi Kit (Qiagen 12863). For *TN04*, the rest of the recovered bacteria were transferred to 1000 ml LB AMP media, incubated overnight in a 30˚C shaking incubator, and isolated as before. The *TN03* library was sequenced to match synthesized tiles to added barcodes (see Matching Synthesized Tiles to Barcodes) first, and then both *TN03* and *TN04* libraries were used to complete the promoter assay screening library, and used to complete the enhancer assay screening library.

Promoter assay cloning, part II

To complete the promoter assay screening library, the KpnI-XbaI ORF fragment from pLess-GFP-APH-donor was ligated into the isolated library (Figure S2B). The isolated plasmid pool and pLess-GFP-APH-donor were digested with KpnI and XbaI (NEB R3142S; R0145S) for three hours, run on a 1% agarose gel, and extracted and purified using QIAEX II Gel Extraction Kit. The fragments were ligated and electroporated as before. Dilutions of recovered bacteria were plated on LB AMP + KAN plates (for counting colony forming units, CFUs), and incubated overnight at 30˚C. The rest of the recovered bacteria were transferred to 1000 ml LB AMP + KAN media and incubated overnight in a 30˚C shaking incubator. 5-10 million CFUs per library were collected. The library was isolated a Qiagen CompactPrep Plasmid Maxi Kit. All members

of the library should be: polyA signal-synthesized tile-synthetic intron-GFP-KAN-barcode-polyA signal.

Enhancer assay cloning

To generate the enhancer assay screening library, the barcoded tiles from the original plasmid pool (before addition of the ORF for promoter assay) was ligated into pTAN08 (Figure S1). The original pool and pTAN08 were digested with SfiI for three hours, run on a 1% agarose gel, and extracted and purified using QIAEX II Gel Extraction Kit. The fragments were ligated and electroporated as before. Dilutions of recovered bacteria were plated on LB AMP plates (for counting CFUs) and incubated overnight at 30˚C. The rest of the recovered bacteria were transferred to 1000 ml LB AMP media and incubated overnight in a 30˚C shaking incubator. 5-10 million CFUs per library were collected.The library was isolated as before. All members of the enhancer library should be: polyA signal-pFOS-synthetic intron-GFP-synthesized tile-barcode-polyA signal.

Creating a tile-barcode key (TN03, TN05)

For the genomic DNA library (TN03) and its derivative (TN05), random barcodes were added after oligo amplification, necessitating the creation of a key to relate barcodes to genomic tiles (Figure S2A). To produce this key, the cloned TN03 library was sequenced to create a barcode-tile key to facilitate cDNA analysis based on barcode sequencing alone. For this sequencing, the original plasmid pool was amplified in an emPCR using 1 ng of template, 5'-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT-3', and 5'-CAAGCAGAAGACGGCATACGAGATXXXXXXGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCACTGGCCGCTTCACTG-3' (where XXXXXX represents sequencing indices) to add adapters for Illumina sequencing. The resulting product was run on a 10% TBE gel, stained with SYBR Gold, extracted, precipitated, and paired-end sequenced at 2x300 on a MiSeq. This sequencing yielded long reads that contained both the

Enhancer tiles were aligned using BWA to the sequences sent for synthesis. Each barcode was matched to the enhancer to which its corresponding sequence aligned best. Barcodes were discarded if they matched to more than one synthesized tile. Barcode-tile pairs that had fewer than 2-5 sequencing reads supporting them were also excluded. Of the pairs that were left, we kept the alignment with the highest mapping quality as the canonical alignment and counted up all of the alignments supporting that barcode-enhancer pair.

Production of viral MPRA libraries

AAV1/AAV2 supernatant was produced as described by (Konermann et al. 2013) with modifications. Lenti-X 293T Cells (Clontech 632180) were grown in DMEM (Gibco 11995-065) + 10% FBS (Gibco 16000-085) + 1% penicillin/streptomycin (Gibco 15140-122) at 37˚C and 5% CO_2 and passaged daily. Cells were never grown beyond 85% confluence and number of passages was kept below 15. The day before transfection, 10^6 cells were passaged in

transfection media – DMEM + 10% Heat-Inactivated FBS (Gibco 10082-147) + 1% penicillin/streptomycin – and plated onto 15 cm dishes. Cells were transfected at 80% confluence. Polyethylenimine (PEI) “Max”, MW 40,000 (Polysciences 24765-2) was made up to 1 mg/ml solution in water, pH brought to 7.0, and filter sterilized for use as a transfection reagent. For AAV production, 20 µg pDF6 helper plasmid, 3.5 µg pAAV1 serotype plasmid, 3.5 µg pAAV2 serotype plasmid, 7 µg ITR containing vector (promoter or enhancer assay screening libraries), and 105 µg of PEI were added to 2 ml plain DMEM, mixed and incubated at room temperature for 15 minutes. This mix was added to 23 ml of pre-warmed transfection media, mixed, and used to replace existing media on cells. After 48 hours in the incubator, the supernatant was collected, spun at 500 xg and 4˚C for 5 minutes to pellet debris, filtered through a 0.45 µm PES filter (Nalgene 16211-068), and frozen in aliquots at -80˚C.

Primary Cortical Neuron Cultures

Primary cultures were prepared from embryonic day 16 CD-1 mouse embryos (Charles River Labs). Cortices were dissected in ice cold dissection media (DM) – HBSS (Invitrogen 14175-145) + 10 mM $MgCl_2$ + 1 mM kynurenic acid (Sigma K3375) + 10 mM HEPES, pH to 7.2. Cortices were then digested in 5 ml of 3.5% papain (Worthington LS003126) in DM for 3 minutes at 37˚C, washed twice with 3 mls 1 mg/ml trypsin inhibitor (Sigma T9253) in DM, and incubated in 5 ml trypsin inhibitor solution for 4 minutes at 37˚C. Cortices were then triturated in plating media – DMEM + 10% FBS + 1% gentamicin (Sigma G1397-10ML), and 15 x 10^6 cells were plated on poly-ornithine (Sigma P3655) coated 15 cm plates. After 3 hours, the media was changed to and cells were maintained in complete Neurobasal (CNB) – Neurobasal (Gibco 12348-017) + 2% B-27 Supplement (Gibco 17504-044) + 1% GlutaMAX Supplement (Gibco 35050-061) + 1% penicillin/streptomycin.

AAV transduction was performed on DIV4 with the addition of 15 ml AAV supernatant to existing media on cells, and a complete media change to fresh CNB was performed 24 hours later. For depolarization of neurons, DIV6 neurons were quieted overnight in 1mM tetrodotoxin (TTX, Tocris 1069) and 100 mM DL-2-amino-5-phosphonopentanoic acid (AP5, Tocris 3693). On DIV7, neurons were depolarized with a media change to CNB + 55 mM KCl and incubated for 0 or 12 hours.

