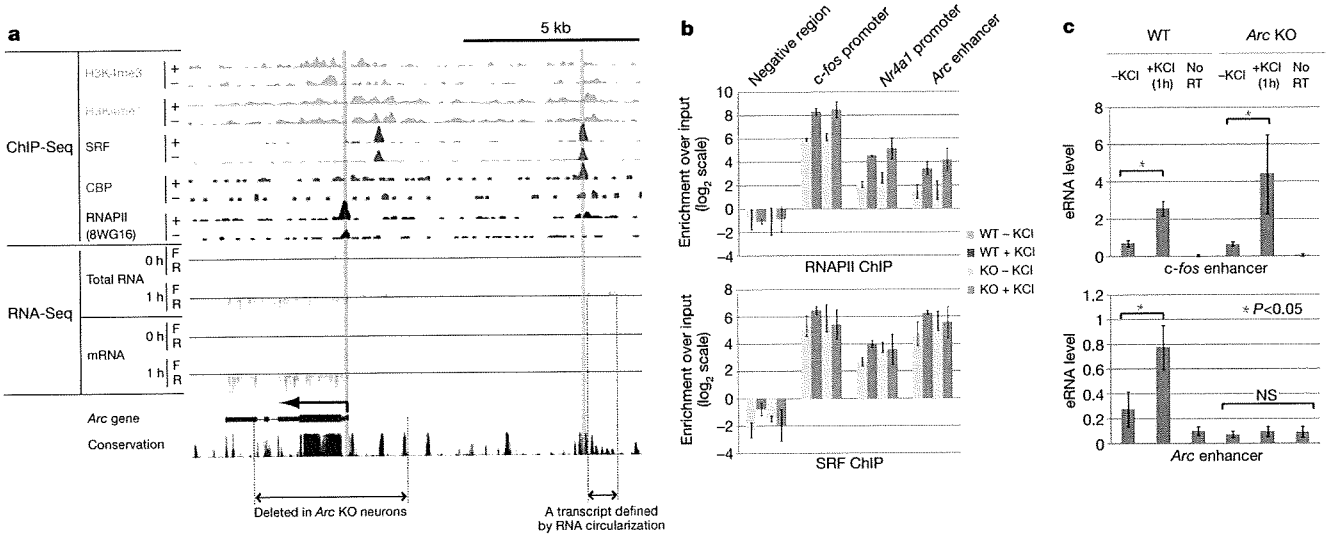


Figure 6 | eRNA synthesis but not RNAPII binding at the *Arc* enhancer requires the presence of the *Arc* promoter. a, The mouse *Arc* genomic locus with ChIP-Seq and RNA-Seq data as in Fig. 1. Also shown are the region deleted in the *Arc* knockout (*Arc* KO) mouse and a non-polyadenylated eRNA transcript defined by the RNA circularization method (Methods). **b**, Binding profiles of RNAPII and SRF at various loci determined by ChIP-qPCR from both wild-type and *Arc* knockout neurons. Error bars indicate s.e.m. ($n = 2$ biological replicates) **c**, RT–qPCR detection of the presence of eRNAs from wild-type and *Arc* knockout neurons. 'No RT' represents the qPCR signal from cDNA samples generated from reactions in which reverse transcriptase was omitted. Error bars are s.e.m. ($n = 3$ biological replicates); *P*-values are from *t*-test. NS, not significant.

by ChIP-Seq to be bound to the *Arc* enhancer (Fig. 6a, b)[13,14]. In *Arc* knockout neurons, both SRF and RNAPII remained bound at the *Arc* enhancer at levels equivalent to those observed in wild-type neurons, indicating that the binding of SRF and RNAPII to the *Arc* enhancer is independent of the *Arc* promoter. However, in the absence of the *Arc* promoter, we were not able to detect eRNA synthesis at the *Arc* enhancer (Fig. 6c). This absence of eRNA was specific to the *Arc* enhancer, as we observed robust induction of eRNA at a c-*fos* enhancer in the *Arc* knockout neurons. These results demonstrate that the recruitment of RNAPII to the *Arc* enhancer is not sufficient to drive eRNA synthesis and suggest that, like mRNA synthesis, eRNA synthesis may require an interaction of the enhancer with a promoter.

Discussion

We provide genome-wide evidence that thousands of neuronal activity-regulated enhancers that are defined by activity-independent H3K4me1 marks and activity-dependent CBP binding also recruit RNAPII and produce eRNAs. The observation of widespread RNAPII binding at enhancers suggests that a general mechanism of activity-dependent enhancer function in neurons may involve recruitment of RNAPII to enhancer loci, followed by subsequent transfer of RNAPII to promoters. Previous studies of a few individual enhancer loci have proposed several models for delivery of RNAPII from an enhancer to a promoter, including tracking of RNAPII along DNA and direct transfer of RNAPII via DNA looping[20]. Our observation that eRNAs are produced only within 2-kb enhancer domains and not along the entire distance between enhancers and promoters indicates that transcription-dependent RNAPII tracking is not likely to be a widespread mechanism of RNAPII delivery.

Our finding that large numbers of neuronal activity-regulated enhancers recruit RNAPII implies that enhancers may be more similar to promoters than previously appreciated. However, our analysis of the *Arc* enhancer in neurons lacking the *Arc* promoter demonstrates that the transcriptional machinery assembled at the *Arc* enhancer is not able to drive transcriptional initiation without the *Arc* promoter. This finding may explain why the level of eRNA synthesis is correlated with the level of transcription at the nearest promoter, and it suggests that eRNA synthesis at many enhancers may require a dynamic interaction between an enhancer and a promoter.

A remaining question is whether eRNAs have a specific biological function. In one model, the RNAPII-dependent transcriptional process at enhancers itself, rather than the eRNA transcripts it produces, may be important for enhancer function. For example, RNAPII has previously been shown to recruit chromatin-modifying enzymes such as histone methyltransferases[26]. In this regard, it is noteworthy that eRNAs are observed only within the H3K4me1-modified enhancer domain, and the level of the H3K4me1 modification and the level of eRNA synthesis are tightly correlated (compare Figs 2b, top, and 4c). Thus, the process of eRNA synthesis could be required to establish and maintain a chromatin landscape at enhancers that is required for enhancer function. However, it is also possible that the eRNA transcripts themselves are functionally important. The ability of enhancers to be transcribed in a regulated manner may provide an evolutionary mechanism by which new, functionally important genes or non-coding RNAs are generated.

METHODS SUMMARY

Directionality index at promoters and enhancers. A directionality index was defined as $|f - r|/(f + r)$, where f is the number of divergent reads on the forward strand and r is the number of divergent reads on the reverse strand within 1.5 kb of the CBP peak or TSS. (See Fig. 5b.)

Calculating the number of extragenic enhancers that produce eRNAs. The level of eRNA for each enhancer locus is calculated by counting all RNA-Seq reads found within a 1.5-kb region on both sides of the CBP peak. As a control, we consider the number of reads found in the adjacent regions (−3.5 kb to −2 kb) and (+2 kb to +3.5 kb) relative to the CBP peak, and in random regions. If one requires >7 reads for detection, 2,267 or 44% of the enhancers have eRNAs, compared to 16% of the flanking regions and 2% of the random regions. (See Fig. 5c.)

Changes in eRNA levels and RNAPII binding at enhancers. For changes in RNAPII binding at enhancers, we counted the number of ChIP-Seq reads within 300 bp of the enhancer centre at each time point. For eRNAs, we used the same procedure, including all reads within 1.5 kb of the enhancer. We defined the normalized induction index as $(s - u)/(s + u)$, where s and u are the number of normalized reads from the stimulated and unstimulated conditions, respectively. (See Fig. 5d.)

Correlations between enhancer features and mRNA expression levels at nearby genes. We paired each enhancer with the nearest TSS, provided that the distance was <1 Mb. The induction index for RefSeq genes was calculated as before for RNAPII, but based on the average read density throughout the coding region for mRNA. Genes were grouped by induction ratio quantiles into 25 bins before plotting. (See Fig. 5d.)

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 21 October 2009; accepted 25 March 2010.
Published online 14 April 2010.

1. Banerji, J., Rusconi, S. & Schaffner, W. Expression of a β-globin gene is enhanced by remote SV40 DNA sequences. *Cell* **27**, 299–308 (1981).
2. Greer, P. L. & Greenberg, M. E. From synapse to nucleus: calcium-dependent gene transcription in the control of synapse development and function. *Neuron* **59**, 846–860 (2008).
3. Heintzman, N. D. *et al.* Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nature Genet.* **39**, 311–318 (2007).
4. Visel, A. *et al.* ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* **457**, 854–858 (2009).
5. Xi, H. *et al.* Identification and characterization of cell type-specific and ubiquitous chromatin regulatory structures in the human genome. *PLoS Genet.* **3**, e136 (2007).
6. Park, P. J. ChIP-seq: advantages and challenges of a maturing technology. *Nature Rev. Genet.* **10**, 669–680 (2009).
7. Barski, A. *et al.* High-resolution profiling of histone methylations in the human genome. *Cell* **129**, 823–837 (2007).
8. Robertson, A. G. *et al.* Genome-wide relationship between histone H3 lysine 4 mono- and tri-methylation and transcription factor binding. *Genome Res.* **18**, 1906–1917 (2008).
9. Chowdhury, S. *et al.* Arc/Arg3.1 interacts with the endocytic machinery to regulate AMPA receptor trafficking. *Neuron* **52**, 445–459 (2006).
10. Plath, N. *et al.* Arc/Arg3.1 is essential for the consolidation of synaptic plasticity and memories. *Neuron* **52**, 437–444 (2006).
11. Shepherd, J. D. *et al.* Arc/Arg3.1 mediates homeostatic synaptic scaling of AMPA receptors. *Neuron* **52**, 475–484 (2006).
12. Waltereit, R. *et al.* Arg3.1/Arc mRNA induction by Ca^{2+} and cAMP requires protein kinase A and mitogen-activated protein kinase/extracellular regulated kinase activation. *J. Neurosci.* **21**, 5484–5493 (2001).
13. Kawashima, T. *et al.* Synaptic activity-responsive element in the Arc/Arg3.1 promoter essential for synapse-to-nucleus signaling in activated neurons. *Proc. Natl Acad. Sci. USA* **106**, 316–321 (2009).
14. Pintchovski, S. A., Peebles, C. L., Kim, H. J., Verdin, E. & Finkbeiner, S. The serum response factor and a putative novel transcription factor regulate expression of the immediate-early gene Arc/Arg3.1 in neurons. *J. Neurosci.* **29**, 1525–1537 (2009).
15. Flavell, S. W. & Greenberg, M. E. Signaling mechanisms linking neuronal activity to gene expression and plasticity of the nervous system. *Annu. Rev. Neurosci.* **31**, 563–590 (2008).
16. Lin, Y. *et al.* Activity-dependent regulation of inhibitory synapse development by Npas4. *Nature* **455**, 1198–1204 (2008).
17. Kee, B. L., Arias, J. & Montminy, M. R. Adaptor-mediated recruitment of RNA polymerase II to a signal-dependent activator. *J. Biol. Chem.* **271**, 2373–2375 (1996).
18. Brodsky, A. S. *et al.* Genomic mapping of RNA polymerase II reveals sites of co-transcriptional regulation in human cells. *Genome Biol.* **6**, R64 (2005).
19. Kim, T. H. *et al.* A high-resolution map of active promoters in the human genome. *Nature* **436**, 876–880 (2005).
20. Koch, F., Jourquin, F., Ferrier, P. & Andrau, J. C. Genome-wide RNA polymerase II: not genes only! *Trends Biochem. Sci.* **33**, 265–273 (2008).
21. Szutorisz, H., Dillon, N. & Tora, L. The role of enhancers as centres for general transcription factor recruitment. *Trends Biochem. Sci.* **30**, 593–599 (2005).
22. Abeel, T., Saeys, Y., Rouze, P. & Van de Peer, Y. ProSOM: core promoter prediction based on unsupervised clustering of DNA physical profiles. *Bioinformatics* **24**, i24–i31 (2008).
23. Ling, J. *et al.* HS2 enhancer function is blocked by a transcriptional terminator inserted between the enhancer and the promoter. *J. Biol. Chem.* **279**, 51704–51713 (2004).
24. Zhao, H. & Dean, A. An insulator blocks spreading of histone acetylation and interferes with RNA polymerase II transfer between an enhancer and gene. *Nucleic Acids Res.* **32**, 4903–4919 (2004).
25. Heintzman, N. D. *et al.* Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459**, 108–112 (2009).
26. Gerber, M. & Shilatifard, A. Transcriptional elongation by RNA polymerase II and histone methylation. *J. Biol. Chem.* **278**, 26303–26306 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank members of the Greenberg laboratory for discussions and for critical reading of the manuscript. We thank S. Vasquez for preparing dissociated mouse cortical neurons. We thank L. Hu for generating antibodies. We thank the Molecular Genetics Core Facility at Children's Hospital Boston, including H. Schneider and S. Burgess, for operation of their SOLiD 3.0 sequencer (I.D.D.R.C). We thank the support and R&D teams at Life Technologies including S. Ranade, R. David, J. Ni, C. Barbacioru, M. Barker, G. Costa and K. McKernan. M.E.G. acknowledges the support of the Nancy Lurie Marks Family Foundation. We thank M. Dehoff for technical support in the *Arc* knockout experiments. This work was supported by the National Institutes of Health grants NS028829 (M.E.G.), R21EY019710 (G.K.), DP2OD006461 (G.K.) and MH-053608 (P.F.W.). This work was also supported by The Lefler postdoctoral fellowship (T-K.K.) and The Jane Coffin Childs Memorial Funds (T-K.K.), The Helen Hay Whitney postdoctoral fellowship (J.M.G.), The Children's Hospital Ophthalmology Foundation (G.K.), The Whitehall Foundation (G.K.), and The Klingenstein Fund (G.K.)

