




# Activity-dependent regulome of human GABAergic neurons reveals new patterns of gene regulation and neurological disease heritability

Gabriella L. Boulting <sup>1,8</sup>  , Ershela Durresi <sup>1,2,7,8</sup> , Bulent Ataman<sup>1,8</sup> , Maxwell A. Sherman <sup>3,4,8</sup> , Kevin Mei <sup>1</sup> , David A. Harmin<sup>1</sup> , Ava C. Carter<sup>1</sup> , Daniel R. Hochbaum <sup>1</sup> , Adam J. Granger<sup>1</sup> , Jesse M. Engreitz<sup>5</sup> , Sinisa Hrvatin <sup>1</sup> , Michael R. Blanchard<sup>1,6</sup> , Marty G. Yang<sup>1</sup> , Eric C. Griffith <sup>1</sup> and Michael E. Greenberg <sup>1</sup> 

**Neuronal activity-dependent gene expression is essential for brain development. Although transcriptional and epigenetic effects of neuronal activity have been explored in mice, such an investigation is lacking in humans. Because alterations in GABAergic neuronal circuits are implicated in neurological disorders, we conducted a comprehensive activity-dependent transcriptional and epigenetic profiling of human induced pluripotent stem cell-derived GABAergic neurons similar to those of the early developing striatum. We identified genes whose expression is inducible after membrane depolarization, some of which have specifically evolved in primates and/or are associated with neurological diseases, including schizophrenia and autism spectrum disorder (ASD). We define the genome-wide profile of human neuronal activity-dependent enhancers, promoters and the transcription factors CREB and CRTCl. We found significant heritability enrichment for ASD in the inducible promoters. Our results suggest that sequence variation within activity-inducible promoters of developing human forebrain GABAergic neurons contributes to ASD risk.**

The human brain is complex, with expanded cellular diversity, increased neuronal connectivity and protracted stages of development compared to other mammals<sup>1–4</sup>. These differences might contribute to human-specific characteristics such as refined digital dexterity, symbolic language and abstract thought<sup>2</sup>. However, the manner in which evolutionary changes in development give rise to human-specific or even primate-specific characteristics remains largely uncharacterized.

The development and maturation of the brain is dependent on neuronal activity in a wide range of animal species<sup>5</sup>. In addition to protein modification, receptor trafficking and local translation mediating the effects of neuronal activity, a range of evidence also points to a key role for activity-driven gene expression changes in the control of mammalian neural development. In this regard, synaptic activity-induced depolarization of the post-synaptic cell triggers expression of a largely stereotyped pattern of early-response gene (ERG) transcription factors (TFs), which, in turn, drive second-wave late-response gene (LRG) programs encoding synaptic effector molecules that diverge substantially among neuronal cell types.

Human brain development, which follows a particularly long timeline, provides ample opportunity for sensory input to shape circuit architecture. However, the limited availability of primary human brain tissue in which neuronal activity can be acutely manipulated

has largely precluded molecular characterization of these responses in humans. We previously identified *Osteocrin* (*OSTN*), a gene that encodes a secreted peptide in the bones and muscles of mice, as selectively induced by neuronal activity in human and macaque, but not mouse or rat, cortical excitatory neurons<sup>6</sup>, and showed that the induction of *OSTN* in primates relies on a series of base pair (bp) changes within an otherwise conserved enhancer sequence.

Forebrain GABAergic neurons (GNs) are another important neuronal subclass that are thought to facilitate circuit computations and behavioral adaptations unique to humans. GNs regulate activity-dependent synapse formation during developmental critical periods, maintain excitatory–inhibitory circuit balance and generate network oscillations, among other critical functions<sup>7</sup>. Nuclei of the basal ganglia are composed almost exclusively of GNs and are critical for behaviors such as motivation, motor coordination, planning, attention, memory and learning<sup>8</sup>. Moreover, basal ganglia dysfunction has long been implicated in movement disorders, including Huntington's disease and Parkinson's disease. More recently, circuits involving the basal ganglia have been associated with behavioral disorders such as Tourette's syndrome, obsessive compulsive disorder and ASD<sup>9</sup>.

Characterization of human GN activity-dependent chromatin dynamics and transcriptional responses might, thus, be particularly important to the understanding of key aspects of human brain

<sup>1</sup>Department of Neurobiology, Harvard Medical School, Boston, MA, USA. <sup>2</sup>MD-PhD Program, Harvard Medical School, Boston, MA, USA. <sup>3</sup>Division of Genetics & Center for Data Sciences, Brigham and Women's Hospital, Boston, MA, USA. <sup>4</sup>Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA, USA. <sup>5</sup>Department of Genetics, Stanford University School of Medicine and BASE Initiative, Betty Irene Moore Children's Heart Center, Lucile Packard Children's Hospital, Stanford, CA, USA. <sup>6</sup>Neurobiology Imaging Facility, Department of Neurobiology, Harvard Medical School, Boston, MA, USA. <sup>7</sup>Present address: Department of Internal Medicine, NYU Langone Medical Center and NYU School of Medicine, New York, NY, USA. <sup>8</sup>These authors contributed equally: Gabriella L. Boulting, Ershela Durresi, Bulent Ataman, Maxwell A. Sherman.