Preparing cells for sequencing of MPRA cDNA barcodes

Samples were collected and stored in 15 ml Trizol Reagent (Invitrogen 15596-018). For RNA extraction, 3 ml of chloroform was added to each tube, vortexed briefly, and centrifuged at 12,000 xg for 15 minutes at 4˚C for phase separation. The RNA containing aqueous layer was then transferred to a new tube and an equal volume of 100% EtOH was added. RNA purification was completed using a Qiagen RNeasy Midi Kit (Qiagen 75142) following on-column DNase treatment with RNase-Free DNase (Qiagen 79254) according to manufacturer’s protocol. mRNA selection was then performed using the Poly(A)Purist MAG Kit (Ambion AM1922) according to manufacturer’s protocol. An in-solution DNase treatment was then performed using RNase Free

DNase I (Roche 4716728001) for 15 minutes at 37˚C. The mRNA was cleaned up and eluted in 12 µl of water using a Qiagen MinElute RNeasy Kit (Qiagen 74204). Up to 2 µg of mRNA in 10 µl of water was converted to cDNA using a High-Capacity RNA-to-cDNA Kit Supermix (Applied Biosystems 4387406).

For the promoter assay, barcode sequences were amplified from cDNA and isolated plasmid DNA samples using NEB Phusion HF Master Mix (NEB M0531S) according to manufacturer's protocol for 26 cycles using 5'-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT-3', and 5'-CAAGCAGAAGACGGCATACGAGATXXXXXXGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCAGGACATAGCGTTGGCTACC-3' (where XXXXXX represents sequencing indices) to add adapters for Illumina sequencing (Figure S2B). For the enhancer assay, barcode sequences were amplified as before, but using 5'-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT-3', and 5'-CAAGCAGAAGACGGCATACGAGATXXXXXXGTGACTGGAGTTCAGACGTGT-3'. The resulting product was run on a 10% TBE gel, stained with SYBR Gold, extracted, precipitated. The final libraries were checked on Agilent Bioanalyzer 2100 chips before sequencing on a HiSeq.

Luciferase assays

Construction of luciferase assays

To construct the enhancer assay, tiles were individually cloned into pTAN02, an ITR-containing AAV screening vector containing minimal human pFos upstream of the Firefly luciferase gene, using KpnI and XhoI sites. The promoter assay constructs were produced in the same way as in the MPRA experiments, but using an altered pLess-GFP-APH-donor plasmid in which the GFP was replaced with Firefly luciferase from pGL3-Basic (Promega) using BmgB1 and BspE1 sites.

Transfection of luciferase constructs

Primary cortical neuron cultures (250,000 neurons) were transfected using PEI (3:1, PEI:DNA mass ratio) on DIV5. The Firefly luciferase constructs were co-transfected with an internal control Renilla luciferase construct, pTK-RN, at a fixed mass ratio of 9:1 (450ng Firefly construct, 50ng Renilla construct). Each experiment was run over three replicates, and each condition run in triplicate. The silencing (DIV6) and stimulating (DIV7) of cultures were performed as in the MPRA experiments. Cultures were collected on the night of DIV7, and prepared using the Dual-Luciferase Reporter Assay System (Promega) according to the manufacturer's protocol. The lysate was assayed over a 10 second period using the GloMax 20/20 Single Tube Luminometer (Promega), and the luciferase activity was calculated as a ratio of the Firefly to Renilla output values. Each value was normalized to the average of all calculated ratios for the two negative control tiles (MYBL2, NFY_short).

Processing of MPRA read counts to transcriptional activities

Quality control

Before analysis of any MPRA data, we removed all sequence tiles that had fewer than 20 reads in any sample.

Definition of enhancer and promoter activities

We quantified enhancer and promoter activities by performing two normalization operations on raw MPRA read counts. We first normalized each tile's cDNA read count to its corresponding DNA read count, then divided that ratio by the corresponding ratio for the negative controls to obtain enhancer or promoter activity. For the genomic tile libraries, the controls were 498 "nonsense" sequences that show no indication of encoding enhancer or promoter activity (Supplemental Fig S3A-B). After normalization, enhancer activity represents a fold-increase over the activity of the basal p*Fos* promoter.
Promoter activity reflects a fold increase over background promoter activity, which is likely generated from the AAV inverted terminal repeats (Flotte et al., 1993; Haberman et al., 2000). Activity values of one (zero on a log scale) are indistinguishable from negative controls.

Computing FDRs

Empirical p-values were calculated as the fraction of the negative control activities that fall at or above a given experimental enhancer or promoter activity. Each p-value was then converted to a corresponding q-value using the p.adjust function in the *stats* package in R (Team RC, 2013).

Motifs enriched in selected MPRA tiles

We began with a list of 1845 motifs in 727 clusters (Xie et al. 2005; 2007). To identify enrichment of motifs in in selected tiles (Figure 4E), we required > 2.5-fold enrichment and a Bonferroni adjusted binomial p-value of < 0.05. For tiles with more enhancer than promoter activity, we compared two tile sets. The first had promoter activity > 1.8-fold and enhancer activity < 1.3-fold. The second had promoter activity < 1.2-fold and enhancer activity > 2-fold. AP1 was the only motif enriched in set 2 compared to set 1. For high CpG vs reversed high CpG tiles, we required > 0.087 CpGs per dinucleotide.

Motifs enriched at gene promoters versus distal CBP enhancers

We began with 1225 motifs (Xie et al. 2005; Jolma et al. 2013), with only 1 representative motif from each Xie *et al.* cluster. The enrichment of motifs at all gene promoters and 12k CBP-bound distal enhancers was determined as the ratio of the occurrences of a given motif within a

promoter or distal enhancer (center 200 bp) to the counts within 800 bp of flanking sequences (-2400 to 2000 bp and +2000 to +2400 bp). Only those motifs with at least an enrichment ratio of at least 2 at either promoters or distal enhancers and at least 10 sequencing counts in promoters, distal enhancers, and flanking sequences for both passed filter.

5' splice signal (5pSS) identification

For identification of 5pSSs we used the sequence RGGTRAG.

Correlation of CAGE tags with promoter activity

We obtained mouse neuron CAGE data (FANTOM Consortium and the RIKEN PMI and CLST (DGT) et al. 2014) and computed Pearson's correlation coefficient between promoter activity of mouse genomic tiles (TN03) and the CAGE tag abundance for tags found within the corresponding genomic sequences.

ChIP-Seq and ChIP-Seq peak analysis

We performed ChIP-Seq in RFX-DN transduced neurons using an anti-MYC A14 antibody (Santa Cruz, sc-789) after 0 or 1 hour of depolarization in 55mM KCl as described in (Kim et al. 2010), except performing sonication on a Covaris. As a control, we also performed ChIP-Seq in untransduced neurons, detecting orders of magnitude fewer peaks (table below). We aligned using Bowtie and called peaks as in (Kim et al. 2010).

Each sample was performed in biological duplicate. We distinguished distal enhancers and gene promoters based on a distance threshold of 1 kb from 25,562 RefSeq TSSs. For testing significance of promoter or enhancer enrichment with a hypergeometric test, we assumed 345,000 possible distal binding sites, based on the number of cerebrum DHS loci from ENCODE minus RefSeq TSSs.