Author Contributions T-K.K., J.M.G. and M.E.G. conceived and designed experiments. T-K.K., J.M.G., M.H., G.K. and M.E.G. wrote the manuscript. T-K.K. optimized the protocol for ChIP-Seq library preparation to be suitable for the SOLiD sequencer and made all ChIP-Seq libraries used in this study. S.K. invented the library construction methodology used for all RNA sequencing reported here. J.M.G., A.M.C. and E.M.-P. made all RNA-Seq libraries. M.H., J.M.G. and D.A.H. performed bioinformatic analyses. K.B.-H. carried out the SOLiD bead preparation and sequencing. T-K.K., J.W., P.F.W. and A.M.C. performed the *Arc* knockout experiment. D.M.B. performed the luciferase experiments. M.L. performed the RNA circularization experiment. H.B. provided the pArc7000 plasmid. D.K. provided the *Arc* knockout mouse. All authors reviewed the manuscript.

Author Information Sequencing data have been submitted to the GEO repository under accession numbers GSE21161 (for all ChIP-Seq and RNA-Seq data) and HM047267 (for circularized *Arc* enhancer RNA). The bigWig files for genome browser visualization are posted online (see Supplementary Table 6). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.E.G. (michael_greenberg@hms.harvard.edu).

METHODS

Mouse cortical cultures. E16.5 C57BL/6 mouse embryo cortices were dissected and then dissociated in 1× Hank's Balanced Salt Solution (HBSS), 20 mg ml^{-1} trypsin (Worthington Biochemicals), and 0.32 mg ml^{-1} L-cysteine (Sigma) for 10 min. Trypsin treatment was terminated with three 2-min washes in 1× HBSS with 10 mg ml^{-1} trypsin inhibitor (Sigma). Trituration of cells was performed with a flame-narrowed Pasteur pipette to dissociate cells fully. Neurons were seeded at an approximate density of 4×10^7 on 15-cm dishes. The dishes were pre-coated overnight with poly-ornithine (30 μg ml^{-1}, Sigma) in water, washed three times with water, and washed once with Neurobasal Medium (Life Technologies) before use. Neurons were maintained in 30 ml Neurobasal Medium containing B27 supplement (2%; Invitrogen), penicillin-streptomycin (50 μg ml^{-1} penicillin, 50 U ml^{-1} streptomycin, Sigma) and glutamine (1 mM, Sigma). Neurons were grown *in vitro* for 7 days. Eight millilitres of the medium was replaced with 10 ml fresh warm medium on the 4th and 6th days *in vitro*.

Membrane depolarization by applying extracellular potassium chloride (KCl). For KCl depolarization of neurons, days *in vitro* (DIV) 6 neurons were quieted overnight in 1 μM tetrodotoxin (TTX, Tocris) and 100 μM D(-)-2-amino-5-phosphonopentanoic acid (D-AP5, Tocris), and they were incubated for 0, 1, 2, or 6 h in 55 mM KCl. ChIP-Seq experiments made use of 0 or 2 h KCl treatment. RNA-Seq experiments were performed with 0, 1, or 6 h of KCl treatment. Luciferase assays were done at 6 h KCl.

Luciferase reporter assays. For testing enhancer functionality, we randomly selected seven extragenic enhancers and tested their ability to function as activity-regulated enhancers in a transfection assay in cultured rat cortical neurons. We transfected DIV 5 neurons using the calcium phosphate method. As a positive control, we used the pGL4.11-Arc7000-luc2P plasmid[27], hereafter referred to as pArc7000, consisting of 7 kb of sequence upstream of the *Arc* coding region. As a negative control, we used pGL4.11-Arc7000-del no 1.luc2P, hereafter referred to as pArc7000-del1, which lacks a 5′ ~250-bp fragment corresponding to the synaptic activity response element (SARE)[27] located 7 kb upstream of the *Arc* gene. To test candidate enhancers, we replaced the SARE (XhoI-SwaI) with ~1-kb fragments centred on the CBP peaks at candidate enhancers. For the negative control loci, we choose three genomic regions where we observed no factor binding (ChIP-Seq) and no RNA expression (RNA-Seq). To excise the *Arc* proximal promoter from each of the candidate enhancer plasmids, we cut with HindIII and re-circularized, which removes two HindIII fragments totalling 1.4 kb.

qPCR validation of eRNA expression and activity induction. For validating eRNA expression, we chose 18 eRNA transcripts that were detected within 1.5 kb of the CBP site. We isolated RNA from three biological replicate KCl experiments at 0, 1 and 6 h of KCl using Trizol (Invitrogen). Each RNA sample was treated with DNaseI (Invitrogen, amplification-grade DNaseI). After the treatment each RNA sample was brought to a volume of 300 μl by the addition of Nuclease Free Water (Ambion), and precipitated with Glycogen (Ambion 5 mg ml^{-1},1:100), 3 M NaOAc (Ambion, 1:10) and 2.5 volumes of 100% ethanol. Reverse transcription of the precipitated RNA was performed using the High Capacity cDNA synthesis kit (Applied Biosystems) with random priming. The cDNA was the source of input for quantitative PCR, using a Step One Plus Real-Time PCR Instrument and SYBR Green reagents (Applied Biosystems). Each primer set used in the analysis was validated using a standard curve obtained from serial dilutions of genomic DNA. Each primer set included in the analysis had melt curves that were consistent with the amplification of a single product in the expected size range. The detectability plot was constructed with concentration values normalized to genomic DNA in the case of eRNA primers, or dilutions of cDNA for gene primers. The inducibility plot was constructed using concentration values that were normalized to corresponding tubulin concentrations.

RNA isolation. RNA was isolated from cultures using 30 ml Trizol on each 15-cm culture dish according to the manufacturer's instructions (Life Technologies). Up to four dishes were solubilized using the same 30 ml Trizol for a yield of 500–1,000 μg total RNA.

To define the 5′ and 3′ termini of uncharacterized RNAs, we circularized RNA, reverse-transcribed circular RNAs using random priming, and PCR-amplified the ligation junctions for conventional sequencing. We generally followed the protocol from ref. 28 with exceptions noted below.

(1) Total RNA from 1-h KCl-treated neurons was first treated with DNaseI as follows: 100 μg of RNA was incubated with 100 units of DNaseI, Amplication Grade (Invitrogen) in 1 ml of the supplied 1× buffer. Then 100 μl of 25 mM EDTA was added and the DNaseI was deactivated at 75 °C for 10 min. RNA was extracted with phenol and chloroform and precipitated with ethanol.

(2) Next we decapped the DNaseI-treated RNA as follows: 10 μg of the RNA was incubated with 25 units of tobacco acid pyrophosphatase (that is, TAP) (Epicentre) in the provided 1× buffer at 37 °C for 1 h. RNA was extracted with phenol and chloroform and precipitated with ethanol.

(3) RNA was circularized using T4 DNA ligase as follows: 8 μg of RNA in 2 ml of 50 mM Tris-HCl (pH 7.6), 10 mM $MgCl_2$, 1 mM ATP, 5% PEG-8000 was incubated with 20 units of T4 RNA ligase (Invitrogen) for 18 h at 37 °C. The ligase was inactivated at 65 °C for 10 min. RNA was extracted with phenol and chloroform and precipitated with ethanol.

(4) Reverse transcription was performed as described with the following modifications: we synthesized cDNA using 1 μg of circularized RNA as a template for cDNA synthesis and random hexameric primers following the Superscipt III reverse transcriptase kit (Invitrogen) in 20 μl reactions.

(5) PCR amplification was performed using a nested PCR approach. Primers were designed to amplify across possible 5′–3′ junctions. First, primary reactions were as follows: 1 μl of cDNA was cycled with 1 μM of each primary primer (Arc_R2new: AGGGTACAAGTAAACAAATACCTGA and Arc_L9new: AGT TCTCTAGCTAAGGCAAGCA) in Power SYBR (Applied Biosystems) mix in a total of 20 μl. Reactions were cycled as follows: 95 °C for 10 min and 20 cycles of 95 °C for 15 s, 60 °C for 30 s and 72 °C for 3 min. Second, the primary PCR reaction was further amplified on a qPCR machine using nested primers as follows: 10 μl of the primary PCR was cycled with 1 μM of each nested primer (Arc_R1new: TTAAGAGTCACAAAGCCACCAAT, Arc_L10new: GTCTCTAC CATTGATGGATCTC) in a 2 ml Power SYBR reaction divided into multiple wells. Nested PCR was performed as follows: 95 °C for 10 min and 40 cycles of 95 °C for 15 s, 60 °C for 1 min. Product was purified using QIAQuick PCR purification columns and sequenced conventionally. The product was mapped to the eRNA region using the UCSC Genome Browser (BLAT).

The sequence of the transcript obtained from this method is: ArcE1 + strand transcript, GGAGAGGTGGGGACCAGAGTCCCTGGCTGGAGACTGGTGAC ATTGTCCCTGCCATTGGTGGCTTTGTGACTCTTAAACCAGACCTGCACA AAGATCTTGTATCAGGTATTTGTTTACTTGTACCCTAGAGCTCTGGTTC CAGGAGAAAGCAGATGGCCCCCCGGGTGGGGGGGCCCTGGGCAGTAG TAGCTCCTCAGTCCTGTAAATAAATCCCTAGGAACAGCGTTCAGGCTG AAGGGTTCGAGCTCTGGGCTGGGCGTACACCAGCGCCAAGATGCAGA CAGGTAAGAATGCTTAGAATTCCTGTGCCTGACATTTCTCATTCTGTCA CAAAGGGGAGTGGGGTACCAATAGGGATGGAGCACAGTGCCCTGAAA GAGTTCAGATTCACACAGAGAACCAGGAAGGGCTTCCTAGAGGTGGAG CCTGTGGGTAGAAGGCAAAGAGCACAGGTTGAGCAGGGGAGGCCAGT GCTGTGTCCCCTCCACTCTTTTGGCTCCCTAATGGCCTTCAAGCGTGG TTACCCTCCTCTGGCTGGTGACTCCTCCGTTTTCTCTCTGGGCGGGGA GGGTGTTGGATCTGGACCTCTTTCTTTCTCCGATGTCTCCTCCTACCAG AGGCAGCTCATCTGAGTTCTCAAGCCTTTGCCCTGGGCTTTGAAACTG ACAACCAGCTCCCAGTGAGTGACCTGTGGAGCTGCAGACATGTGGAGA TCCAAGTTGTCTCCCCACTTCCTGTAGTTCTCTAGCTAAGGCAAGCAG GTCTCTACCATTGATGGATCTCACAGGTACCAGGCACGACTCTCGGTC CCTCGACCCACTGGAAAAGGTTGTGCATGGGTTCAGGGT.

SOLiD sequencing. SOLiD sequencing of ChIP-Seq and RNA-Seq libraries were performed on a SOLiD instrument (1, 2, or 3.0 version) with 35-bp reads according to manufacturer's instructions (Life Technologies). All experiments were performed on full sequencing slides with barcodes used to distinguish up to 16 sequencing libraries on a slide. Libraries were quantified by SYBR green quantitative-PCR (qPCR) to determine appropriate mixing ratios, which also depended on the desired sequencing depth for each of the libraries in the mixture.

Chromatin immunoprecipitation sequencing (ChIP-Seq). Forty million mouse cortical neurons cultured to *in vitro* day 7 were used for each ChIP-Seq library construction. ChIP was performed as described[29] using antibodies listed above. The immunoprecipitated DNA fragments were repaired by the End-It DNA End Repair Kit (Epicentre Biotechnology) according to the manufacturer's instructions. The end-repaired ChIP DNA fragments were purified by MinElute Reaction Cleanup Kit (Qiagen) and eluted in 20 μl in EB buffer. The resulting DNA fragments were ligated with P1 and P2 adaptors for SOLiD genome analyser (adaptor sequences can be made available upon request) for 20 min at room temperature using the Quick Ligase Kit (NEB), followed by purification using the MinElute Reaction Cleanup Kit (Qiagen). The purified, adaptor-ligated ChIP DNA fragments were subject to 6% native-PAGE for an in-gel PCR reaction. A gel slice containing 175–200 bp adaptor-ligated ChIP DNA fragments (corresponding to 125–150 bp genomic fragment sizes) was cut and shredded. PCR Platinum Supermix (100–200 μl, Invitrogen), 50 pmol of PCR primers (available upon request), 0.5 μl Taq DNA polymerase (NEB), and 0.15 μl p.f.u. Turbo DNA polymerase (Stratagene) were added into the shredded gel slice. The adaptor-ligated ChIP DNA fragments were amplified by 15 cycles of in-Gel PCR. After the PCR reaction, gel pieces were filtered out by 0.45 μm filter spin column, and the amplified ChIP-Seq library was purified by the MinElute PCR purification kit (Qiagen). The library was purified by one more round of 6% PAGE. A gel slice containing 200–250 bp PCR products (110–150 bp fragment size) was cut and shredded, and the amplified library was extract out of the gel by passive elution in elution buffer (1.5 M ammonium acetate in 1× TE). Gel pieces were

filtered out by filter spin column, and the resulting ChIP-Seq library was purified using the Qiaquick PCR purification kit (Qiagen).