✉e-mail: [gabriella\\_boulting@hms.harvard.edu](mailto:gabriella_boulting@hms.harvard.edu); [michael\\_greenberg@hms.harvard.edu](mailto:michael_greenberg@hms.harvard.edu)

function. To this end, we employed human induced pluripotent stem cell (hiPSC)-derived cultures highly enriched for developing forebrain GABAergic neurons (hGNs) to profile activity-induced changes in gene expression, histone modifications and TF binding at various time points after membrane depolarization. In addition, to gain insights into neurological disorder etiology, we leveraged these datasets to model neurological disease heritability enrichment in subsets of hGN promoters and enhancers, uncovering a significant enrichment for ASD heritability specifically in activity-inducible promoters that are also enriched for cyclic AMP-responsive element-binding protein 1 (CREB) binding. These results suggest that disruptions to activity-dependent gene expression programs within hGNs, particularly within developing striatal cell types, might play a role in ASD, and studies of activity-dependent gene regulation could facilitate functional understanding of human disease-associated genetic variation.

## Results

**hGNs model developing forebrain GNs.** We found that hGNs differentiated from the single hiPSC line 1434 (single genotype (1G); Supplementary Table 1) are strongly enriched for membrane depolarization-responsive GNs. Immunostaining demonstrated that hGNs were highly enriched for neuronal as opposed to glial and neural progenitor cell types ( $92.9\% \pm 7\%$ ,  $n=2$ , MAP2<sup>+</sup> cells; Fig. 1a,b) ( $1.5\% \pm 1.1\%$ ,  $n=3$ , NESTIN<sup>+</sup> cells; 0% GFAP<sup>+</sup> cells,  $n=1$ , >10,000 cells; Extended Data Fig. 1a). RNA sequencing (RNAseq) of hGN cultures confirmed strong expression of canonical GN marker genes (Fig. 1e), with a gene expression profile typical of neurons of the ventral telencephalon and only minimal expression of non-neuronal gene markers (Fig. 1e and Extended Data Fig. 1c).

hGNs could be grown reproducibly, and were stably viable over more than ten different production lots for at least 3 weeks in culture. Upon exposure to 55 mM potassium chloride (KCl)-induced membrane depolarization, hGNs exhibited robust induction of activity-dependent gene expression as measured by quantitative real-time polymerase chain reaction (qPCR) (Fig. 1c). By whole-cell patch-clamp electrophysiology, we confirmed that hGNs displayed polarized membranes, fired at least single action potentials in response to current injection and formed GABAergic synapses (Fig. 1d and Extended Data Fig. 1b), consistent with previous reports<sup>10–12</sup>.

We used single-cell RNA sequencing (scRNAseq) to characterize the diversity of GN subtypes present in the hGN cultures and obtained data from 37,101 cells derived from three independent differentiation lots (Fig. 1f). This analysis confirmed broad expression of neuronal, ventrally patterned telencephalic marker genes across almost all cell clusters (Fig. 1g and Extended Data Fig. 2a,g). Neuronal clusters 0–2, 4, 5 and 9–11 displayed gene expression patterns typical of the developing striatum, with a subset of those clusters also enriched for gene expression characteristic of more mature spiny projection neurons (SPNs) (Fig. 1f,g and Extended Data Fig. 2b,c,h,i)<sup>13,14</sup>. A minority of hGN cell clusters expressed genes indicative of pallial (clusters 6 and 14), neural mitotic progenitor (clusters 8 and 12) or medial ganglionic eminence-derived neural subtypes (cluster 13) (Fig. 1f,g and Extended Data Fig. 2d,e,j)<sup>14,15</sup>. Two neuronal clusters also exhibited co-enrichment of *ZIC* gene expression (clusters 3 and 7), a characteristic of GNs of the septum regions of the developing forebrain<sup>16</sup> (Fig. 1f and Extended Data Fig. 2f).

To confirm the reproducibility of this culture system across multiple human genotypes, we compared total RNAseq from hGN cultures derived from the original hiPSC line with analogous cultures prepared from hiPSC lines obtained from three independent human donors (four genotypes: 4G; Supplementary Table 1). hGN cultures from all four independent genotypes expressed markers of ventral telencephalon-derived GNs (Extended Data Fig. 1c). Spearman correlation after quantile normalization of expressed genes showed high correlations across all pairwise comparisons of human

genotypes ( $r_s=0.9695–0.9853$ ). Uniformity across genotypes was also confirmed by immunostaining, which showed that most cells across the four genotypes were neuronal (MAP2<sup>+</sup>), with only a small fraction of the cells expressing the neuronal progenitor marker NESTIN (0–5.1%) (Extended Data Fig. 1d).