TF	Promoters	Distal enhancers	Total
RFX, Rep1 (-RFX-DN)	NA	NA	42
RFX, Rep2 (-RFX-DN)	NA	NA	11
RFX, 0h Rep1 (+RFX-DN)	8967	146990	
RFX, 0h Rep2 (+RFX-DN)	8926	113985	
RFX, 1h Rep1 (+RFX-DN)	6684	67699	
RFX, 1h Rep2 (+RFX-DN)	9831	104824	
RFX_0-1hr mean	8602	108374	
CREB (Kim et al. 2010)	845	165	
FOS (Malik et al. 2014)	12062	532	

Detection of eRNA expression

At the subset of distal enhancers for which eRNA expression can be assessed unambiguously by RNA-seq (the 118 extragenic ones), 117 have detectable eRNA expression. We defined eRNA-expressing extragenic enhancers as those with at least one RNA-Seq read after 1hr KCl stimulation (Kim et al. 2010).

Analysis of DHS sizes

We compared the widths of distal enhancer versus promoter DHS peaks from 487 mouse cerebrum DHS peaks corresponding to loci we tiled in array TN03. We used DHS data from wgEncodeUwDnaseCerebrumC57bl6MAdult8wksHotspotsRep1 downloaded from the UCSC Genome Table Browser.

Comparison of eRNA expression at distal enhancers with or without RFX motifs.

We used the RFX5 motif (Jolma et al. 2013) and required a PWM match of at least 0.85 using the matchPWM function in BioStrings. We used RNA-Seq data (Kim et al. 2010) for eRNA quantification and considered 5117 distal enhancers defined in (Kim et al. 2010). 4491 of these lacked the RFX motif and had a median eRNA read count of 6, summed across three conditions: 0, 1, and 6 hours of KCl stimulation. 626 distal enhancers possessed the RFX motif and had a median read count of 11.

Plasmid sequences

>pTAN08 STARR TN03

ctgcgcgctcgctcgctcactgaggccgcccgggcaaagcccgggcgtcgggcgacctttggtcgcccggcctcagtgagcgagc
gagcgcgcagagagggagtggccaactccatcactaggggttccttgtagttaatgattaacccgccatgctacttatctacgtagcca
tgctctaggaagatcggaattcgcccttaagAATAAAATATCTTTATTTTCATTACATCTGTGTGTTGGTTTT
TTGTGTGAATCGATAGTACTAACATACGCTCTCCATCAAAACAAAACGAAACAAAACAAACT
AGCAAAATAGGCTGTCCCCAGTGCAAGTGCAGGTGCCAGAACATTTCTCTgcaatgcctatttaata
tgagcatgatcttagagttgtgtattCCCTCGAGATCTGCACTCATTCATAAAACGCTTGTTATAAAAGCA
GTGGCTGCGGCGCCTCGTACTCCAACCGCATCTGCAGCGAGCAACTGAGAAGCCAAGACT
GA<pFos>GCCGGCGGCCGGGCCCGAATTAATTCGCTGTCTGCGAGGGCCAGCTGTTGGG
GTGAGTACTCCCTCTCAAAAGCGGGCATGACTTCTGCGCTAAGATTGTCAGTTTCCAAAAA
CGAGGAGGATTTGATATTCACCTGGCCCGCGGTGATGCCTTTGAGGGTGGCCGCGTCCAT
CTGGTCAGAAAAGACAATCTTTTTGTTGTCAAGCTTGAGGTGTGGCAGGCTTGAGATCTGG
CCATACACTTGAGTGACAATGACATCCACTTTGCCTTTCTCTCCACAGGTGTCCACTCCCA
GGTCCAACTGCAGGTCGCCTGCAGGCTTAgccatggtgagcaagggcgaggagctgttcaccggggtggtgcc
catcctggtcgagctggacggcgacgtaaacggccacaagttcagcgtgtccggcgagggcgagggcgatgccacctacggcaa
gctgaccctgaagttcatctgcaccaccggcaagctgcccgtgccctggcccaccctcgtgaccaccctgacctacggcgtgcagtg

cttcagccgctaccccgaccacatgaagcagcacgacttcttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatctt
cttcaaggacgacggcaactacaagacccgcgccgaggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagg
gcatcgacttcaaggaggacggcaacatcctggggcacaagctggagtacaactacaacagccacaacgtctatatcatggccga
caagcagaagaacggcatcaaggtgaacttcaagatccgccacaacatcgaggacggcagcgtgcagctcgccgaccactacc
agcagaacaccccatcggcgacggccccgtgctgctgcccgacaaccactacctgagcacccagtccgccctgagcaaagacc
ccaacgagaagcgcgatcacatggtcctgctggagttcgtgaccgccgccgggatcactctcggcatggacgagctgtacaagtaat
aagcttgCCTTGCTCCTGCCGAGAAAGTATCGTGACTGGAGTTCAGACGTGTGCTCTTCCGAT
CTACTGCAAGACGAGGCAGCGCGGCTATGGCCTAACTGGCCGCTTCACTGNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNCACTGCGGCTCCTCAGGTACCTCTAGANNNNNNNNNNNNNNNAGA
TCGGAAGAGCGTCGGCCTCGGAGGCCGCTTCGAGCAGACATGATAAGATACATTGATGAG
TTTGGACAAACCACAACTAGAATGCAGTGAAAAAAATGCTTTATTTGTGAAATTTGTGATGC
TATTGCTTTATTTGTAACCATTATAAGCTGCAATAAACAAGTTAACAACAACAATTGCATTCA
TTTTATGTTTCAGGTTCAGGGGGAGGTGTGGGAGGTTTTTTAAAGCAAGTAAAACCTCTAC
AAATGTGGTAAAATCGATAAGGATCCGACCGATGCCCTTGAGAGCCTTCAACCCAGTCAGC
TCCTTCCGGTGGGCGCGGGGCATGACTATCGTCGCCGCACTTATGACTGTCTTgttaagggcg
aattcccgataaggatcttcctagagcatggctacgtagataagtagcatggcgggttaatcattaactacaaggaacccctagtgatg
gagttggccactccctctctgcgcgctcgctcgctcactgaggccgggcgaccaaaggtcgcccgacgcccgggctttgcccgggcg
gcctcagtgagcgagcgagcgcgcagccttaattaacctaattcactggccgtcgttttacaacgtcgtgactgggaaaaccctggcgt
tacccaacttaatcgccttgcagcacatccccctttcgccagctggcgtaatagcgaagaggcccgcaccgatcgcccttcccaacag
ttgcgcagcctgaatggcgaatgggacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgc
tacacttgccagcgccctagcgcccgctcctttcgctttcttcccttcctttctcgccacgttcgccggctttccccgtcaagctctaaatcgg
gggctcccttagggttccgatttagtgctttacggcacctcgaccccaaaaaacttgattagggtgatggttcacgtagtgggccatcgc
cctgatagacggtttttcgcccctttgacgttggagtccacgttctttaatagtggactcttgttccaaactggaacaacactcaaccctatctc
ggtctattcttttgatttataagggattttgccgatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaattttaaca
aaatattaacgtttataatttcaggtggcatctttcggggaaatgtgcgcggaacccctatttgtttatttttctaaatacattcaaatatgtatc
cgctcatgagacaataaccctgataaatgcttcaataatattgaaaaaggaagagtatgagtattcaacatttccgtgtcgcccttattcc
ctttttgcggcattttgccttcctgtttttgctcacccagaaacgctggtgaaagtaaaagatgctgaagatcagttgggtgcacgagtgg
gttacatcgaactggatctcaatagtggtaagatccttgagagttttcgccccgaagaacgttttccaatgatgagcacttttaaagttctg
ctatgtggcgcggtattatcccgtattgacgccgggcaagagcaactcggtcgccgcatacactattctcagaatgacttggttgagtac
tcaccagtcacagaaaagcatcttacggatggcatgacagtaagagaattatgcagtgctgccataaccatgagtgataacactgcg
gccaacttacttctgacaacgatcggaggaccgaaggagctaaccgcttttttgcacaacatgggggatcatgtaactcgccttgatcg
ttgggaaccggagctgaatgaagccataccaaacgacgagcgtgacaccacgatgcctgtagtaatggtaacaacgttgcgcaaa
ctattaactggcgaactacttactctagcttcccggcaacaattaatagactggatggaggcggataaagttgcaggaccacttctgcg
ctcggcccttccggctggctggtttattgctgataaatctggagccggtgagcgtgggtctcgcggtatcattgcagcactggggccaga
tggtaagccctcccgtatcgtagttatctacacgacggggagtcaggcaactatggatgaacgaaatagacagatcgctgagatagg
tgcctcactgattaagcattggtaactgtcagaccaagtttactcatatatactttagattgatttaaaacttcatttttaatttaaaaggatcta
ggtgaagatccttttgataatctcatgaccaaaatcccttaacgtgagttttcgttccactgagcgtcagaccccgtagaaaagatcaaa
ggatcttcttgagatcctttttttctgcgcgtaatctgctgcttgcaaacaaaaaaaccaccgctaccagcggtggtttgtttgccggatcaa
gagctaccaactctttttccgaaggtaactggcttcagcagagcgcagataccaaatactgtccttctagtgtagccgtagttaggccac
cacttcaagaactctgtagcaccgcctacatacctcgctctgctaatcctgttaccagtggctgctgccagtggcgataagtcgtgtctta
ccgggttggactcaagacgatagttaccggataaggcgcagcggtcgggctgaacgggggttcgtgcacacagcccagcttgga
gcgaacgacctacaccgaactgagatacctacagcgtgagctatgagaaagcgccacgcttcccgaagggagaaaggcggaca