Whole-transcriptome sequencing (WT-Seq: sequencing of total RNA). WT-Seq was performed according to a protocol/kit now available from Life Technologies, with minor modifications that are included below. Briefly, 5–10 μg of RNA isolated from mouse cortical cultures was depleted of ribosomal RNAs using two rounds of Human/Mouse RiboMinus treatment (Life Technologies) with overnight ethanol precipitations for sample re-concentration. The removal of ribosomal RNAs was confirmed on a Bioanalyser Nano Chip (Agilent). A total of 500–1,000 ng of riboRNA-depleted total RNA was fragmented with 10–18 min at 37 °C RNaseIII treatment, and 10 min of RNaseIII inactivation at 65 °C. Fragmentation was followed by size selection of ~50 to ~150 bp fragments using the flashPAGE denaturing PAGE-fractionator (Life Technologies) and ethanol precipitation overnight. The resulting RNA was directionally ligated, reverse-transcribed and RNaseH treated.

After trial PCR to assess library quality and quantity, 30 μl cDNA was run on a native 6% PAGE gel. The 90–120-bp size window (corresponding to 50–80-bp RNA insert size) was cut from the gel, shredded and inserted directly into a 400 μl PCR reaction using standard WT-Seq kit components and submitted to 11–15 cycles of PCR. The PCR product was phenol-chloroform extracted, ethanol precipitated and re-suspended in 20 μl WT-Seq gel loading buffer. The resulting sample was run on a 6% native PAGE gel, and the 150–175-bp size range (corresponding to 60–85 bp) was cut from the gel, shredded, and extracted overnight in WT-Seq PAGE elution buffer. The resulting library was filtered through 0.45 μm spin filters (Life Technologies) to remove gel pieces and ethanol precipitated.

We note that WT-Seq can detect neither the 5′-most fragment from transcripts with 5′-modified ends (such as mRNA 5′ 7-methyl-guanosine caps) nor the 3′-most fragment from transcripts with 3′-modified ends. However, for transcripts long enough to produce multiple ≥50-bp fragments, WT-Seq should detect the remaining fragments.

Mouse cortical neuron WT-Seq data presented in this manuscript are from one specific biological replicate, but each result was confirmed in at least one additional replicate.

mRNA sequencing (mRNA-Seq). mRNA-Seq was performed exactly as WT-Seq, except that the ribosomal RNA-removal steps were replaced by two rounds of polyA purification using the FastTrack MAG mRNA isolation kit (Life Technologies). The removal of ribosomal RNAs was confirmed on a Bioanalyser Nano Chip (Agilent).

Annotation version and mRNA TSS collection. For filtering CBP peaks to remove TSSs, we used all TSSs from UCSC known genes, Ensembl genes and RefSeq genes. The NCBI reference sequence (RefSeq) collection of mouse gene annotation, version 37_1, was used for analysing TSSs for comparison with enhancers (for example, Fig. 2a). A subset of RefSeq genes has multiple annotated TSSs per gene, and we used for these analyses only the 5′-most TSSs from these genes. Thus, 25,562 TSSs were used instead of the full set of 27,854 RefSeq TSSs.

Read alignment (mapping sequencing reads to the genome and splice junctions). ChIP-Seq, WT-Seq and mRNA-Seq sequencing reads were aligned using the large genomes matching pipeline from Life Technologies with parameters –e 3 –t 35 –z 10. These parameters dictate that 0–3 colour-space mismatches are allowed, a 35-bp read is aligned, and after 10 hits on a given chromosome, the aligner no longer looks for further matches. ChIP-Seq reads were aligned to the mouse NCBI genome version 37. WT-Seq and mRNA-Seq reads were also aligned to mouse NCBI 37, but in their cases, the genome was expanded with addition of a pseudo-chromosome consisting of exon–exon splice junctions, although reads aligning to splice junctions were used solely to assess the strand specificity of WT-Seq (next section). Only reads aligning to a single genomic position with a tolerance of 0–3 colour-space mismatches were used for findings reported in this manuscript.

ChIP-Seq read-length extension. After ChIP-Seq reads were aligned, they were extended to 120 bp to match the length of the DNA fragments that were sequenced. We chose to extend to 120 bp based on experimental considerations (see ChIP-Seq procedure, above), but we also confirmed that this was a reasonable extension distance bioinformatically using the procedure introduced by ref. 30.

Peak finding from ChIP-Seq. Chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-Seq) experiments produces a large number of short (~120 bp) DNA fragments which are enriched in the regions where the transcription factor of interest was bound to the DNA. The resulting profile of the mapped short sequence reads to the DNA shows which regions are enriched for the transcription factor in question. As an example of such a profile, consider Fig. 1. The next challenge is to determine which of the enriched regions ('peaks') are statistically significant at a given threshold and what regions correspond to genomic background. The background is given by sequencing the input genomic DNA fragments. Thus, a peak is defined as a region that contains significantly more reads from the ChIP experiment than from the input control. We first determined the false discovery rate by using a sliding window with a width of 240 bp for every 10 bp in the mouse genome. Owing to repetitive sequences, it is impossible to assign 35-bp reads uniquely to some regions[31,32]. Consequently, such regions are excluded from our analysis. For each window, we calculated the statistic $D = R - N$ where R is the number of reads from ChIP, and N is the number of reads from an input sample. By considering the marginal distributions of R and N, we note that they both can be well approximated by a Poisson distribution with parameters λ_R and λ_N, respectively. It follows that D is a Skellam distribution[33]:

$$\Pr(D=d) \approx \mathrm{Ske}(d;\lambda_R,\lambda_N) = e^{-(\lambda_R+\lambda_N)}\left(\frac{\lambda_R}{\lambda_N}\right)^{d/2} I_{|d|}\left(2\sqrt{\lambda_R\lambda_N}\right) \quad (1)$$

where $I_{|d|}(z)$ is the modified Bessel function of the first kind of order $|d|$. When comparing the number of reads in a given window from two different samples, care must be taken to correct for differences in the total amount of sequenced reads. The construction of the null distribution from equation (1) takes unequal numbers of reads in the two samples into account by shifting the mode of the distribution so that mode will be positive if there are more ChIP reads than input control reads and negative if there are more reads for the input control sample. To determine the number of reads required for a 240-bp window to be significant, we use the local false discovery rate (locFDR) framework[34]. Using this methodology, we assume that the density of D, $f(D)$, can be written as the mixture $f(D) = p_0 f_0(D) + p_1 f_1(D)$, where f_0 is the null density, f_1 is the density of windows corresponding to true peaks and $p_0 + p_1 = 1$ with $p_0 \geq 0.9$. The locFDR, FDR(d), is related to the more familiar FDR[35] through

$$\mathrm{FDR}(d) = E[\mathrm{locFDR}(d) | D \geq d] \quad (2)$$

where E is the expectation with respect to the mixture density f. Taking CBP as an example, with a fixed threshold $f_p = 0.01$ and empirically estimated $\lambda_R = 0.346$ and $\lambda_N = 0.419$, we find that the critical difference is $d_0 = 5$ fragments. Inserting the empirical distribution and the Skellam null distribution from equation (1), we find that the FDR is $\sim 3 \times 10^{-6}$. For the H3K4me1 peaks, the procedure described above was used, except that the window size was changed to 1,000 bp at increments of 100 bp. The motivation for using a larger window size is that histone modifications are typically much broader than transcription factor binding sites and using a larger window allows us to consider a signal from a larger genomic region and the process becomes less sensitive to noise. The parameters for the Skellam distribution were $\lambda_R = 11.88$ and $\lambda_N = 1.33$, and from equation (2) we estimate the FDR to be $\sim 2 \times 10^{-5}$.

Reproducibility. To verify that our findings are robust and reproducible, we sequenced at least one biological replicate of each ChIP-Seq and RNA-Seq experiment. The correlations of enrichment (ChIP-Seq) or read numbers (RNA-Seq) are high, except in cases (such as enhancers with few RNA-Seq reads) where sensitivity is poor. We present scatter plots of replicates in Supplementary Fig. 10. In the specific case of CBP, we had one ChIP-Seq experiment using the Millipore antibody and several others using an Abcam antibody. As is often the case in ChIP experiments, the quality of the two experiments varied, with the Millipore antibody performing much better in terms of the proportion of reads under peaks. The quality of the CBP peaks are important for subsequent analysis, both in terms of their number (higher numbers result in better statistical power in downstream analyses) and their quality (false positive CBP sites will reduce the apparent fractional number of enhancers with RNAPII binding and eRNA expression). Thus, we took extra care to select high-confidence CBP sites. First, we used our highest quality ChIP-Seq (using Millipore CBP antibody) to call 41,148 peaks. Of these, 4,567 were also called as peaks in several Abcam ChIP-Seq experiments. In addition, we found that the enrichment of reads from the Abcam ChIP-Seq at the remaining 36,601 loci was substantially above the levels seen at randomly chosen regions. Thus, to gain additional sensitivity in identifying replicated peaks, we set a 0.1 false discovery rate (FDR) based on the 90th percentile of CBP Abcam enrichment at random loci. 23,346 CBP Millipore peaks not called as peaks in the Abcam IP were nonetheless found to be above the 0.1 FDR (Supplementary Fig. 11a). Thus, our final set of validated CBP peaks is 4,567 + 23,457 = 28,024 (Supplementary Fig. 11b).

Alternative peak-finding method. Because the location of the CBP peaks is key to all further analyses for this paper, we validated the results using another peak-finding algorithm called Sissrs[36]. We assumed a fragment length of 120 bp and we used an FDR of 0.001, but other than that, the default settings were not changed. Sissrs reported 32,656 CBP peaks (CBP Millipore, 2 h KCl), and we found that 31,842 (97.5%) of those were located within 1 kb of the peaks detected using our method. Using a more stringent threshold of 100 bp, 83% of the Sissrs peaks were found near one of our peaks. For the unstimulated condition, Sissrs reported

8,980 peaks, significantly more than we found. Closer inspection reveals that all of our peaks were within 1 kb of a Sissrs peak and 78% were within 100 bp. All peaks discovered by both our method and Sissrs for the unstimulated condition show a high degree of induction (Supplementary Fig. 9), indicating that the degree of CBP binding before KCl stimulation is very low.

Normalization of ChIP-Seq and RNA-Seq read numbers. To compensate for differences in total sequencing read depth among samples, all ChIP-Seq read counts were first normalized to their equivalent numbers assuming 10 million total reads per sample. Next, the normalized number of reads in the IP was subtracted from the normalized number of reads in the input within a 240-bp scanning window, and the subtracted value was used for further analysis and plotting. We refer to these two procedures respectively as normalization and input-subtraction. A similar procedure was performed for RNA-Seq data, where read counts were normalized to their equivalent numbers assuming 50 million reads per sample, and the normalized values were used for further analysis and plotting. Because ChIP-Seq reads correspond to the ends of larger DNA fragments produced by sonication, we extended each ChIP-Seq read to 120 bp.

Selecting random control regions. To generate a set of random peaks matching the CBP peaks we did the following for each intragenic peak: (1) calculate the distance d to the TSS of the overlapping gene; (2) select a gene g at random; (3) place the new random region d base pairs downstream of the TSS of gene g; (4) if d is larger than the length of g then the random location is invalid, and we draw a new g.

For extragenic peaks the procedure is similar: (1) calculate the distance d to the nearest TSS (at position t); (2) select a gene g at random; (3) place the new random region at $t+d$; (4) ensure that the location is extragenic; if it is not, then we draw a new g.

When selecting random control regions for comparing the eRNA levels found at the putative enhancers, we wanted to control for the genomic location relative to TSSs as well as for the difficulty of mapping reads to repetitive regions. We defined m as the fraction of 35-mers in a given region that are unique to the mouse genome. For both extra- and intragenic, we also make sure that m is greater than 0.8 and we exclude the sites that do not fulfil this criterion.

Defining enhancers based on ChIP-Seq data. To locate enhancers, we started from the set of 41,148 CBP (Millipore) peaks located by our peak-finding algorithm. Supplementary Table 1 indicates the number of CBP peaks at each stage of this filtering. (We note that the analysis in the main text and main figures is primarily focused on the extragenic peaks.) To be considered as enhancers, individual CBP peaks had to meet the following criteria: (1) the CBP peak had to be replicated with the Abcam CBP antibody (see reproducibility section). (2) The CBP peak had to be at least 1 kb away from all annotated TSSs. (3) CBP peaks with abnormally high levels of both H3K4me1 and H3K4me3 enrichment in a 2-kb bin centred on the peak were also removed. There were 159 sites that were disallowed based on this criterion. When examined in the UCSC Genome browser, these sites showed unusually long (>500 bp) regions of very strong enrichment for multiple transcription factors, suggesting that they were not true binding sites. (4) On the basis of evidence from ESTs and other annotations, it is reasonable to suspect that some of the loci in our enhancer sites could correspond to true promoters. We wanted to take a conservative approach, and hence we removed all CBP peaks that have a 5′-sequenced EST from the UCSC Genome Browser spliced EST track that has a 5′ end within 2 kb of the CBP peak and that spans an annotated TSS. (5) We removed peaks that showed evidence of initiating long transcripts. We compared the regions −4 to −2 kb relative to the CBP peak with the region +2 to +4 kb on the forward strand. If the density in the downstream region was significantly higher than in the upstream region, we took this as evidence that a longer, possibly coding transcript was initiated at the loci. (6) An H3K4me1 peak had to be present within 2 kb in both replicates of stimulated H3K4me1. (7) The enrichment of H3K4me3 within a 2 kb window centred on the peak had to be less than 2 in both unstimulated and stimulated cells (Supplementary Fig. 8). (8) A very small number of CBP peaks (for example, those found near ribosomal RNA genes) had >10,000 RNA-Seq reads mapped within 1.5 kb. We removed these loci to simplify our analysis of RNA-Seq reads at enhancers.