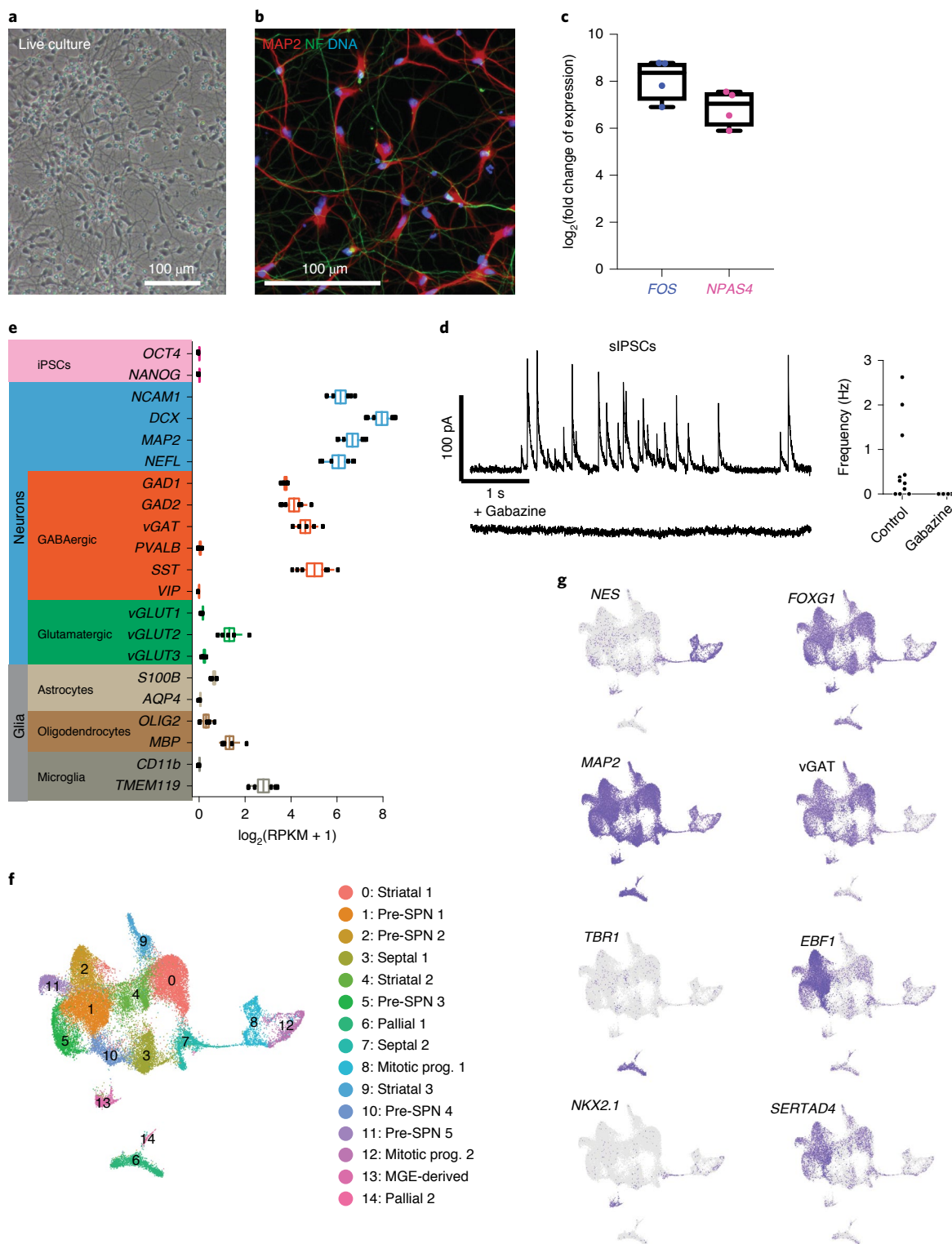
Together, our analyses identify hGNs as a depolarization-responsive predominantly post-mitotic GABAergic neuronal culture composed of cells that are similar to the embryonic ganglionic eminence- and septum-derived cell types that ultimately populate the cortex and basal ganglia as interneurons and striatal SPNs. Consistent with this conclusion, transcriptomic correlation analysis with 18 human brain regions across eight early developmental stages (<https://www.brainspan.org>) showed highest similarity to post-conceptional week (p.c.w.) 16 striatum (1G  $s_r=0.8366$  and 4G  $s_r=0.8474$ ; Supplementary Table 2). Thus, hGNs are a useful cell culture model for identifying activity-dependent genes in GNs of the human developing subcortical forebrain.

**The depolarization-responsive transcriptome of hGNs.** We identified activity-regulated transcripts in hGN cultures using RNAseq, comparing samples from unstimulated cultures to those harvested 15 min, 1, 2 and 4 h after exposure to elevated KCl (55 mM) (Fig. 2). Previous studies show that this stimulation paradigm leads to the induction of genes in cultured neurons that are also activated in the intact mouse brain in response to various physiological stimuli<sup>6,17,18</sup>.

We performed differential gene expression analysis using multiple independent lots of hGNs from the single original hiPSC line as biological replicates (1G; Fig. 2 and Supplementary Data 1), as well as using hGNs derived from the four hiPSC lines from independent donors as biological replicates (4G; Extended Data Fig. 3 and Supplementary Data 2). Calculating gene expression changes at each stimulation time point compared to the unstimulated condition and using the criteria of a  $q$  value consistent with a false discovery rate (FDR) of 0.05 and a fold-change threshold of 1.5, we identified a total of 1,077 unique inducible genes (1G = 655, 4G = 854, 432 genes in common,  $z$  score = 84.59; Supplementary Data 3).

Depolarization-induced gene expression in hGNs was found to broadly recapitulate that observed in other types of neuronal culture. Well-characterized ERG TFs, such as *NPAS4*, *FOS* and *NR4A1*, were induced in both the 1G and 4G datasets within 1 h of membrane depolarization (Fig. 2a and Extended Data Fig. 3a,b). We also observed rapid induction of the primate lineage-restricted genes *LINC00473* and *ZNF331*, previously identified in human neuronal cultures (Fig. 2a,b and Extended Data Fig. 3a)<sup>6,19</sup>. A set of LRGs was observed 2–4 h after membrane depolarization that includes genes encoding known inducible effector proteins, such as *BDNF*, *RGS2* and *SCG2* (Supplementary Data 1–3)<sup>20,21</sup>. Consistent with previous observations, most ERGs encoded nuclear proteins, whereas a large portion of LRGs encoded membrane-bound or secreted factors (Extended Data Fig. 3d). The induction of a subset of these transcripts was subsequently verified by qPCR (Extended Data Fig. 3c).