ggtatccggtaagcggcagggtcggaacaggagagcgcacgagggagcttccagggggaaacgcctggtatctttatagtcctgtc
gggtttcgccacctctgacttgagcgtcgatttttgtgatgctcgtcaggggggcggagcctatggaaaaacgccagcaacgcggcctt
tttacggttcctggccttttgctgcggttttgctcacatgttctttcctgcgttatcccctgattctgtggataaccgtattaccgcctttgagtgag
ctgataccgctcgccgcagccgaacgaccgagcgcagcgagtcagtgagcgaggaagcggaagagcgcccaatacgcaaacc
gcctctccccgcgcgttggccgattcattaatgcagctggcacgacaggtttcccgactggaaagcgggcagtgagcgcaacgcaat
taatgtgagttagctcactcattaggcaccccaggctttacactttatgcttccggctcgtatgttgtgtggaattgtgagcggataacaatttt
cacacaggaaacagctatgaccatgattacgccagatttaattaagg

>pTAN01 TN03 PromoterLess

gccgcgttgctggcgtttttccataggctccgccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccg
acaggactataaagataccaggcgtttcccccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccggatacctgtcc
gcctttctcccttcgggaagcgtggcgctttctatagctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaagctgggctgt
gtgcacgaaccccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggtaagacacgacttatcgc
cactggcagcagccactggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttgaagtggtggcctaactacg
gctacactagaagaacagtatttggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaaac
aaaccaccgctggtagcggtggtttttttgtttgcaagcagcagattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttc
tacggggtctgacgctcagtggaacgaaaactcacgttaagggattttggtcatgagattatcaaaaaggatcttcacctagatccttttta
aattaaaaatgaagttttaaatcaatctaaagtatatatgagtaaacttggtctgacagttaccaatgcttaatcagtgaggcacctatctc
agcgatctgtctatttcgttcatccatagttgcctgactccccgtcgtgtagataactacgatacgggagggcttaccatctggccccagtg
ctgcaatgataccgcgagacccacgctcaccggctccagatttatcagcaataaaccagccagccggaagggccgagcgcagaa
gtggtcctgcaactttatccgcctccatccagtctattaattgttgccgggaagctagagtaagtagttcgccagttaatagtttgcgcaac
gttgttgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttcccaacgatcaaggcgagttac
atgatcccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggccgcagtgttatcactcatg
gttatggcagcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgagaat
agtgtatgcggcgaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgctcatcatt
ggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaactgatct
tcagcatcttttactttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggaataagggcgacacg
gaaatgttgaatactcatactcttccttttcaatattattgaagcatttatcagggttattgtctcatgagcggatacatatttgaatgtatttag
aaaaataaacaaatagggttccgcgcacatttccccgaaaagtgccacctaaattgtaagcgttaatattttgttaaaattcgcgttaa
atttttgttaaatcagctcatttttaaccaataggccgaaatcggcaaaatcccttataaatcaaaagaatagaccgagatagggttgag
tgttgttccagtttggaacaagagtccactattaaagaacgtggactccaacgtcaaagggcgaaaaaccgtctatcagggcgatgg
cccactacgtgaaccatcaccctaatcaagttttttggggtcgaggtgccgtaaagcactaaatcggaaccctaaagggagcccccg
atttagagcttgacggggaaagccggcgaacgtggcgagaaaggaagggaagaaagcgaaaggagcgggcgctagggcgct
ggcaagtgtagcggtcacgctgcgcgtaaccaccacacccgccgcgcttaatgcgccgctacagggcgcgtcccattcgccattca
ggctgcgcaactgttgggaagggcgatcggtgcgggcctcttcgctattacgccagctgcgcgctcgctcgctcactgaggccgcccg
ggcaaagcccgggcgtcgggcgacctttggtcgcccggcctcagtgagcgagcgagcgcgcagagagggagtggccaactccat
cactaggggttccttgtagttaatgattaacccgccatgctacttatctacgtagccatgctctaggaagatcgTACGGGAGGTAT
TGGACAGGCCGCAATAAAATATCTTTATTTTCATTACATCTGTGTGTTGGTTTTTTGTGTGAA
TCGATAGTACTAACATACGCTCTCCATCAAAACAAAACGAAACAAAACAAACTAGCAAAATA
GGCTGTCCCCAGTGCAAGTGCAGGTGCCAGAACATTTCTCTGGCCTAACTGGCCGCTTCA
CTGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNCACTGCGGCTCCTCA<TN03/TN05 only>GGTACC