Selection and clustering of enhancers. For each of the transcription factors CBP, CREB, NPAS4, SRF and RNAPII, the 1,000 enhancers with the highest levels of input-normalized ChIP-Seq reads within 200 bp of the enhancer centre were selected as well as the 1,000 enhancers with the highest levels of H3K4me1 within 1 kb. The enhancers were pooled and the ones without any divergent reads within 1.5 kb were removed. Subsequently, 315 enhancers were selected at random and ordered by row on the basis of k-means clustering performed in R, based on the transcription factors mentioned above, the amount of H3K4me1 and the amount of divergent RNA-Seq signal (Fig. 5a).

Directionality index. A directionality index used in Fig. 5b was defined as $|f-r|/(f+r)$, where f is the number of divergent reads on the forward strand and r is the number of divergent reads on the reverse strand within 1.5 kb of the CBP peak or TSS. The estimation of the directionality index is complicated by the fact that the number of reads found at enhancers is much smaller than the number of reads found at promoters. To make sure that the observed difference is not an artefact due to the lower levels of eRNAs, we down-sampled the number of reads at promoters to match the eRNA read numbers.

Detection of enhancers with eRNAs. As a negative control for the eRNAs, we compared with the adjacent regions −3.5 to −2 kb and +2 to +3.5 kb. A less stringent control is provided by the random control loci where we used the −1.5 to +1.5 kb regions. If one requires >7 reads for detection, 2,267 or 44% of the enhancers have eRNAs, compared to 16% of the flanking regions and 2% of the random regions (Fig. 5c).

For the intragenic enhancers, in examining eRNA transcription we were limited to considering the antisense strand. Using a similar strategy to that shown in Fig. 5c, we found that 22% of intragenic enhancers had more than four antisense reads within 1.5 kb of the CBP peak, whereas 14% of the enhancers had more than four antisense reads in the flanking regions.

Changes in eRNA levels and RNAPII binding at enhancers. For changes in RNAPII binding at enhancers, we counted the number of ChIP-Seq reads within 300 bp of the enhancer centre at each time point. For eRNAs, we used the same procedure, including all reads within 1.5 kb of the enhancer. We defined the normalized induction index as $(s-u)/(s+u)$, where s and u are the number of normalized reads from the stimulated and unstimulated conditions, respectively (Fig. 5d).

Correlations between enhancer features and mRNA expression levels at nearby genes. We paired each enhancer with the nearest TSS, provided that the distance was <1 Mb. The induction index for RefSeq genes was calculated as before for RNAPII, but based on the average read density throughout the coding region for mRNA. Genes were grouped by induction ratio quantiles into 25 bins before plotting (Fig. 5d).

Searching for known consensus motifs. For each enhancer and TSS where an NPAS4 peak was found we searched for perfect matches to the NPAS4 consensus motifs 'CACGC' and 'CACGTA'[37] on both strands in a 300-bp window centred around the CBP peak or the TSS. Surprisingly, at least one of the motifs was found in 66% of the promoter peaks but only in 27% of the peaks found at enhancers. The fraction of enhancers with the NPAS4 motifs is only marginally larger than the 22% found at TSSs lacking NPAS4 binding or the 14% found at random loci. Given the fact that there is no significant difference between the peak sizes found at enhancers and promoters (Supplementary Table 1), this result suggests that the mechanism of binding of NPAS4 to promoters and enhancers may differ, as has been observed for other factors.

Induced genes. We provide Supplementary Table 8, showing a list of KCl-regulated genes used for analysis in Fig. 5. The list was generated using our most deeply sequenced RNA-Seq experiment and filtered for fold-change and a z-score ensuring adequate read number. The list of genes is consistent with previous experiments published using array technologies[27,38], and is well correlated with replicate RNA-Seq experiments.

Gene ontology analysis using DAVID. To examine whether particular gene classes were enriched either for CREB binding at their promoters or for regulation by KCl, we took KCl-regulated (Supplementary Table 8) or CREB-bound genes and asked whether particular gene classes were enriched. The results obtained using the software DAVID[39,40] are in Supplementary Table 7 (CREB) and Supplementary Table 9 (KCl-regulated genes).

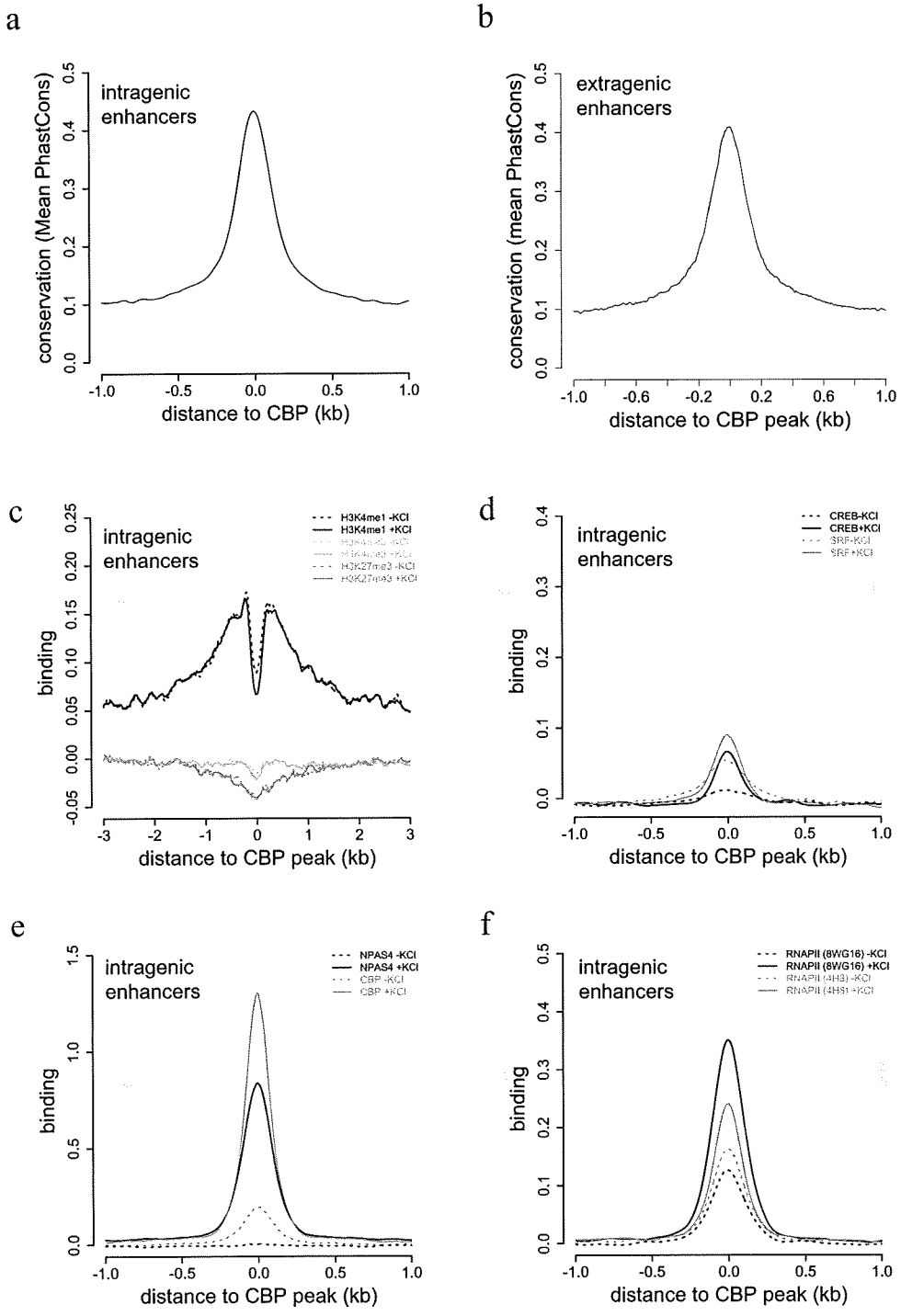
Additional bioinformatics. Analysis of aligned reads was performed with a combination of custom Perl, Java, R, and MATLAB scripts. Additional details are available upon request.

Note on plots. In the case of plots with RefSeq promoters aligned by their mRNA TSSs: in each case, promoters are aligned so that the positions +1 to +1,000 bp along the x axis correspond to the first 1 kb of each reference sequence annotated pre-mRNA, and the positions −1,000 to −1 bp consist of the first 1 kb upstream of each mRNA TSS. For plots with enhancers aligned at the centre of their CBP binding site, the same logic applies with the CBP peak centre substituting for the mRNA TSS. Owing to different sequencing depths, different scales are required in Figs 1 and 6a for displaying different ChIP results. For Fig. 1, the scales are H3K4me3 (0–1), H3K4me1 (0–1), H3K27me3 (0–1), SRF (0–3.5), CBP (0–3), CREB (0–2), NPAS4 (0–3), RNAPII (0–3), RNA-Seq (0–10), where the numbers in parentheses are normalized read counts. The corresponding numbers for Fig. 6a are H3K4me3 (0–0.5), H3K4me1 (0–2), SRF (0–7), CBP (0–1), RNAPII (0–1.8), RNA-Seq (0–10). Note that no input subtraction was performed for these plots. The conservation track shows 30-way Multiz alignment and conservation scores (PhastCons).

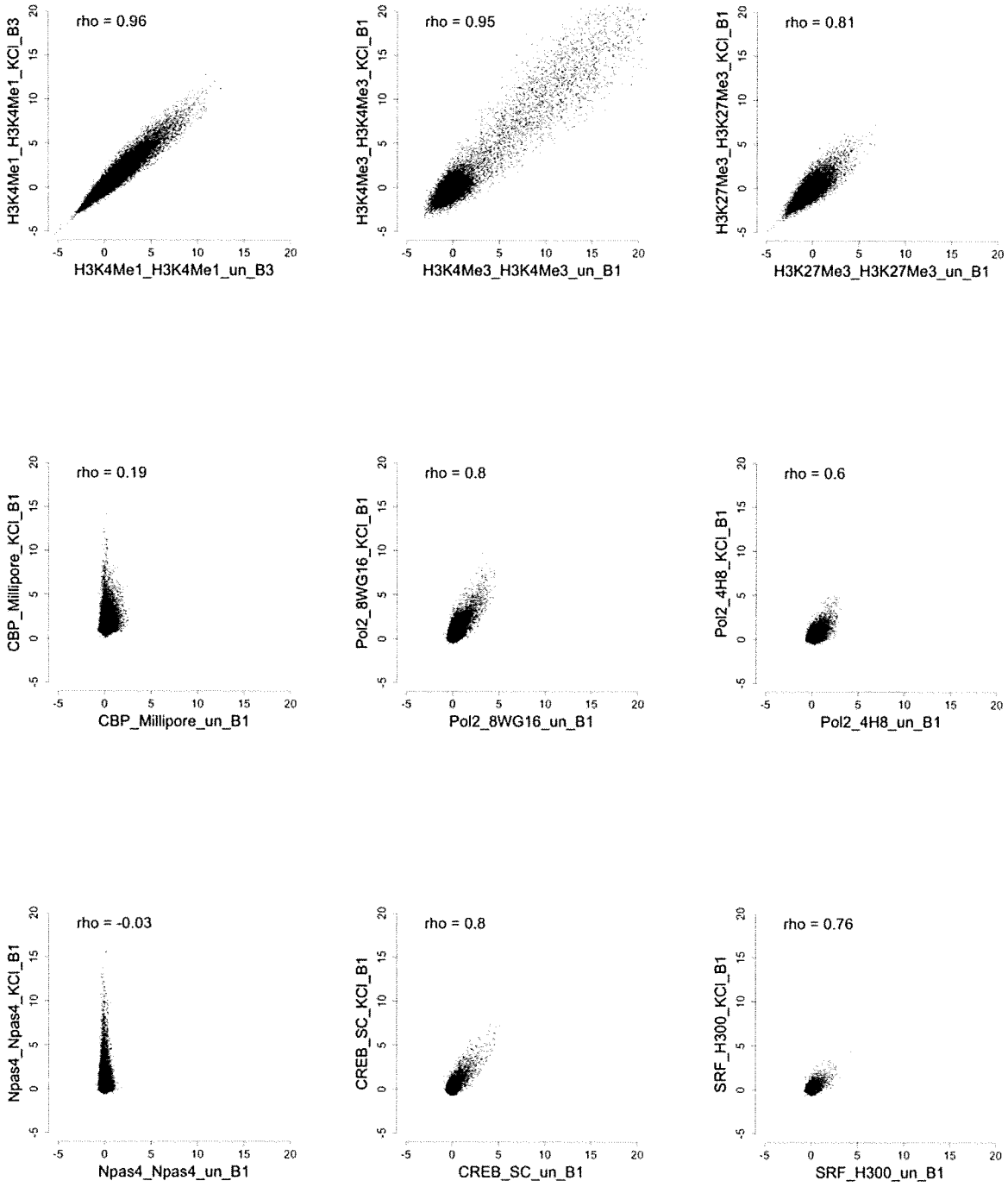
27. Kawashima, T. *et al.* Synaptic activity-responsive element in the Arc/Arg3.1 promoter essential for synapse-to-nucleus signaling in activated neurons. *Proc. Natl Acad. Sci. USA* 106, 316–321 (2009).