This analysis also identified membrane depolarization-regulated genes that appear to be selectively induced in developing primate or human GNs. Examples include *XIRP1*, an F-actin-binding protein, and *MMP1*, a member of the secreted matrix metalloproteinase family that degrades extracellular matrix components during development and tissue remodeling<sup>22</sup> (Fig. 2b, Extended Data Fig. 3c and Supplementary Data 1 and 3). Subsets of midbrain and hindbrain neurons have been reported to express *Xirp1* in the mouse brain (<http://mousebrain.org>)<sup>23</sup>, but stimulus-responsive *Xirp1* expression in neurons has not been previously reported. *MMP1* is expressed across the human brain during development (<http://hbatlas.org/>)<sup>22,24</sup>, including in the subplate and intermediate zones of the midgestational developing human cortex (<http://brain-span.org>)<sup>25</sup>, but expression of the *MMP1* mouse homologues *Mmp-1a* and *Mmp-1b* has not been previously observed in the



**Fig. 1 | hGNs express RNA signatures of developing forebrain GNs. a, b** Live (**a**; DIV 7) and immunostained (**b**; DIV 15) hGN cultures exhibit neuronal morphology and express the neuronal proteins MAP2 and NF. Images are representative of more than ten independent experiments. **c**, Fold induction of mRNA of the immediate-early genes *FOS* and *NPAS4* at 1 h after membrane depolarization of hGNs compared to unstimulated mRNA levels, measured by qPCR ( $n=4$ ; median and interquartile range). **d**, hGNs form functional synapses in culture and show spontaneous inhibitory post-synaptic currents (sIPSCs), which are blocked by gabazine. **e**, mRNA expression levels (from 1G total RNAseq) of hGNs, consistent with an enrichment of GABAergic neuronal identities as opposed to other brain cell types or induced pluripotent stem cells ( $n=6$ ; box center, mean; box minima and maxima, mean  $\pm 1$  s.e.; whiskers minima and maxima, mean  $\pm 1$  s.d.). **f**, UMAP visualization of hGN scRNAseq displaying eight striatal-like GABAergic neuronal cell clusters, three clusters of MGE- or septal-like derived GNs, and minority populations of glutamatergic neurons and mitotic progenitor cells. **g**, UMAP visualizations of hGN scRNAseq with cells colored purple if RNA transcripts were detected for developmental marker genes of neural progenitors (*NES*), post-mitotic neurons (*MAP2*), forebrain identity (*FOXG1*), GNs (*vGAT* or *SLC32A1*), pallial glutamatergic neurons (*TBR1*), MGE-derived neurons (*NKX2.1*) and developing striatal neurons (*EBF1* and *SERTAD4*).

brain (<http://mouse.brain-map.org>). *MMP1* and *XIRP1* could, thus, be important during the extended period of neuronal activity-dependent maturation that human neurons undergo during development; however, when and how during evolution these loci acquired this form of regulation still remains to be determined.

Particularly abundant among hGN activity-regulated genes were long intergenic non-coding (LINC) RNAs and genes encoding proteins containing a DNA-binding zinc finger (ZNF) domain, and several loci within these two gene families were found to be specific to primate genomes (Supplementary Data 4 and 5). Comparative genomic studies show that both LINC RNA and ZNF gene families underwent an expansion during primate evolution, primarily through the action of lineage-specific transposable elements and tandem duplication events, respectively<sup>26,27</sup>. In principle, these genes could enable primate- or human-specific feedback mechanisms that contribute to the rapid modulation of transcription in response to neuronal activity.

**scRNAseq reveals cell-type-specific membrane depolarization-inducible gene expression.** To ask how activity-regulated gene expression varies across the different cell types within our cultures, we performed scRNAseq on hGN cultures after exposure to elevated levels of KCl for 0, 1, 2 and 4 h (Supplementary Data 6). Across the four time points, the relative numbers of cells in each of the 15 clusters were similar within each replicate, indicating minimal batch effects (37,101 cells from 11 InDrops libraries; Extended Data Fig. 4a). Subsequent analysis showed that hGNs mirrored other cell types<sup>18</sup> in that ERG TFs are consistently induced across nearly all cell types, with LRGs exhibiting more cell-type-specific induction (Mann–Whitney  $P \leq 0.0055$ ; Fig. 3a). *FOS* was induced in all clusters, whereas *BDNF*, which is known to be inducible in excitatory glutamatergic neurons but not GNs in rodent models, was selectively induced in hGN pallial neuron cluster 6 (Extended Data Fig. 4b). In total, this analysis identified 1,810 genes whose expression is induced in a cluster-specific manner within hGN cultures (Supplementary Data 7).

Although scRNAseq detected the expression of some messenger RNAs (mRNAs) in too few cells to enable activity-dependent differential gene expression analysis (*XIRP1*), many genes were identified as having activity-inducible expression by both total and scRNAseq, allowing for comparisons between the two sequencing modalities (Fig. 3b and Supplementary Data 8). For example, scRNAseq revealed that *MMP1* induction is unique to striatal neuron cluster 9 of hGNs, whereas *LINC00473* and *ZNF331* are induced in most neuronal clusters but not mitotic progenitor clusters 8 and 12 (Extended Data Fig. 4b). In addition, scRNAseq analysis uncovered depolarization-responsive genes not detected as inducible by total RNAseq, including the previously reported inducible genes *SST* and *NPTX2*, whose expression was restricted to a subset of neuronal hGN clusters (Fig. 3b, Extended Data Fig. 4b and Supplementary Data 8).