<red=same seq as ENH assay>
ATCTCCCCCCgtaatatcttggcaccagatattttgagattcaGGCCC<cloning fragment>GAATTAATTCGCTGTCTGCGAGGGCCAGCTGTTGGGGTGAGTACTCCCTCTCAA
AAGCGGGCATGACTTCTGCGCTAAGATTGTCAGTTTCCAAAAACGAGGAGGATTTGATATT
CACCTGGCCCGCGGTGATGCCTTTGAGGGTGGCCGCGTCCATCTGGTCAGAAAAGACAAT
CTTTTTGTTGTCAAGCTTGAGGTGTGGCAGGCTTGAGATCTGGCCATACACTTGAGTGACA
ATGACATCCACTTTGCCTTTCTCTCCACAGGTGTCCACTCCCAGGTCCAACTGCAGGTCGC
CTGCAGGCTTAGAC<synthetic_intron>TTGACAATTAATCATCGGCATAGTATATCGGCATAG
TATAATACGA<pEM7>CTCACTATAGGAGGGCCACCATGAGCGGGGGCGAGGAGCTGTTC
GCCGGCATCGTGCCCGTGCTGATCGAGCTGGACGGCGACGTGCACGGCCACAAGTTCAG
CGTGCGCGGCGAGGGCGAGGGCGACGCCGACTACGGCAAGCTGGAGATCAAGTTCATCT
GCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTGGTGACCACCCTCTGCTACGGC
ATCCAGTGCTTCGCCCGCTACCCCGAGCACATGAAGATGAACGACTTCTTCAAGAGCGCC
ATGCCCGAGGGCTACATCCAGGAGCGCACCATCCAGTTCCAGGACGACGGCAAGTACAA
GACCCGCGGCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGG
GCAAGGACTTCAAGGAGGACGGCAACATCCTGGGCCACAAGCTGGAGTACAGCTTCAACA
GCCACAACGTGTACATCCGCCCCGACAAGGCCAACAACGGCCTGGAGGCTAACTTCAAGA
CCCGCCACAACATCGAGGGCGGCGGCGTGCAGCTGGCCGACCACTACCAGACCAACGTG
CCCCTGGGCGACGGCCCCGTGCTGATCCCCATCAACCACTACCTGAGCACTCAGACCAAG
ATCAGCAAGGACCCGCAACGAGGCCCGCGACCACATGGTGCTCCTGGAGTCCTTCAGCGC
CTGCTGCCACACCCACGGCATGGACGAGCTGTACAGGTCCGGAGGAGGCTCTGGCGGTG
GAAGCGGCATGATTGAACAAGATGGATTGCACGCAGGTTCTCCGGCCGCTTGGGTGGAGA
GGCTATTCGGCTATGACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCCGTGTTCC
GGCTGTCAGCGCAGGGGCGCCCGGTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCCTGA
ATGAACTGCAAGACGAGGCAGCGCGGCTATCGTGGCTGGCCACGACGGGCGTTCCTTGC
GCAGCTGTGCTCGACGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGT
GCCGGGGCAGGATCTCCTGTCATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCATGGC
TGATGCAATGCGGCGGCTGCATACGCTTGATCCGGCTACCTGCCCATTCGACCACCAAGC
GAAACATCGCATCGAGCGAGCACGTACTCGGATGGAAGCCGGTCTTGTCGATCAGGATGA
TCTGGACGAAGAACATCAGGGGCTCGCGCCAGCCGAACTGTTCGCCAGGCTCAAGGCGA
GCATGCCCGACGGCGAGGATCTCGTCGTGACCCATGGCGATGCCTGCTTGCCGAATATCA
TGGTGGAAAATGGCCGCTTTTCTGGATTCATCGACTGTGGCCGGCTGGGTGTGGCGGACC
GCTATCAGGACATAGCGTTGGCTACCCGTGATATTGCTGAAGAACTTGGCGGCGAATGGG
CTGACCGCTTCCTCGTGCTTTACGGTATCGCCGCTCCCGATTCGCAGCGCATCGCCTTCTA
TCGCCTTCTTGACGAGTTCTTCTGATAATTCTAGANNNNNNNNNNNNNNNN<barcode>AGA
TCGGAAGAGCGTCGGCCTCGGAGGCCAAGCTAGTCGGGCGGCCgcttcgagcagacatgataagat
acattgatgagtttggacaaaccacaactagaatgcagtgaaaaaaatgctttatttgtgaaatttgtgatgctattgctttatttgtaaccat
tataagctgcaataaacaagttaacaacaacaattgcattcattttatgtttcaggttcagggggagatgtgggaggttttttaaagcaagt
aaaacctctacaaatgtggtaaaatcgataaggatcttcctagagcatggctacgtagataagtagcatggcgggttaatcattaacta
caaggaacccctagtgatggagttggccactccctctctgcgcgctcgctcgctcactgaggccgggcgaccaaaggtcgcccgac
gcccgggctttgcccgggcggcctcagtgagcgagcgagcgcgcagctgcattaatgaatcggccaacgcgcggggagaggcgg
tttgcgtattgggcgctcttccgcttcctcgctcactgactcgctgcgctcggtcgttcggctgcggcgagcggtatcagctcactcaaagg
cggtaatacggttatccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgt
aaaaag

> RDJ_lucif_prom_assay_RA12_pTAN01_pLess (promoter test with luciferase)

gccgcgttgctggcgtttttccataggctccgccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccg
acaggactataaagataccaggcgtttcccccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccggatacctgtcc
gcctttctcccttcgggaagcgtggcgctttctcatagctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaagctgggctgt
gtgcacgaaccccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggtaagacacgacttatcgc
cactggcagcagccactggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttgaagtggtggcctaactacg
gctacactagaagaacagtatttggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaaac
aaaccaccgctggtagcggtggtttttttgtttgcaagcagcagattacgcgcagaaaaaaggatctcaagaagatcctttgatcttttc
tacggggtctgacgctcagtggaacgaaaactcacgttaagggattttggtcatgagattatcaaaaaggatcttcacctagatccttta
aattaaaaatgaagttttaaatcaatctaaagtatatatgagtaaacttggtctgacagttaccaatgcttaatcagtgaggcacctatctc
agcgatctgtctatttcgttcatccatagttgcctgactccccgtcgtgtagataactacgatacgggagggcttaccatctggccccagtg
ctgcaatgataccgcgagacccacgctcaccggctccagatttatcagcaataaaccagccagccggaagggccgagcgcagaa
gtggtcctgcaactttatccgcctccatccagtctattaattgttgccgggaagctagagtaagtagttcgccagttaatagtttgcgcaac
gttgttgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttcccaacgatcaaggcgagttac
atgatcccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggccgcagtgttatcactcatg
gttatggcagcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgagaat
agtgtatgcggcgaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgctcatcatt
ggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaactgatct
tcagcatcttttactttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggaataagggcgacacg
gaaatgttgaatactcatactcttccttttcaatattattgaagcatttatcagggttattgtctcatgagcggatacatatttgaatgtatttag
aaaaataaacaaatagggttccgcgcacatttccccgaaaagtgccacctaaattgtaagcgttaatattttgttaaaattcgcgttaa
atttttgttaaatcagctcatttttaaccaataggccgaaatcggcaaaatcccttataaatcaaaagaatagaccgagatagggttgag
tgttgttccagtttggaacaagagtccactattaaagaacgtggactccaacgtcaaagggcgaaaaaccgtctatcagggcgatgg
cccactacgtgaaccatcaccctaatcaagttttttggggtcgaggtgccgtaaagcactaaatcggaaccctaaagggagcccccg
atttagagcttgacggggaaagccggcgaacgtggcgagaaaggaagggaagaaagcgaaaggagcgggcgctagggcgct
ggcaagtgtagcggtcacgctgcgcgtaaccaccacacccgccgcgcttaatgcgccgctacagggcgcgtcccattcgccattca
ggctgcgcaactgttgggaagggcgatcggtgcgggcctcttcgctattacgccagctgcgcgctcgctcgctcactgaggccgcccg
ggcaaagcccgggcgtcgggcgacctttggtcgcccggcctcagtgagcgagcgagcgcgcagagagggagtggccaactccat
cactaggggttccttgtagttaatgattaacccgccatgctacttatctacgtagccatgctctaggaagatcgTACGGGAGGTAT
TGGACAGGCCGCAATAAAATATCTTTATTTTCATTACATCTGTGTGTTGGTTTTTTGTGTGAA
TCGATAGTACTAACATACGCTCTCCATCAAAACAAAACGAAACAAAACAAACTAGCAAAATA
GGCTGTCCCCAGTGCAAGTGCAGGTGCCAGAACATTTCTCTGGCCTAACTGGCCNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNGGTACCATCTCCCCCCgtaatatcttggcaccagatattttgagattc
aGGCCCGAATTAATTCGCTGTCTGCGAGGGCCAGCTGTTGGGGTGAGTACTCCCTCTCAA
AAGCGGGCATGACTTCTGCGCTAAGATTGTCAGTTTCCAAAAACGAGGAGGATTTGATATT
CACCTGGCCCGCGGTGATGCCTTTGAGGGTGGCCGCGTCCATCTGGTCAGAAAAGACAAT
CTTTTTGTTGTCAAGCTTGAGGTGTGGCAGGCTTGAGATCTGGCCATACACTTGAGTGACA
ATGACATCCACTTTGCCTTTCTCTCCACAGGTGTCCACTCCCAGGTCCAACTGCAGGTCGC
CTGCAGGCTTAGACTTGACAATTAATCATCGGCATAGTATATCGGCATAGTATAATACGACT
CACTATAGGAGGGCCACCatggaagacgccaaaaacataaagaaaggcccggcgccattctatccgctggaagatg
gaaccgctggagagcaactgcataaggctatgaagagatacgccctggttcctggaacaattgcttttacagatgcacatatcgaggt

ggacatcacttacgctgagtacttcgaaatgtccgttcggttggcagaagctatgaaacgatatgggctgaatacaaatcacagaatc
gtcgtatgcagtgaaaactctcttcaattctttatgccggtgttgggcgcgttatttatcggagttgcagttgcgcccgcgaacgacatttata
atgaacgtgaattgctcaacagtatgggcatttcgcagcctaccgtggtgttcgtttccaaaaaggggttgcaaaaaattttgaacgtgc
aaaaaaagctcccaatcatccaaaaaattattatcatggattctaaaacggattaccagggatttcagtcgatgtacacgttcgtcacat
ctcatctacctcccggttttaatgaatacgattttgtgccagagtccttcgatagggacaagacaattgcactgatcatgaactcctctgga
tctactggtctgcctaaaggtgtcgctctgcctcatagaactgcctgcgtgagattctcgcatgccagagatcctatttttggcaatcaaatc
attccggatactgcgattttaagtgttgttccattccatcacggttttggaatgtttactacactcggatatttgatatgtggatttcgagtcgtctt
aatgtatagatttgaagaagagctgtttctgaggagccttcaggattacaagattcaaagtgcgctgctggtgccaaccctattctccttct
tcgccaaaagcactctgattgacaaatacgatttatctaatttacacgaaattgcttctggtggcgctcccctctctaaggaagtcgggga
agcggttgccaagaggttccatctgccaggtatcaggcaaggatatgggctcactgagactacatcagctattctgattacacccgag
ggggatgataaaccgggcgcggtcggtaaagttgttccatttttgaagcgaaggttgtggatctggataccgggaaaacgctgggcg
ttaatcaaagaggcgaactgtgtgtgagaggtcctatgattatgtccggttatgtaaacaatccggaagcgaccaacgccttgattgac
aaggatggatggctacattctggagacatagcttactgggacgaagacgaacacttcttcatcgttgaccgcctgaagtctctgattaa
gtacaaaggctatcaggtggctcccgctgaattggaatccatcttgctccaacaccccaacatcttcgacgcaggtgtcgcaggtcttc
ccgacgatgacgccggtgaacttcccgccgccgttgttgttttggagcacggaaagacgatgacggaaaaagagatcgtggattacg
tcgccagtcaagtaacaaccgcgaaaaagttgcgcggaggagttgtgtttgtggacgaagtaccgaaaggtcttaccggaaaactc
gacgcaagaaaaatcagagagatcctcataaaggccaagaagggcggaaagatcgccgtgtaattctagagtcggggcggccg
gccgcttcgagcagacatgataagatacattgatgagtttggacaaaccacaactagaatgcagtgaaaaaaatgctttatttgtgaa
atttgtgatgctattgctttatttgtaaccattataagctgcaataaacaagttaacaacaacaattgcattcattttatgtttcaggttcaggg
ggaggtgtgggaggttttttaaagcaagtaaaacctctacaaatgtggtaaaatcgataaggatccgtcgaccgatgcccttgagagc
cttcaacccagtcagctccttTCCGGAGGAGGCTCTGGCGGTGGAAGCGGCATGATTGAACAAGATG
GATTGCACGCAGGTTCTCCGGCCGCTTGGGTGGAGAGGCTATTCGGCTATGACTGGGCAC
AACAGACAATCGGCTGCTCTGATGCCGCCGTGTTCCGGCTGTCAGCGCAGGGGCGCCCG
GTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCCTGAATGAACTGCAAGACGAGGCAGCG
CGGCTATCGTGGCTGGCCACGACGGGCGTTCCTTGCGCAGCTGTGCTCGACGTTGTCACT
GAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGGGCAGGATCTCCTGTCATC
TCACCTTGCTCCTGCCGAGAAAGTATCCATCATGGCTGATGCAATGCGGCGGCTGCATAC
GCTTGATCCGGCTACCTGCCCATTCGACCACCAAGCGAAACATCGCATCGAGCGAGCACG
TACTCGGATGGAAGCCGGTCTTGTCGATCAGGATGATCTGGACGAAGAACATCAGGGGCT
CGCGCCAGCCGAACTGTTCGCCAGGCTCAAGGCGAGCATGCCCGACGGCGAGGATCTCG
TCGTGACCCATGGCGATGCCTGCTTGCCGAATATCATGGTGGAAAATGGCCGCTTTTCTG
GATTCATCGACTGTGGCCGGCTGGGTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTA
CCCGTGATATTGCTGAAGAACTTGGCGGCGAATGGGCTGACCGCTTCCTCGTGCTTTACG
GTATCGCCGCTCCCGATTCGCAGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTG
ATAATTCTAGAGGTGTCGGTTATCTGCAGATCGGAAGAGCGTCGGCCTCGGCGGCCAAGC
TAGTCGGGCGGCCgcttcgagcagacatgataagatacattgatgagtttggacaaaccacaactagaatgcagtgaaaa
aaatgctttatttgtgaaatttgtgatgctattgctttatttgtaaccattataagctgcaataaacaagttaacaacaacaattgcattcatttt
atgtttcaggttcaggggagatgtgggaggttttttaaagcaagtaaaacctctacaaatgtggtaaaatcgataaggatcttcctaga
gcatggctacgtagataagtagcatggcgggttaatcattaactacaaggaacccctagtgatggagttggccactccctctctgcgcg
ctcgctcgctcactgaggccgggcgaccaaaggtcgcccgacgcccgggctttgcccgggcggcctcagtgagcgagcgagcgc
gcagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgctcttccgcttcctcgctcactgactcgctgc
gctcggtcgttcggctgcggcgagcggtatcagctcactcaaaggcggtaatacggttatccacagaatcaggggataacgcagga
aagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaag

> RDJ_lucif_enh_assay_RA12_pTAN02 (enhancer test with luciferase)
cgcgttgctggcgtttttccataggctccgcccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgac
aggactataaagataccaggcgtttccccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccggatacctgtccgc
ctttctcccttcgggaagcgtggcgctttctcatagctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaagctgggctgtgtg
cacgaaccccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggtaagacacgacttatcgccact
ggcagcagccactggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttgaagtggtggcctaactacggcta
cactagaagaacagtatttggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaaacaaa
ccaccgctggtagcggtggtttttttgtttgcaagcagcagattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctac
ggggtctgacgctcagtggaacgaaaactcacgttaagggattttggtcatgagattatcaaaaaggatcttcacctagatccttttaaat
taaaaatgaagttttaaatcaatctaaagtatatatgagtaaacttggtctgacagttaccaatgcttaatcagtgaggcacctatctcagc
gatctgtctatttcgttcatccatagttgcctgactccccgtcgtgtagataactacgatacgggagggcttaccatctggccccagtgctg
caatgataccgcgagacccacgctcaccggctccagatttatcagcaataaaccagccagccggaagggccgagcgcagaagtg
gtcctgcaactttatccgcctccatccagtctattaattgttgccgggaagctagagtaagtagttcgccagttaatagtttgcgcaacgttg
ttgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttcccaacgatcaaggcgagttacatga
tcccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggccgcagtgttatcactcatggttat
ggcagcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgagaatagtg
tatgcggcgaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgctcatcattgga
aaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaactgatcttca
gcatcttttactttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggaataagggcgacacgga
aatgttgaatactcatactcttcctttttcaatattattgaagcatttatcagggttattgtctcatgagcggatacatatttgaatgtatttagaa
aaataaacaaataggggttccgcgcacatttccccgaaaagtgccacctaaattgtaagcgttaatattttgttaaaattcgcgttaaattt
ttgttaaatcagctcattttttaaccaataggccgaaatcggcaaaatcccttataaatcaaaagaatagaccgagatagggttgagtgtt
gttccagtttggaacaagagtccactattaaagaacgtggactccaacgtcaaagggcgaaaaaccgtctatcagggcgatggccc
actacgtgaaccatcaccctaatcaagttttttggggtcgaggtgccgtaaagcactaaatcggaaccctaaagggagcccccgattt
agagcttgacggggaaagccggcgaacgtggcgagaaaggaagggaagaaagcgaaaggagcgggcgctagggcgctggc
aagtgtagcggtcacgctgcgcgtaaccaccacacccgccgcgcttaatgcgccgctacagggcgcgtcccattcgccattcaggct
gcgcaactgttgggaagggcgatcggtgcgggcctcttcgctattacgccagctgcgcgctcgctcgctcactgaggccgcccgggc
aaagcccgggcgtcgggcgacctttggtcgcccggcctcagtgagcgagcgagcgcgcagagagggagtggccaactccatcac
taggggttccttgtagttaatgattaacccgccatgctacttatctacgtagccatgctctaggaagatcgGGTACCNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNgctagcgtaatggtaagccatgcagatccatggaagacgccaaaaacatCTCG
AGATCTGCACTCATTCATAAAACGCTTGTTATAAAAGCAGTGGCTGCGGCGCCTCGTACTC
CAACCGCATCTGCAGCGAGCAACTGAGAAGCCAAGACTGAGCCGGCGGCCTCTAGAGTC
GACTCTAGAGGATCTCCATGGAAGACGCCAAAAACATAAAGAAAGGCCCGGCGCCATTCT
ATCCTCTAGAGGATGGAACCGCTGGAGAGCAACTGCATAAGGCTATGAAGAGATACGCCC
TGGTTCCTGGAACAATTGCTTTTACAGATGCACATATCGAGGTGAACATCACGTACGCGGA
ATACTTCGAAATGTCCGTTCGGTTGGCAGAAGCTATGAAACGATATGGGCTGAATACAAAT
CACAGAATCGTCGTATGCAGTGAAAACTCTCTTCAATTCTTTATGCCGGTGTTGGGCGCGT
TATTTATCGGAGTTGCAGTTGCGCCCGCGAACGACATTTATAATGAACGTGAATTGCTCAA
CAGTATGAACATTTCGCAGCCTACCGTAGTGTTTGTTTCCAAAAAGGGGTTGCAAAAAATTT
TGAACGTGCAAAAAAAATTACCAATAATCCAGAAAATTATTATCATGGATTCTAAAACGGATT
ACCAGGGATTTCAGTCGATGTACACGTTCGTCACATCTCATCTACCTCCCGGTTTTAATGAA
TACGATTTTGTACCAGAGTCCTTTGATCGTGACAAAACAATTGCACTGATAATGAATTCCTC
TGGATCTACTGGGTTACCTAAGGGTGTGGCCCTTCCGCATAGAACTGCCTGCGTCAGATTC