28. Couttlet, P. *et al.* Messenger RNA deadenylation precedes decapping in mammalian cells. *Proc. Natl Acad. Sci. USA* **94,** 5628–5633 (1997).
29. Flavell, S. W. *et al.* Genome-wide analysis of MEF2 transcriptional program reveals synaptic target genes and neuronal activity-dependent polyadenylation site selection. *Neuron* **60,** 1022–1038 (2008).
30. Kharchenko, P. V., Tolstorukov, M. Y. & Park, P. J. Design and analysis of ChIP-seq experiments for DNA-binding proteins. *Nature Biotechnol.* **26,** 1351–1359 (2008).
31. Rozowsky, J. *et al.* PeakSeq enables systematic scoring of ChIP-seq experiments relative to controls. *Nature Biotechnol.* **27,** 66–75 (2009).
32. Robertson, A. G. *et al.* Genome-wide relationship between histone H3 lysine 4 mono- and tri-methylation and transcription factor binding. *Genome Res.* **18,** 1906–1917 (2008).
33. Skellam, J. G. The frequency distribution of the difference between two Poisson variates belonging to different populations. *J. R. Stat. Soc. A* **109,** 296 (1946).
34. Efron, B. Microarrays, empirical Bayes and the two-groups model. *Stat. Sci.* **23,** 1–22 (2008).
35. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B. Methodol.* **57,** 289–300 (1995).
36. Jothi, R., Cuddapah, S., Barski, A., Cui, K. & Zhao, K. Genome-wide identification of *in vivo* protein–DNA binding sites from ChIP-Seq data. *Nucleic Acids Res.* **36,** 5221–5231 (2008).
37. Ooe, N. *et al.* Identification of a novel basic helix-loop-helix-PAS factor, NXF, reveals a Sim2 competitive, positive regulatory role in dendritic-cytoskeleton modulator drebrin gene expression. *Mol. Cell. Biol.* **24,** 608–616 (2004).
38. Lin, Y. *et al.* Activity-dependent regulation of inhibitory synapse development by Npas4. *Nature* **455,** 1198–1204 (2008).
39. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. *Nature Protocols* **4,** 44–57 (2009).
40. Dennis, G. Jr *et al.* DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol.* **4,** 3 (2003).

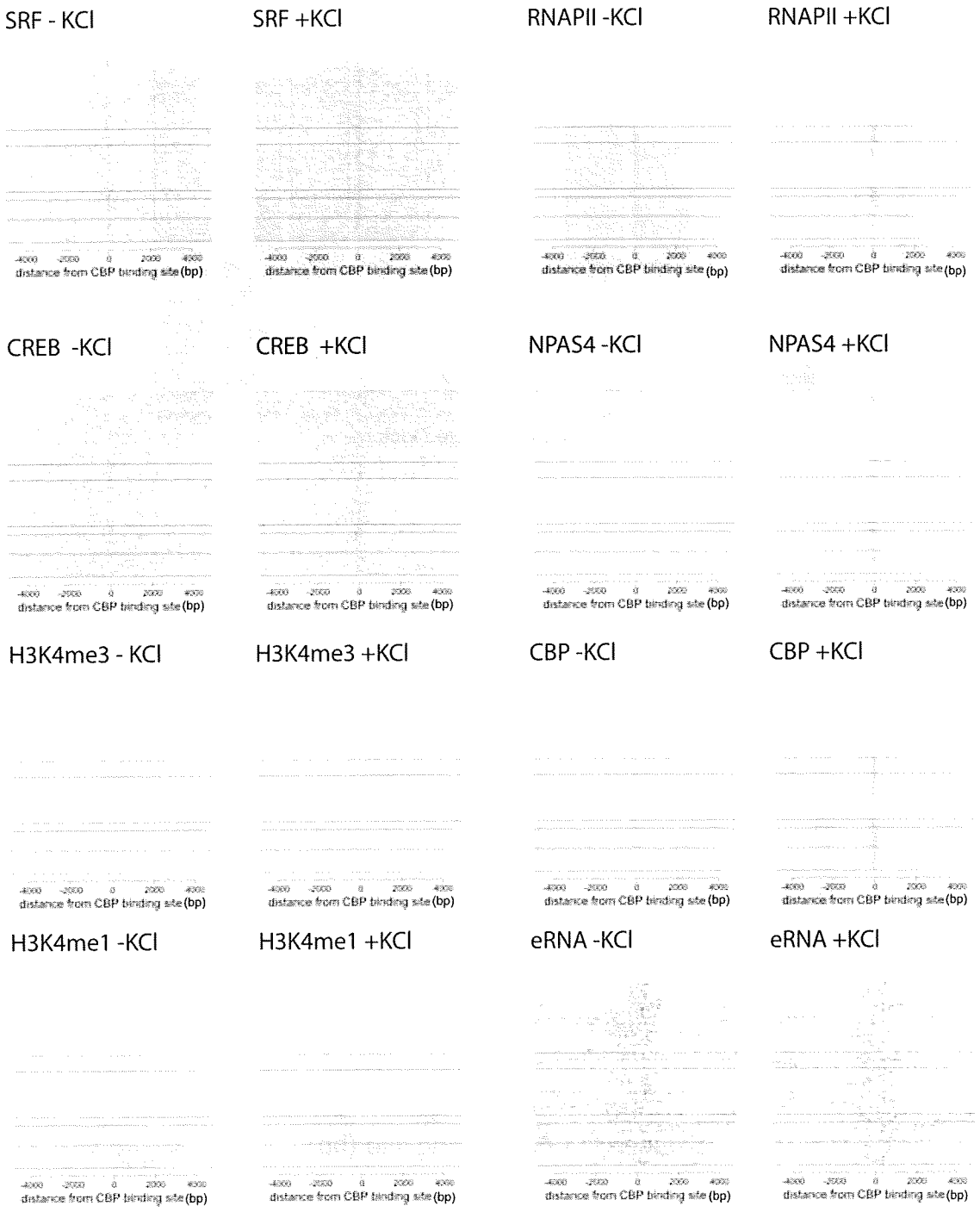
SUPPLEMENTARY INFORMATION



Supplementary Figure 1. ChIP-Seq profiles for histone modifications, TFs, and RNAPII at intragenic enhancers. a-b, Conservation profiles at 6,718 intragenic (**a**) and 5,117 extragenic (**b**) enhancers. **c-f,** Profiles of histone modifications and TFs at enhancers located within intragenic regions are displayed as in Figs. 2b and 4b.

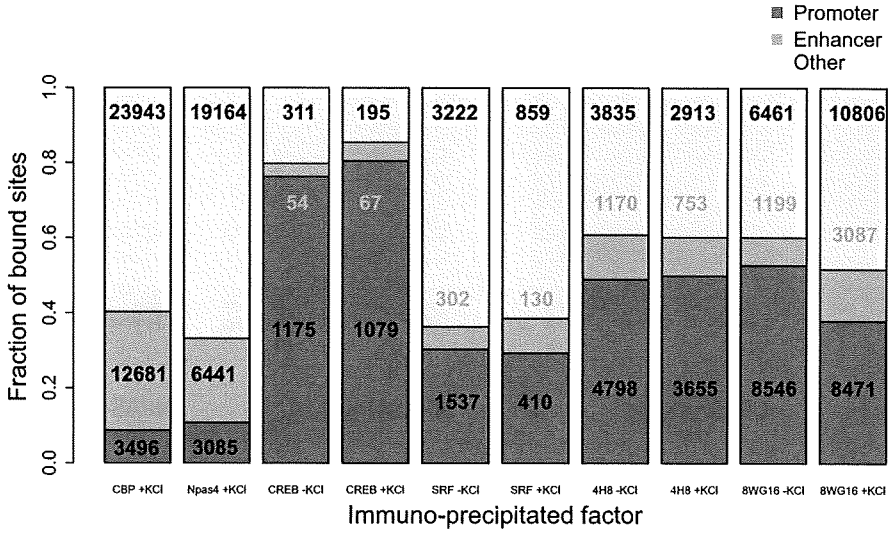


Supplementary Figure 2. Changes in TF binding at individual CBP peaks upon membrane depolarization. In each plot, each dot represents one or more of 41,148 CBP peaks called using the Millipore CBP antibody after 2 hours KCl treatment. The position of each dot on the *x*-axis denotes the number of ChIP-Seq reads in the unstimulated condition, while the position of each dot on the *y*-axis denotes the number of ChIP-Seq reads in the stimulated condition. For histone methylation marks, reads within 1 kb of the CBP-bound enhancer center are included. For the TFs, reads within 100 bp of the enhancer center are included. Read numbers are normalized and input-subtracted as described in Methods. Rho is the Spearman correlation coefficient. B*N* refers to biological replicate *N*.

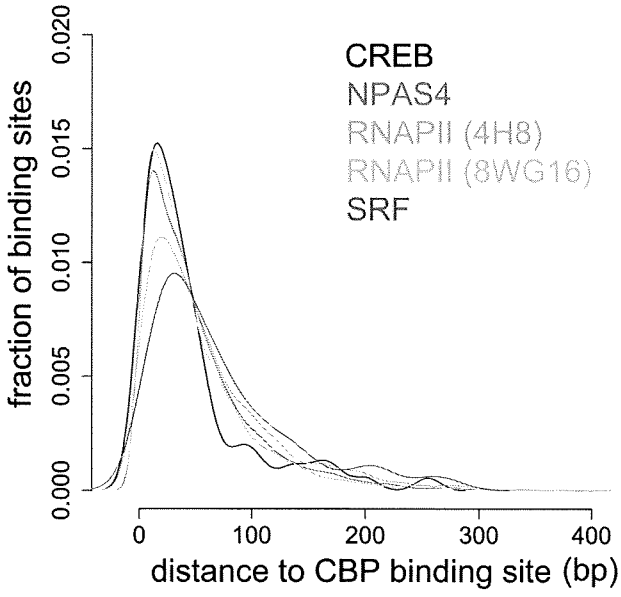


Supplementary Figure 3. Heatmap showing the binding profiles of methylated histones and transcription factors at 315 selected extragenic enhancers. The same 315 enhancers shown in fig. 5a are repeated in the same order for each factor. The color represents the strength of binding, from dark green (strongly de-enriched compared to input) to dark red (strongly enriched compared to input). A similar plot for eRNAs displays forward (blue) and reverse (magenta) reads from the total RNA fraction across the same set of 315 enhancers, as also shown in fig. 5a. Note that in the RNAPII experiments, 8WG16 antibody was used.