**Membrane depolarization-inducible expression of disease-associated genes.** Previous studies indicate that mutations within

genes that encode voltage-gated calcium channel subunits, calcium-dependent signaling molecules and calcium-dependent TFs can lead to severe monogenic neurodevelopmental disorders<sup>28,29</sup>. Moreover, genome-wide association studies (GWASs) suggest that sequence variation in these same gene loci might contribute to ASD, schizophrenia (SCZ) and bipolar (BP) disorder<sup>30–32</sup>. We asked whether the set of hGN activity-inducible genes detected by scRNAseq is enriched for such disease-associated genes, compared to all hGN-expressed genes. We found that Simons Foundation Autism Research Initiative (SFARI) ASD-associated genes, but not SCZ-associated genes, are significantly enriched ( $P = 1.627 \times 10^{-5}$ ) within the hGN scRNAseq inducible gene set (Supplementary Data 9). Nevertheless, many SCZ-associated genes were also found to be regulated by membrane depolarization in hGNs, a finding that might inform future investigations of SCZ disease etiology (Supplementary Data 10).

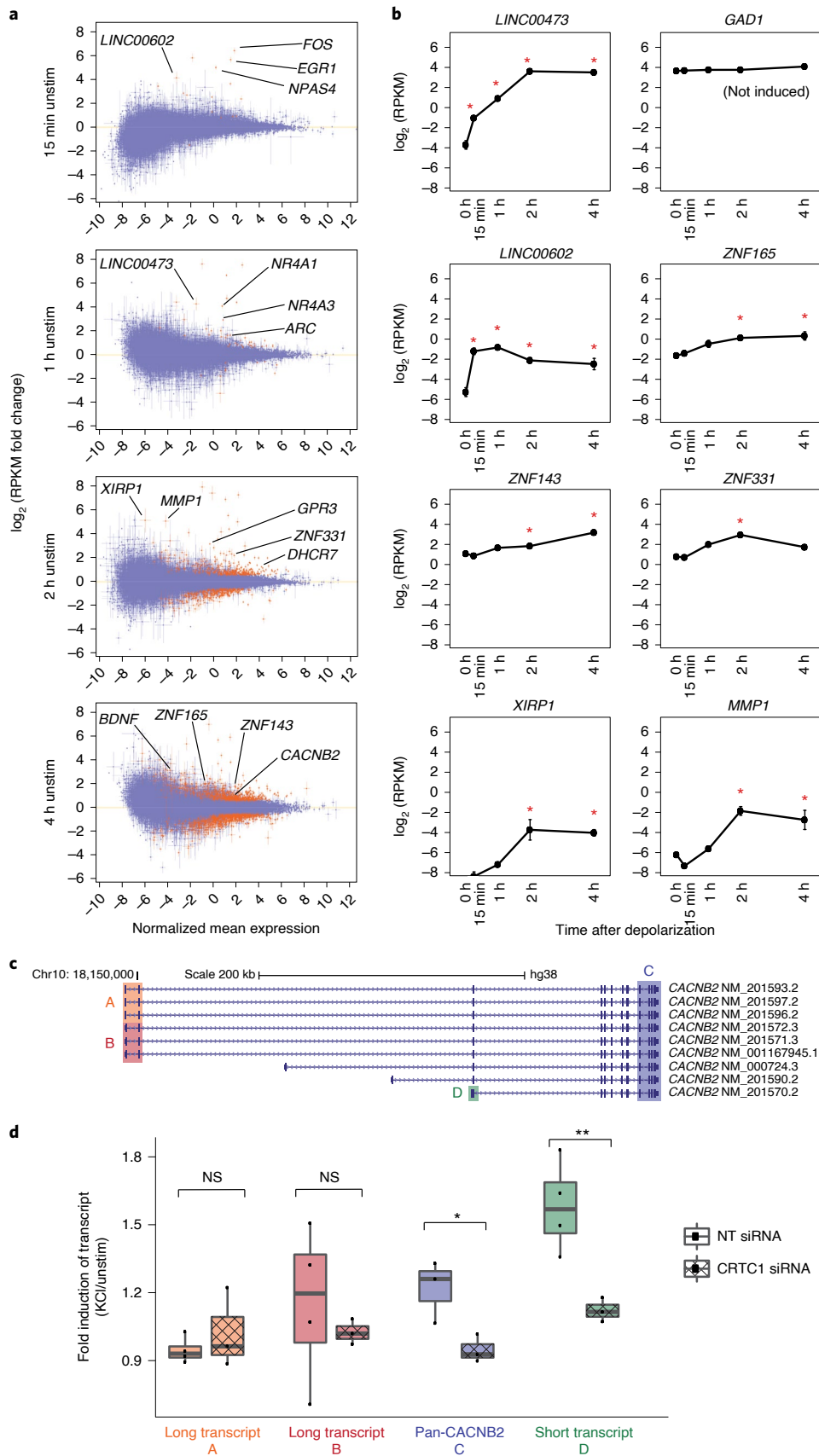
To examine more closely how disease-associated genes are activity regulated within hGN cultures, we looked at the induction patterns of these genes across the scRNAseq clusters. For example, salt-inducible kinase 1 (*SIK1*), an ASD- and epilepsy-associated gene previously reported to be induced by neuronal activity, was found to be activity inducible in nearly all post-mitotic clusters (Fig. 3c). However, several other ASD-associated genes displayed inducible expression in only a subset of hGN clusters. These include genes such as 7-dehydrocholesterol reductase (*DHCR7*) in many neuronal clusters, including all striatal neuron-like clusters; calcium and integrin binding family member 2 (*CIB2*) in only striatal neuron-like cluster 10; and SH3 and multiple ankyrin repeat domains 3 (*SHANK3*) in only striatal neuron-like cluster 9 (Fig. 3c). In addition to being associated with ASD, *DHCR7* and *CIB2* are mutated in Smith–Lemli–Opitz syndrome and Usher syndrome, respectively<sup>33,34</sup>. *CIB2* is likely involved in intracellular calcium regulation and homeostasis, and its mutation causes loss of neuroblasts in zebrafish and photoreceptor cells in *Drosophila*<sup>34</sup>. *DHCR7* reduces 7-dehydrocholesterol to cholesterol, and the accumulation of 7-dehydrocholesterol prevents formation of the Wnt receptor complex, resulting in precocious neural differentiation of hiPSCs<sup>33</sup>. *SHANK3* mutations are a highly penetrant monogenic risk factor for ASD and cause Phelan–McDermid syndrome. *SHANK3* is important for synaptic scaffolding, regulating spine morphology and neurotransmission, and its disruption leads to ASD-like phenotypes in both mouse and monkey<sup>35</sup>. Future examination of the transcriptional dynamics of such genes in vivo and in disease models will be needed to determine whether their induction in developing striatal neurons affects the excitatory–inhibitory balance of circuits involved in ASD pathology<sup>36</sup>.