TCGCATGCCAGAGATCCTATTTTTGGCAATCAAATCATTCCGGATACTGCGATTTTAAGTGT
TGTTCCATTCCATCACGGTTTTGGAATGTTTACTACACTCGGATATTTGATATGTGGATTTC
GAGTCGTCTTAATGTATAGATTTGAAGAAGAGCTGTTTTTACGATCCCTTCAGGATTACAAA
ATTCAAAGTGCGTTGCTAGTACCAACCCTATTTTCATTCTTCGCCAAAAGCACTCTGATTGA
CAAATACGATTTATCTAATTTACACGAAATTGCTTCTGGGGGCGCACCTCTTTCGAAAGAAG
TCGGGGAAGCGGTTGCAAAACGCTTCCATCTTCCAGGGATACGACAAGGATATGGGCTCA
CTGAGACTACATCAGCTATTCTGATTACACCCGAGGGGGATGATAAACCGGGCGCGGTCG
GTAAAGTTGTTCCATTTTTTGAAGCGAAGGTTGTGGATCTGGATACCGGGAAAACGCTGGG
CGTTAATCAGAGAGGCGAATTATGTGTCAGAGGACCTATGATTATGTCCGGTTATGTAAAC
AATCCGGAAGCGACCAACGCCTTGATTGACAAGGATGGATGGCTACATTCTGGAGACATA
GCTTACTGGGACGAAGACGAACACTTCTTCATAGTTGACCGCTTGAAGTCTTTAATTAAATA
CAAAGGATATCAGGTGGCCCCCGCTGAATTGGAATCGATATTGTTACAACACCCCAACATC
TTCGACGCGGGCGTGGCAGGTCTTCCCGACGATGACGCCGGTGAACTTCCCGCCGCCGT
TGTTGTTTTGGAGCACGGAAAGACGATGACGGAAAAAGAGATCGTGGATTACGTCGCCAG
TCAAGTAACAACCGCGAAAAAGTTGCGCGGAGGAGTTGTGTTTGTGGACGAAGTACCGAA
AGGTCTTACCGGAAAACTCGACGCAAGAAAAATCAGAGAGATCCTCATAAAGGCCAAGAAG
GGCGGAAAGTCCAAATTGTAAttcgagcagacatgataagatacattgatgagtttggacaaaccacaactagaatgc
agtgaaaaaaatgctttatttgtgaaatttgtgatgctattgctttatttgtaaccattataagctgcaataaacaagttaacaacaacaattg
cattcattttatgtttcaggttcaggggagatgtgggaggttttttaaagcaagtaaaacctctacaaatgtggtaaaatcgataaggatc
ttcctagagcatggctacgtagataagtagcatggcgggttaatcattaactacaaggaacccctagtgatggagttggccactccctct
ctgcgcgctcgctcgctcactgaggccgggcgaccaaaggtcgcccgacgcccgggctttgcccgggcggcctcagtgagcgagc
gagcgcgcagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgctcttccgcttcctcgctcactgact
cgctgcgctcggtcgttcggctgcggcgagcggtatcagctcactcaaaggcggtaatacggttatccacagaatcaggggataacg
caggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggc

Tiles used for both luciferase experiments and MPRA (Figure 5B)

>RFX_bkg_TTATTTTAAGA
CGTTGCTAGGCAACGTTATTTTAAGACGTTGCTAGGCAACGTTATTTTAAGACGTTGCTAGG
CAACGTTATTTTAAGACGTTGCTAG
> AP1_bkg_CCCGCGCTGCC
TGACTCACCCGCGCTGCCTGACTCACCCGCGCTGCCTGACTCACCCGCGCTGCCTGACTC
ACCCGCGCTGCCTGACTCACCCGCGCT
> MYBL2_bkg_TTATTTTAAGA
AACCGTTAACGGTTATTTTAAGAAACCGTTAACGGTTATTTTAAGAAACCGTTAACGGTTATT
TTAAGAAACCGTTAACGGTTATTT
> NFYshort_bkg_CCCGCGCTGCC
TTTGGCAGACTGCCAAACCCGCGCTGCCTTTGGCAGACTGCCAAACCCGCGCTGCCTTTG
GCAGACTGCCAAACCCGCGCTGCCTTT

SUPPLEMENTAL REFERENCES

Arnold CD, Gerlach D, Stelzer C, Boryń ŁM, Rath M, Stark A. 2013. Genome-wide quantitative

enhancer activity maps identified by STARR-seq. Science 339: 1074–1077.

Blanchette M, Kent WJ, Riemer C, Elnitski L, Smit AFA, Roskin KM, Baertsch R, Rosenbloom K, Clawson H, Green ED, et al. 2004. Aligning multiple genomic sequences with the threaded blockset aligner. Genome Res 14: 708–715.

FANTOM Consortium and the RIKEN PMI and CLST (DGT), Forrest ARR, Kawaji H, Rehli M, Baillie JK, de Hoon MJL, Haberle V, Lassmann T, Kulakovskiy IV, Lizio M, et al. 2014. A promoter-level mammalian expression atlas. Nature 507: 462–470.

Gjoneska E, Pfenning AR, Mathys H, Quon G, Kundaje A, Tsai L-H, Kellis M. 2015. Conserved epigenomic signals in mice and humans reveal immune basis of Alzheimer's disease. Nature 518: 365–369.

Jolma A, Yan J, Whitington T, Toivonen J, Nitta KR, Rastas P, Morgunova E, Enge M, Taipale M, Wei G, et al. 2013. DNA-binding specificities of human transcription factors. Cell 152: 327–339.

Kim T-K, Hemberg M, Gray JM, Costa AM, Bear DM, Wu J, Harmin DA, Laptewicz M, Barbara-Haley K, Kuersten S, et al. 2010. Widespread transcription at neuronal activity-regulated enhancers. Nature 465: 182–187.

Konermann S, Brigham MD, Trevino AE, Hsu PD, Heidenreich M, Cong L, Platt RJ, Scott DA, Church GM, Zhang F. 2013. Optical control of mammalian endogenous transcription and epigenetic states. Nature 500: 472–476.

Malik AN, Vierbuchen T, Hemberg M, Rubin AA, Ling E, Couch CH, Stroud H, Spiegel I, Farh KK-H, Harmin DA, et al. 2014. Genome-wide identification and characterization of functional neuronal activity-dependent enhancers. Nat Neurosci 17: 1330–1339.

Melnikov A, Murugan A, Zhang X, Tesileanu T, Wang L, Rogov P, Feizi S, Gnirke A, Callan CG, Kinney JB, et al. 2012. Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. Nat Biotech 30: 271–277.

Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J, et al. 2015. Integrative analysis of 111 reference human epigenomes. Nature 518: 317–330.

Xie X, Lu J, Kulbokas EJ, Golub TR, Mootha V, Lindblad-Toh K, Lander ES, Kellis M. 2005. Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals. Nature 434: 338–345.

Xie X, Mikkelsen TS, Gnirke A, Lindblad-Toh K, Kellis M, Lander ES. 2007. Systematic discovery of regulatory motifs in conserved regions of the human genome, including thousands of CTCF insulator sites. Proc Natl Acad Sci USA 104: 7145–7150.