a

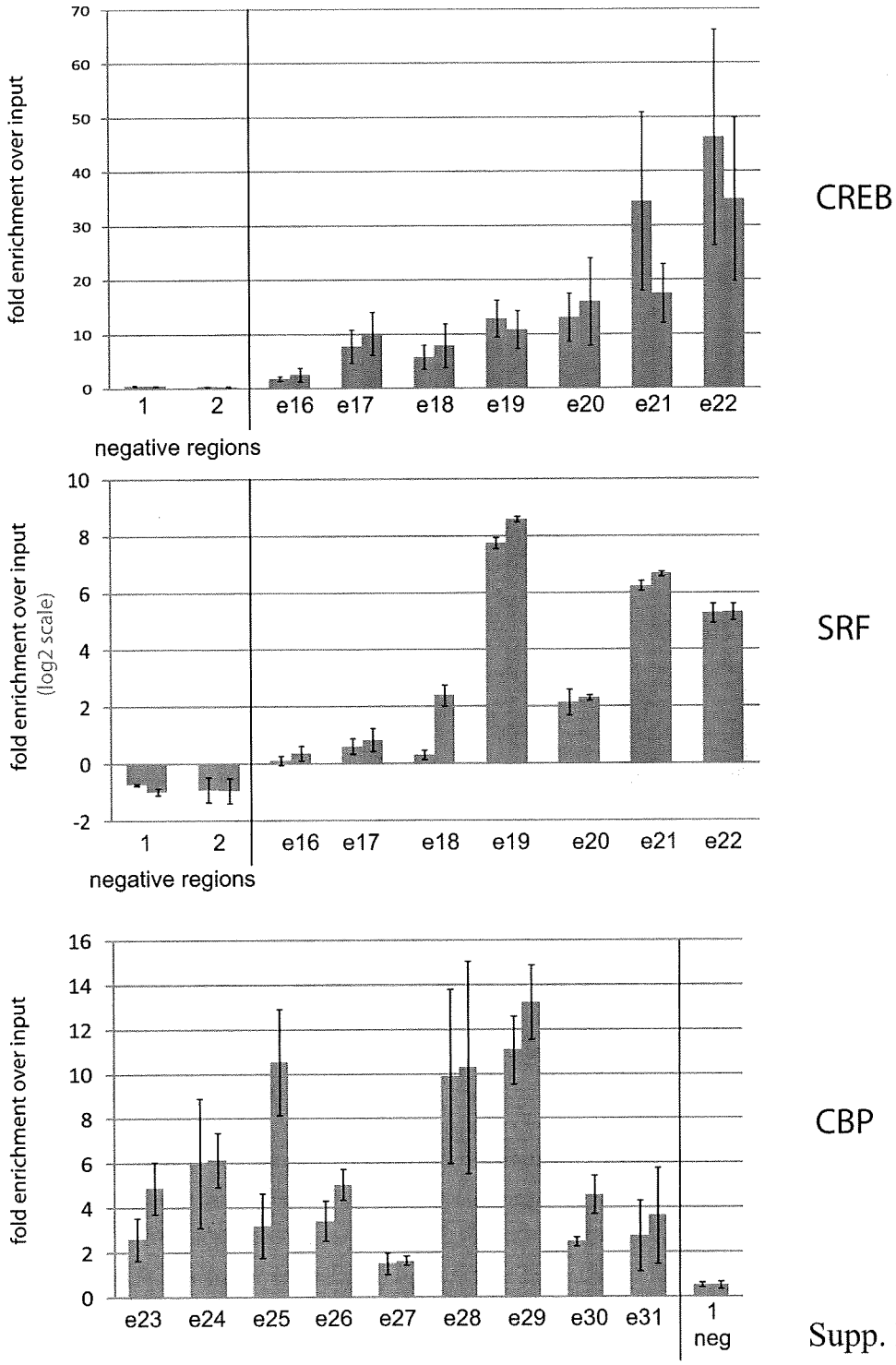


b



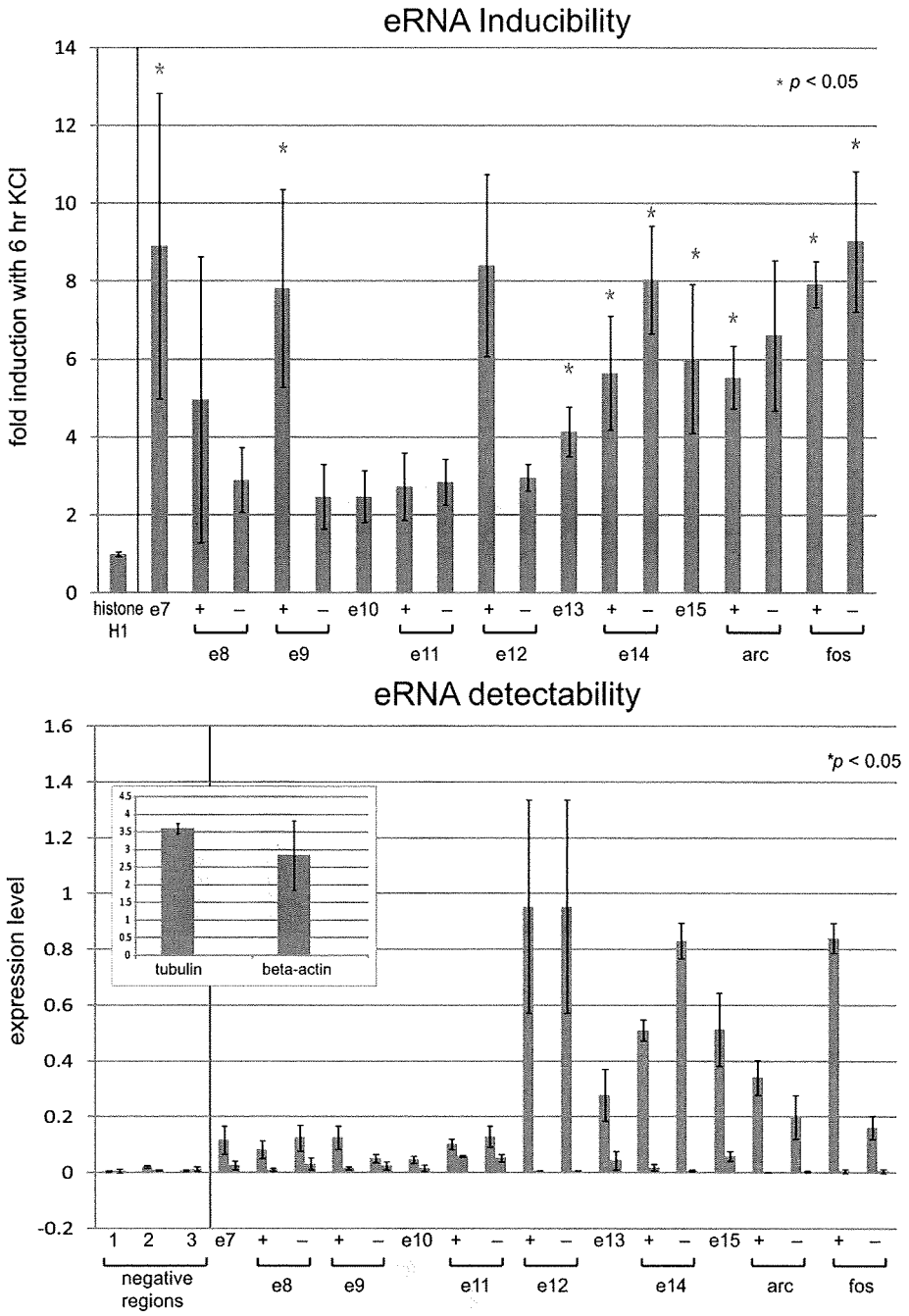
Supp. Fig. 4

Supplementary Figure 4. Locations of TF and RNAPII binding sites. a, Binding sites for each TF (or RNAPII) are grouped into three categories: promoter-proximal (within 1 kb of 25,562 annotated RefSeq TSSs), enhancer-proximal (within 1kb of 11,835 CBP-bound enhancer centers), or other (*i.e.*, not in the previous two categories). The actual number of peaks in each category is also shown. 4H8 and 8WG16 are anti-RNAPII antibodies. **b,** Distribution of the distances from each TF or RNAPII binding site to the CBP peak center. The *x*-axis denotes the distance (bp) between the factor binding site and the CBP binding site. The *y*-axis denotes the fraction of binding sites of a given TF at any given distance to CBP.

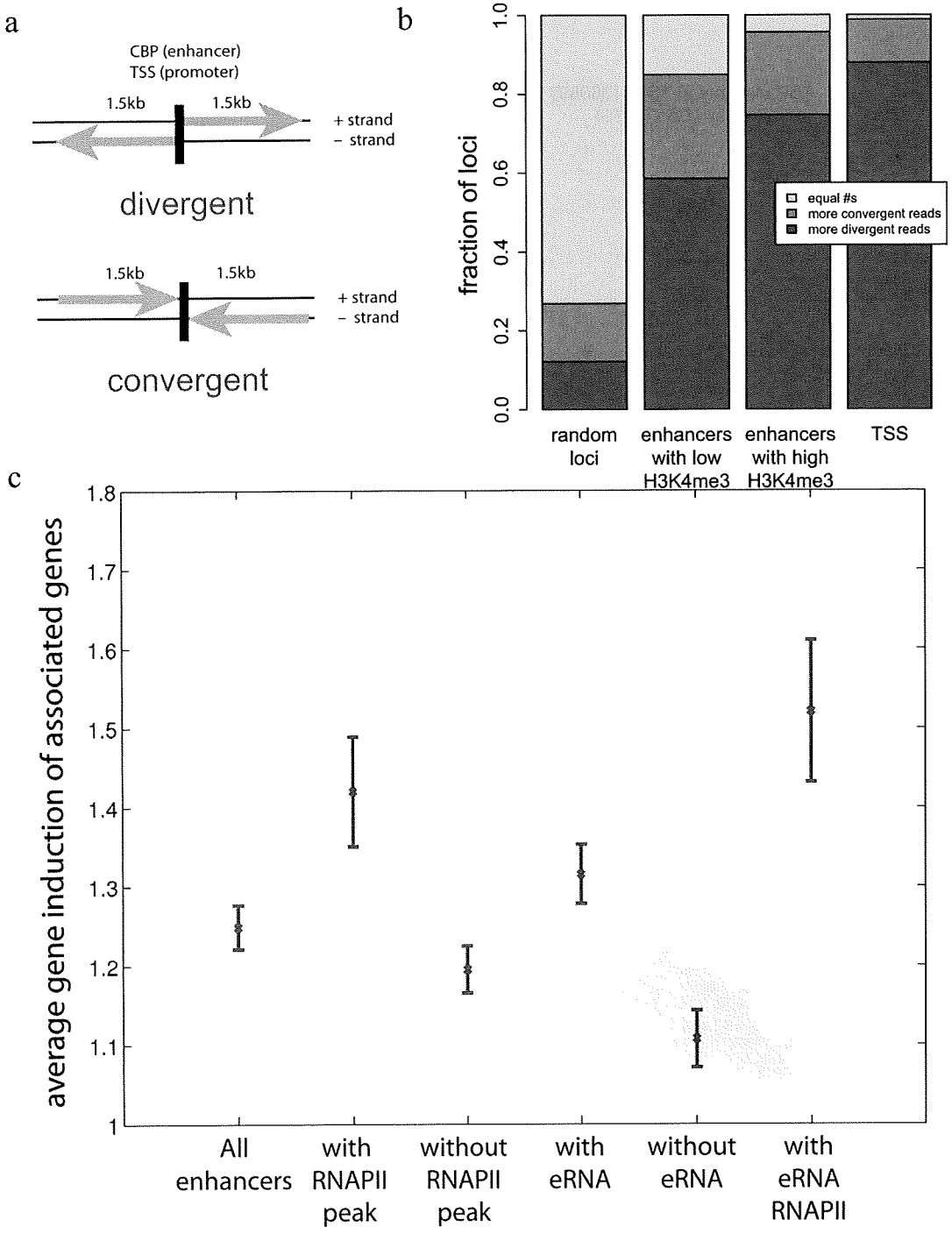


Supp. Fig. 5

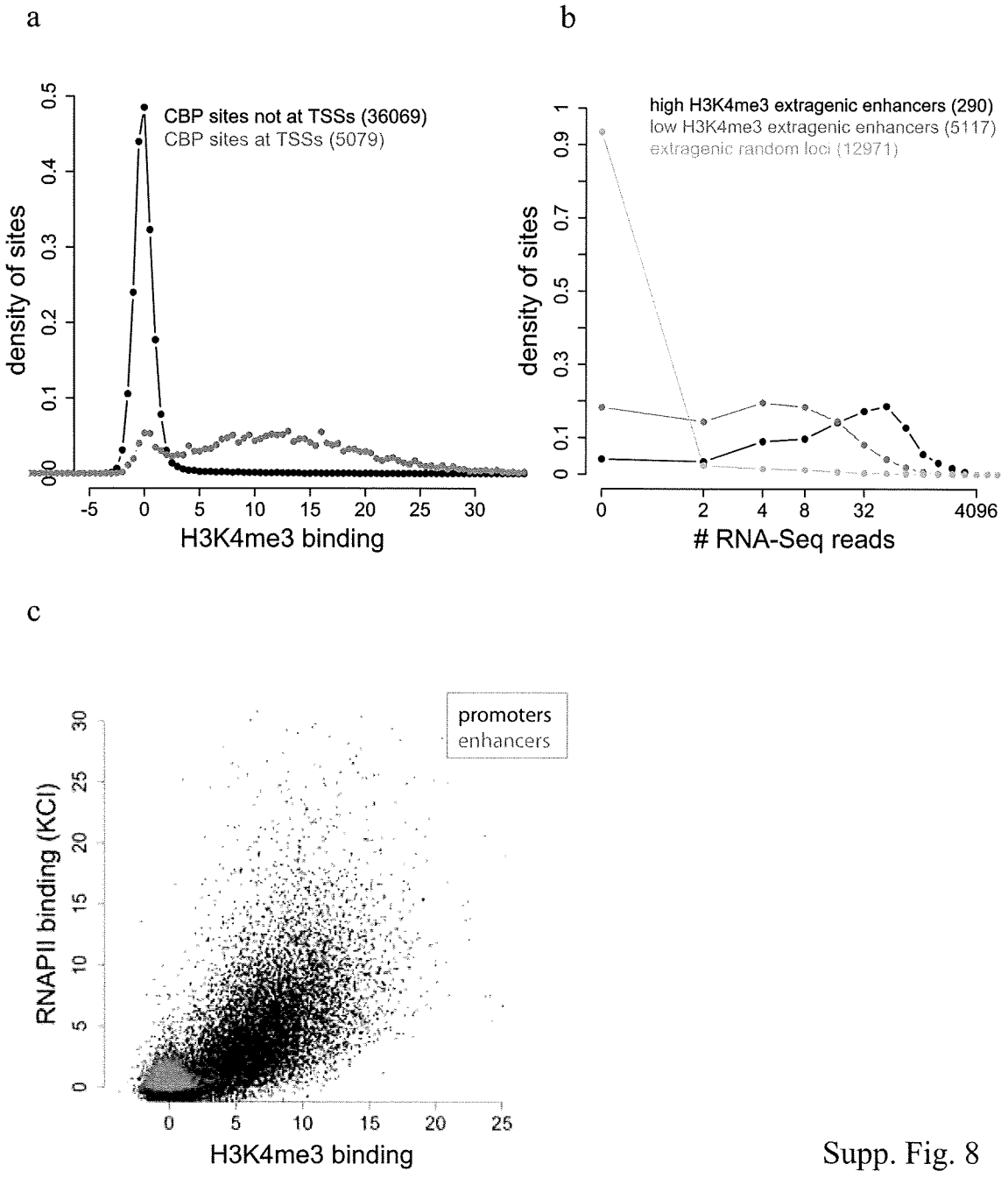
Supplementary Figure 5. ChIP-qPCR validation of randomly chosen CREB, SRF, and CBP binding sites. Blue and red bars represent qPCR signal from immunoprecipitated DNA over input DNA from 0 and 2 hour KCl-treated neurons, respectively. Labels e16-e31 designate the binding sites identified in our ChIP-Seq experiments (Methods, Supplementary Table 10). Two negative regions were chosen for an absence of binding in ChIP-Seq experiments. Errors bars, s.e.m. (n=4 biological replicates).