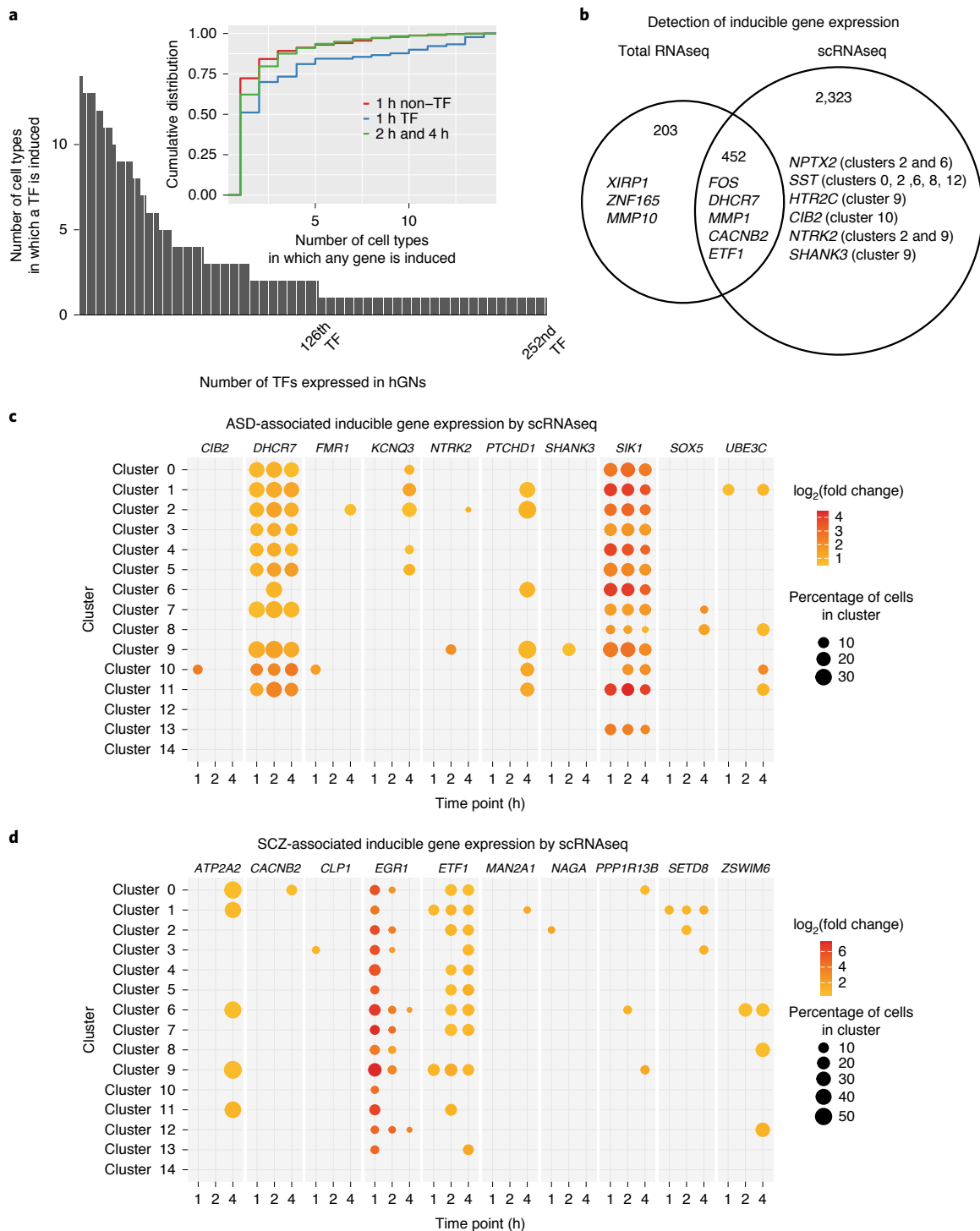
Several activity-regulated genes detected in hGN cultures also mapped to known SCZ-associated loci (Fig. 3d)<sup>30</sup>. Early growth response 1 (*EGR1* or *Zif268*), a SCZ-associated gene known to display neuronal activity-dependent expression, was induced by membrane depolarization in nearly all hGN cell clusters (Fig. 3d). In contrast, the SCZ-associated gene zinc finger SWIM domain-containing protein 6 (*ZSWIM6*) was specifically induced in pallial neuronal cluster

**Fig. 2 | Neuronal activity-dependent gene expression of hGNs by total RNAseq.** mRNA expression level changes measured by total RNAseq of 1G hGNs at (a) 15 min ( $n = 3$ ), 1 h ( $n = 3$ ), 2 h ( $n = 6$ ) and 4 h ( $n = 3$ ) after membrane depolarization, compared to expression in unstimulated cultures ( $n = 6$ ) represented by MA-plot. Genes with a significantly different gene expression level and a minimum fold-change magnitude of 1.5 after depolarization are marked in red. Genes having inducible expression fall above  $y = 0$ , and example gene names are labeled. b, Time courses of gene expression levels measured by total RNAseq for genes not previously known to be stimulation inducible in the brain, *LINC00473* and *ZNF331* (positive controls), and *GAD1* (negative control). Time points at which the transcript was significantly induced compared to unstimulated cultures are marked with an asterisk. ( $n = 6$  (0 h and 2 h) or  $n = 3$  (15 min, 1 h and 4 h) biologically independent samples; plots display mean  $\pm$  s.e.m.) c, Diagram of *CACNB2* transcript variants and exons recognized by TaqMan probes used to query transcript-specific activity-dependent expression. d, Box and whisker plot displaying fold changes in mRNA levels of different *CACNB2* transcript variants after membrane depolarization of hGNs, as measured by TaqMan qPCR ( $n = 3–4$ ; one-sided  $t$ -test:  $*P = 0.0145$  and  $**P = 0.0071$ ). *CACNB2* induction is due to increased levels of 'short transcript D' (NM\_201570.2) in response to membrane depolarization and is dependent on the transcription factor CRTCl. NS, not significantly different.

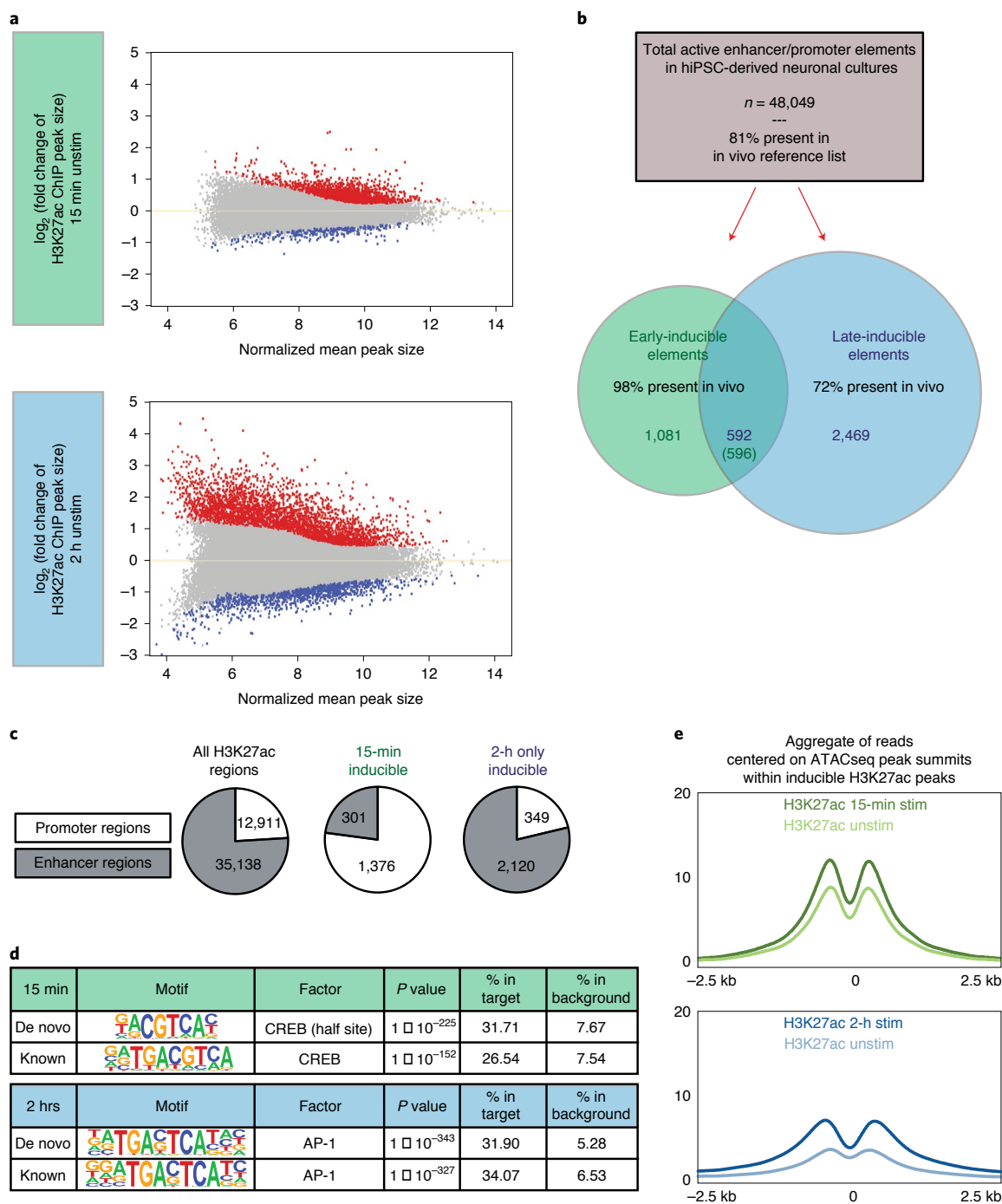
6 and mitotic progenitor clusters 8 and 12 (Fig. 3d). Yet another SCZ-associated gene encoding the calcium voltage-gated channel auxiliary subunit beta 2 (*CACNB2*) showed inducible expression

specifically in striatal cluster 0 (Fig. 3d). *CACNB2* encodes a cytoplasmic channel subunit that enhances  $Ca_v1$  and  $Ca_v2$   $Ca^{2+}$  channel currents by changing the voltage dependence and kinetics of