Supplementary Figure 6. eRNA Validation after DNase I treatment. eRNA inducibility and detectability as measured by RT-qPCR using random hexamer-primed reverse transcription. Candidate eRNA transcripts were selected randomly with the exceptions of the c-*fos* and *arc* enhancers. In the inducibility plot, histone mRNA transcript whose expression is not induced by membrane-depolarization was included as a negative control. + and – refer to regions upstream (-) or downstream (+) of the CBP-bound enhancer center, where upstream is defined as 3' on the – genomic strand and downstream is defined as 3' on the + genomic strand. For the enhancers e7, e10, e13 and e15, only the downstream regions (+) were tested for the presence of eRNA. In the detectability plot, blue bars indicate qPCR signal of each eRNA transcript relative to a genomic DNA standard used across all samples. Thus, the values in *y*-axis represent a relative measure of gene expression. The red bars indicate RT- controls in which the RT reaction lacked the RT enzyme. All tested eRNAs were detectable significantly above the mean background level observed in the negative regions (1-3) based on a *t*-test. Error bars, s.e.m. (n=3 biological replicates)

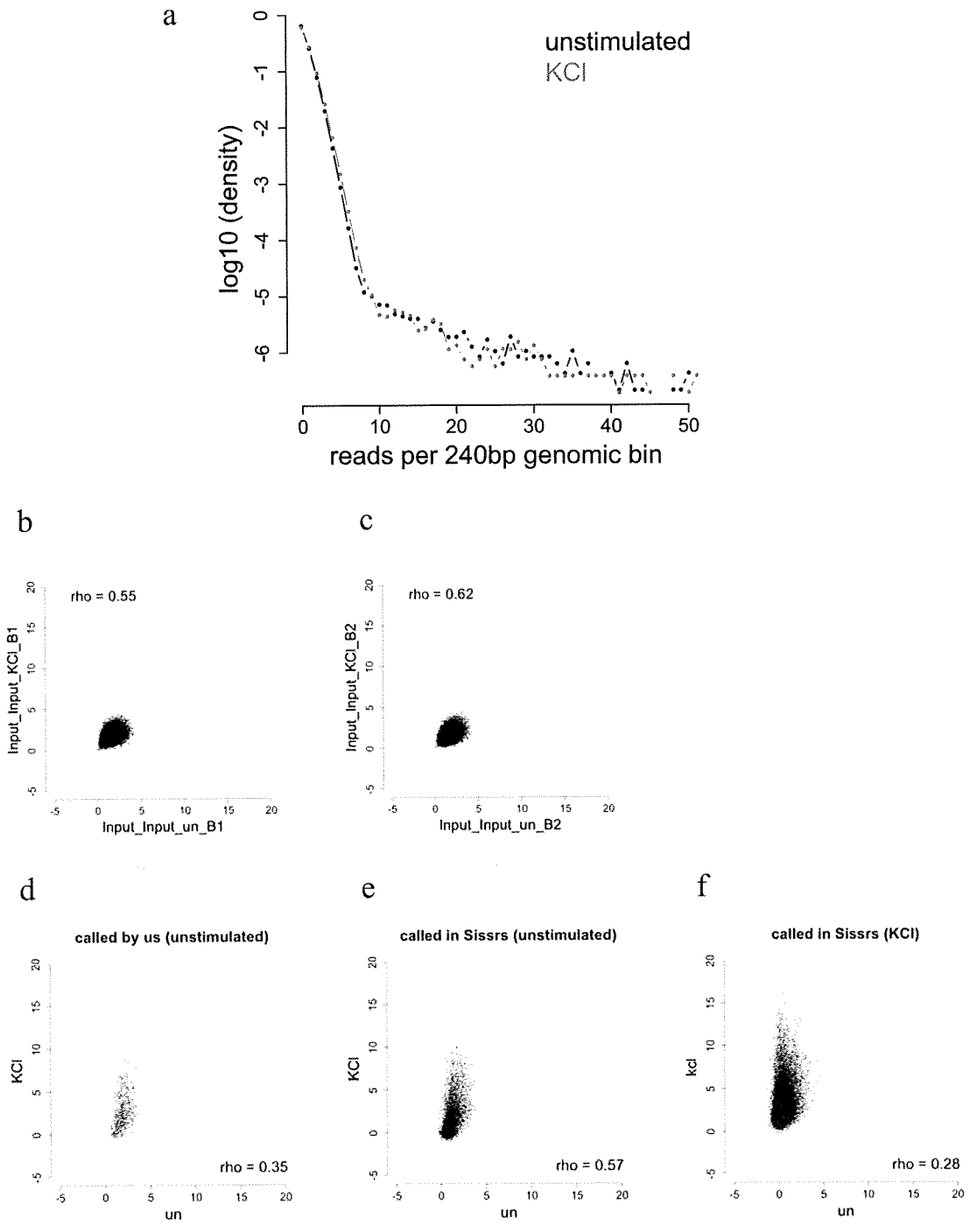


Supplementary Figure 7. Transcription at enhancers appears to initiate from the CBP-bound enhancer center. a, Schematic illustration of convergent and divergent reads with respect to a given locus. "Transcriptional noise" resulting from transcription initiated at other loci is predicted to result in equal numbers of divergent and convergent reads with respect to enhancer centers or TSSs. Transcription initiated from a CBP-bound enhancer center or TSS will be mostly divergent. **b,** Enhancers have more divergent than convergent reads, which is consistent with the notion that they initiate transcription. Only extragenic enhancers and random regions were included in this analysis. **c,** The average induction level for genes associated with various sets of enhancers.



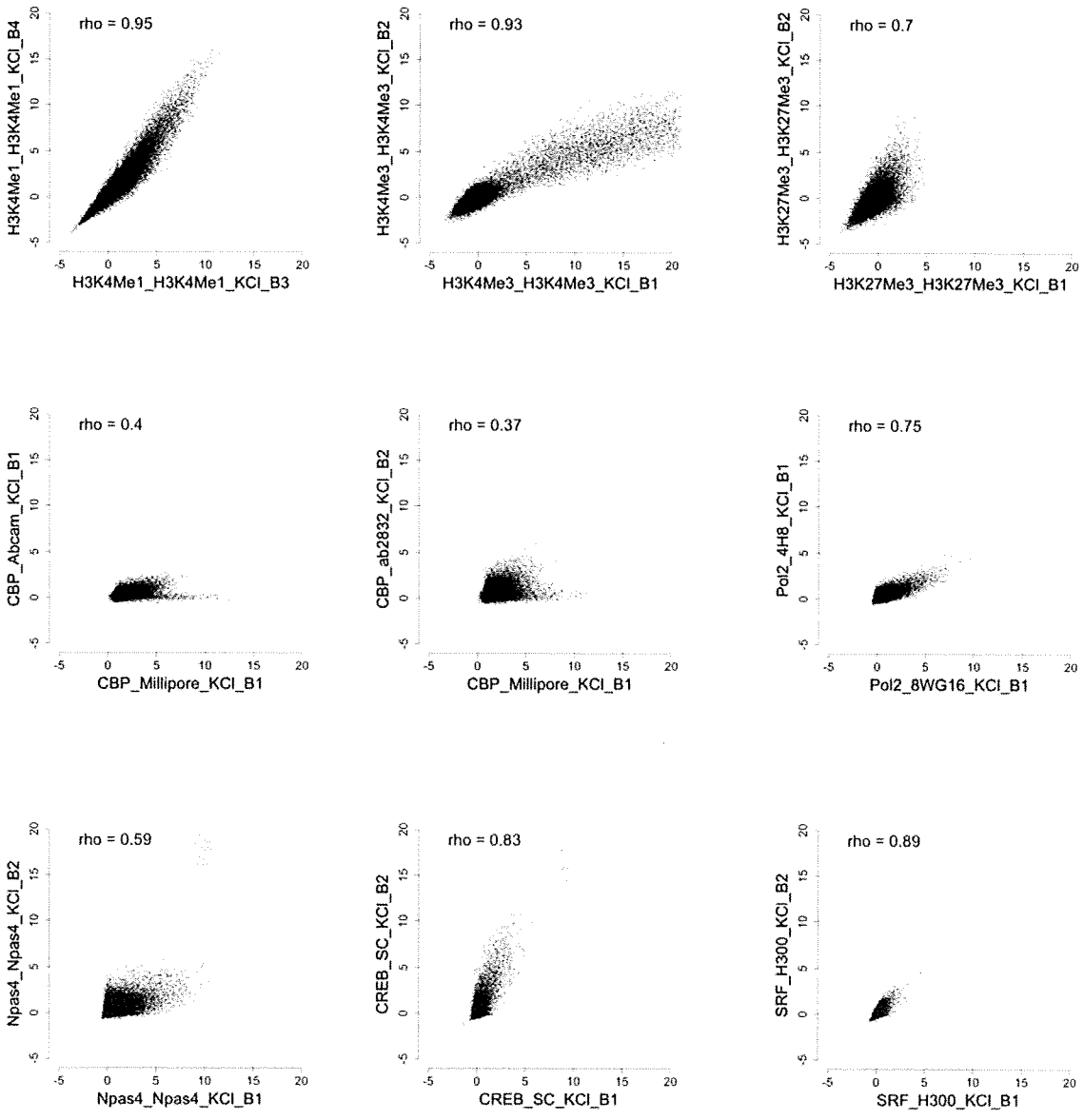
Supp. Fig. 8

Supplementary Figure 8. Characterization of H3K4me3 detected at enhancers. a, Binding profiles of H3K4me3 near CBP (Millipore) peaks (Supp. Tbl 5). The CBP peaks found near RefSeq, Ensembl, or UCSC-annotated TSSs (red) exhibit a bimodal distribution of H3k4me3 levels, whereas most of the other CBP sites exhibit low levels of H3K4me3 binding. The cut-off for separating high and low H3Kme3 peaks at enhancers is defined as the local minimum between the two modes of the red curve found at $x = 2$. **b,** The distribution of the numbers of RNA-Seq reads found within 1.5 kb of random loci (green) or extragenic enhancers with low (red) or high (black) levels of H3K4me3, as in Fig. 5c. **c,** Levels of H3K4me3 and RNAPII binding at promoters. Each black dot represents a TSS, and each red dot represents one of 5,117 extragenic enhancers with low H3K4me3 or one of 290 extragenic enhancers with high H3K4me3. The anti-RNAPII antibody used was 8WG16.



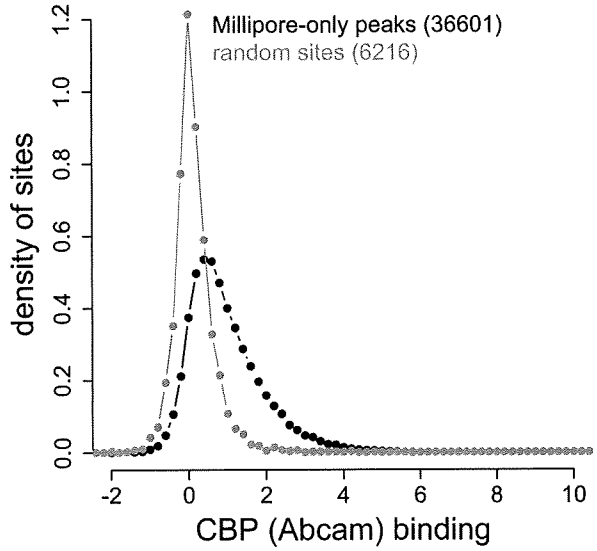
Supplementary Figure 9. Characterization of the input control and CBP peak finding using Sissrs. **a,** The distributions of input ChIP-Seq read numbers from unstimulated and KCl-treated neurons are very close to a Poisson distribution with deviations appearing only in 240 bp bins containing more than 10 reads. **b-c,** The number of input reads found near the 41,148 CBP peaks before and after membrane depolarization for two biological replicates (B1, B2). **d,** Levels of CBP before and after membrane depolarization at the 1,100 CBP peaks detected in the unstimulated condition using our peak finding method on the Millipore CBP data (Methods, Supp. Tbl 5). (These peaks were not included on our enhancer list unless they were

independently detected in the membrane-depolarized condition). **e,** Levels of CBP before and after membrane depolarization at the 8,980 CBP peaks detected in the unstimulated condition (Millipore) using Sissrs. **f,** Levels of CBP before and after membrane depolarization at the 32,656 CBP peaks detected in the stimulated condition (Millipore) using Sissrs. In all three cases for **d-f**, it is clear that the peaks are strongly inducible (e.g., compare to CREB and SRF in Supplementary Figure 2).

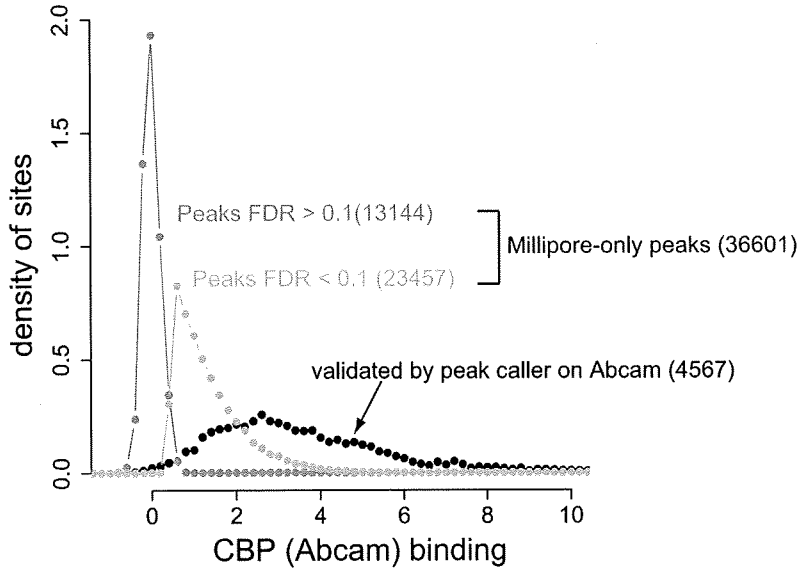


Supplementary Figure 10. Biological replicates of ChIP-Seq experiments. Each dot represents the normalized number of reads found near one of the 41,148 CBP peaks for each of the biological replicates. Rho is the Spearman correlation. The axes represent normalized, input-subtracted ChIP-Seq read counts. B*N* refers to biological replicate *N*.

a



b



Supplementary Figure 11. Reproducibility of CBP peaks. **a,** CBP binding levels determined by ChIP-Seq with the Abcam CBP antibody at 36,601 loci, which were initially defined as CBP binding sites by the Millipore CBP antibody but that were not replicated by peak-calling on the Abcam CBP antibody (black). Many of these loci are still significantly enriched compared to 6,216 random regions (red), suggesting that they are *bona fide* CBP binding sites but were not called by our peak calling method (from the Abcam data) due to insufficient sequencing depth (Methods). **b,** Using a 0.1 FDR for the distribution in (**a**), 23,457 out of 36,601 CBP Millipore-only peaks are statistically validated, and the figure shows the distribution of the level of Abcam binding for the different peak categories.

神経疾患と細胞骨格

竹本-木村さやか*　上田(石原)奈津実*　布施俊光*　上條諭志*　尾藤晴彦*

神経回路網の発達期には，幼弱神経細胞の移動，軸索・樹状突起の形成・伸展がまず起こる．このような神経細胞の形態形成過程ではアクチンや微小管などの細胞骨格系の貢献が大である．さらに，伸展した軸索が投射先の樹状突起と遭遇し，シナプスが形成・成熟する際にも，非常に動的な細胞形態制御が不可欠となる．最近では，成熟回路におけるシナプス機能についても，絶えず再編成しつづけ，動的に維持されていることが明らかとなっている．このようなダイナミックな神経細胞骨格制御がいったん破綻すると，どのような疾患が起こるのか．本稿では細胞骨格機能障害が引き起こす分子病態と最終的な脳高次機能異常について概説する．

はじめに

近年，ヒト遺伝性脳高次機能異常の家系研究から，種々の原因候補遺伝子が単離同定されてきた．興味深いことに，原因遺伝子の多くは，脳発達の途上で神経回路形成を制御すると考えられる細胞骨格制御因子である．神経回路網の発達期には，初期の細胞移動から，後期のシナプス形成期にいたるまで，数多くの神経細胞形態形成制御過程が存在する．さらに，神経回路の成熟後も，シナプス機能が機能的・形態的に維持され，再編成しつづけることが最近の知見により明らかとなっている．このように考えると，アクチン・微小管などの細胞骨格機能の遺伝的障害が，脳高次機能異常に結びついていることは偶然ではなく，必然であるとすらいえよう．本稿では，これまでに明らかになっている神経細胞骨格の制御メカニズムとその意義について簡潔に紹介し，さらに，いくつかの具体的な疾患について，その細胞骨格機能障害が引き起こす分子病態と最終的な脳高次機能異常について概説する．

Key Words
アクチン
微小管
シナプス
レセプタートラフィッキング
Rho

1．神経回路発達における細胞骨格制御の意義

20世紀初頭に，Ramón y Cajalはニューロン・ドクトリンという仮説を著し，そのなかで，脳神経系が数多くの神経細胞種から成り，これら多数の細胞が融合せず，複雑な接合(すなわちシナプス)を互いに形成することにより，その複雑な脳機能を果たしている可能性を明らかにした[1]．この考えは，筋肉という興奮性組織が，むしろ筋細胞融合によって協調的収縮という機能を獲得するのとは，まったく様相を異にするということを喝破した点で画期的であった．

* TAKEMOTO－KIMURA Sayaka, AGETA－ISHIHARA Natsumi, FUSE Toshimitsu, KAMIJO Satoshi, BITO Haruhiko/東京大学大学院医学系研究科　神経生化学教室

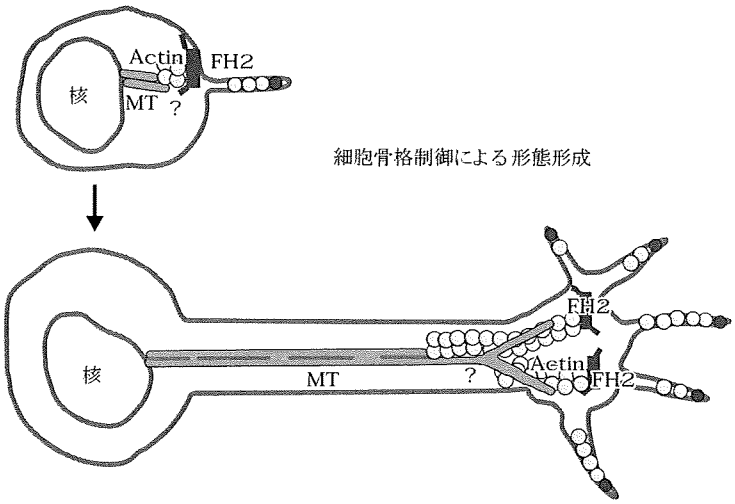


図　微小管 MT)とアクチン細胞骨格 Actin)の相互作用による神経細胞の形態形成制御 Bito H, 2003[18]より改変引用)
両骨格の協調には，フォルミンタンパク質 FH2)などが関与していると考えられている．

それから 100 年以上が過ぎ，シナプスという神経回路接合部分の形成と機能構築に関する知見は飛躍的に進展した．脳神経系の発達の途上では，①終末分化終了後に，細胞種特異的な転写因子群の発現制御により，興奮性・抑制性などの神経細胞の分化が規定され，つづいて，②主として微小管制御にもとづいた細胞移動に依拠した脳領野ならびに大脳皮質層構造の構築が進み，③細胞極性の成立と同時に，軸索ならびに樹状突起とよばれる 2 種類の機能的に分化した突起を伸ばし，軸索先端部が標的となる神経細胞の樹状突起へ目掛けて投射する．このような 3 つの過程が，シナプス形成に先立つ神経回路形成の根幹をなすと考えられている[2]．

そして最終的には，軸索の先端の神経終末には，神経伝達物質放出を司る分子装置が整備され，樹状突起スパイン上の後シナプス肥厚部 post synaptic density: PSD)とよばれる特殊な構造と対をなし，細胞間接着装置であるシナプスを形成する．こうして神経伝達物質の放出と受容を介して情報を一方向的に伝達する機構が完成する[2]．

このように，神経細胞の移動，軸索および樹状突起の形成・伸展や，シナプス形成過程には，アクチン・微小管細胞骨格による動的制御が不可欠と考えられている（図）[18]．

2．成熟神経回路の形成後の動的細胞骨格の役割

一方，成熟神経回路形成が完了した後は，神経細胞の形態やシナプス結合は大きく変化せず，細胞骨格系は静的な構造支持装置であるという考えが長らく主流であった．しかしながらたとえば Donald O Hebb は，シナプス可塑性を予言する考察のなかで，シナプスの増減を説明しうる機構として活動依存的な突起の動的な新生・消退を予見した[3]．長期増強や長期抑圧などのシナプス可塑性の実在が確認された 1980 年以降，実際にシナプスを含む神経形態の活動依存的変化を実証することが急務となった．興奮性シナプスの多くは，スパインとよばれる樹状突起上の微小な突起に形成されているが，最近シナプス伝達の長期増強 long term potentiation: LTP)に伴いスパイン，スパイン頸部，スパイン数などが変化する様相がついに明らかにされた[4]．このような過程においては，シナプス維持とスパイン形態制御とその協調性がきわめて重要であり，シナプス後部の分子構築が，動的なアクチン骨格と PSD 分子間の相互作用により，規定されていることが示されている[5][6]．

3. 脳発達期における神経細胞骨格制御の破綻がもたらす脳機能異常

近年，ヒトにおける遺伝性の脳高次機能異常（精神遅滞，特異的認知障害など）や精神疾患の家系研究から，さまざまな原因候補遺伝子が同定されるようになった．すると驚くべきことに，これらの原因遺伝子のうちの少なからずが，神経細胞骨格制御にとって重要と考えられる分子をコードする例が次々と報告された．

たとえば脳のしわが欠失する広範な脳病変を伴う滑脳症は，胎生期の脳発達における細胞移動の異常を基礎病変としている遺伝性疾患である．その原因遺伝子は近年LIS1やdoublecortinといった微小管結合タンパク質であることが明らかになり，機能的神経回路形成初期における微小管制御の重要性が改めて証明された．さらに，統合失調症家系の解析から注目されている*DISC-1*遺伝子産物が，これら微小管結合タンパクと分子複合体を形成することから，脳発生における神経細胞の移動不全が精神疾患発症に結びつく可能性が示唆されてきている[7]．

この後の神経回路形成最終段階では，軸索末端と樹状突起のあいだでシナプスが形成される．このとき，シナプス前膜とシナプス後膜のあいだでは，細胞接着分子を介した積極的な会合が形成されると考えられている．Neurexin/Neuroligin複合体や，これとシナプス後部において結合するShank 3の変異を伴う家族性自閉症の存在が明らかになり，Shankはcortactinを介してシナプス後部のアクチン骨格と結合していることから，形成されたシナプスの数や位置の制御破綻が，自閉症発症の一因となっているのかもしれない．

一方，古くから，精神遅滞とシナプスが局在している構造であるスパインの形態の異常の関係を示唆する病理所見の報告が知られていた．しかしその因果関係がもっとも最初に明らかになったのは，伴性遺伝性精神遅滞と低分子量Gタンパク質制御の関係の研究を通じてである．伴性遺伝性精神遅滞の候補原因遺伝子としてRhoGAP（GTPase activating protein）分子であるoligophrenin-1，Cdc42/Racの下流のエフェクターであるPak3（p21-activated kinase3），RhoGEF（guanine nucleotide exchange factor）であるARHGEF6などが同定された[8]が，実際これら遺伝子を改変したマウスにおいて，脳発達期におけるスパイン形態の異変が顕著に認められた[9][10]．さらに，外界からの多様な感覚入力を担う神経回路の形成と成熟には，神経突起のダイナミックな進展退縮が必要であり，アクチン重合制御やアクトミオシン系による物質輸送制御が不可欠と考えられている．たとえば，知覚回路の異常の例としては，無症候性遺伝性難聴DFNA1があげられる．この原因遺伝子としてRhoエフェクターのmDia1が同定されている[11]．同様に，Williams症候群は，多くの発達段階にかかわる遺伝子群の欠失により生じるが，とくに*LIMK1*遺伝子の欠損を含む患者では，視空間形成認知機能の特異的な障害が生じる．LIMK1は，Rho結合キナーゼROCKの基質であり，スパインや軸索先端の成長円錐におけるアクチン重合を制御していると考えられている．実際に，LIMK1ノックアウトマウスにおいては，樹状突起形成異常と神経可塑性異常などが報告されている[12]．

このような知見から，神経回路網がマクロ的にもミクロ的にも正確に形成される過程に関与する多くの分子の破綻により，脳高次機能異常が引き起こされることが明らかとなり，脳発達期における正確な細胞骨格制御の重要性が裏付けられた．

4. 成熟神経回路における細胞骨格制御破綻のもたらす病態

では，微小管あるいはアクチンからなる神経細胞骨格の機能異常が，成熟脳にて発生した場合は，何が起こるのであろうか．もっとも顕著な例として，アルツハイマー病があげられる．この病変の本質が神経原線維変化における過剰リン酸化tauの蓄積であるという発見以来，神経細胞の病態・代謝異常の最終結末として，重篤な細胞骨格障害をきたし，結果として神経変性や脳機能異常が生じると考えられた．

さらに，アルツハイマー病発症初期のベータアミロイド蓄積においても，アクチン細胞骨格の関与が想定されている．たとえば，LIMKによるcofilinリン酸化が増強され，その結果神経変性が増悪する発端となる可能性が報告されている[13]．ほぼ同様の機構により，アクチン重合制御因子profilinのROCKによるリン酸化により，