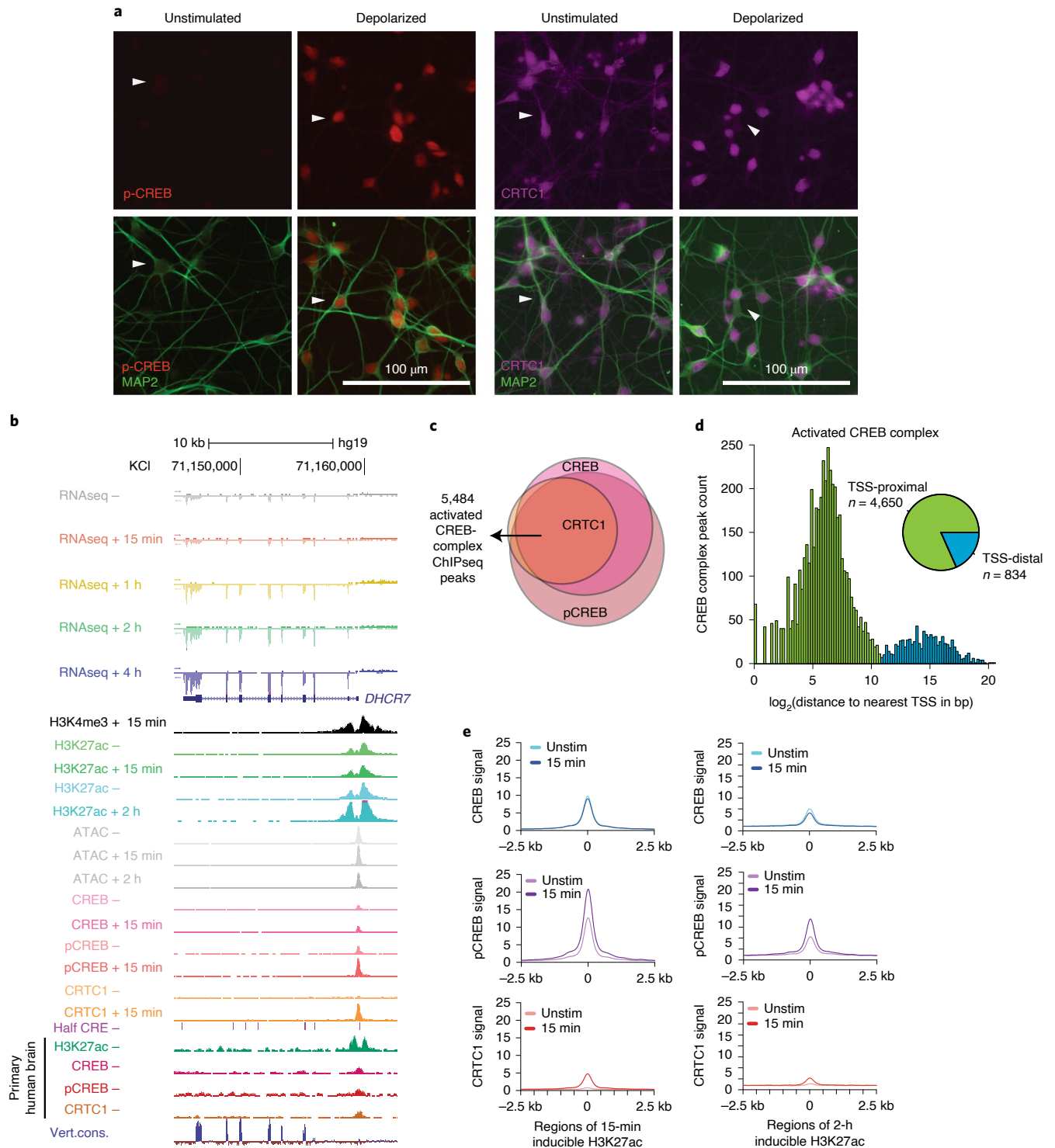
**Fig. 3 | Neuronal activity-dependent gene expression of hGNs by scRNAseq.** **a**, scRNAseq of 1G hGNs reveals ubiquitous and cell-type-specific TF gene induction within hGN cultures. Early-inducible (1 h) TF genes are induced in more cell types than late-inducible (2 h and 4 h) TF genes. **b**, The genes detected to have inducible expression are partially overlapping between total and scRNAseq methods. Total RNAseq can detect induction of lowly expressed genes (for example, *XIRP1*), whereas scRNAseq can detect induction of genes inducible in only a subset of hGN cell types (for example, *NPTX2*). For total RNAseq,  $n = 6$  (0 h and 2 h) or  $n = 3$  (15 min, 1 h and 4 h) biologically independent samples. **c**, Selected SFARI ASD-associated genes having inducible expression in hGNs, as measured by scRNAseq, including genes with induction across many hGN cell clusters (for example, *DHCR7*) and genes showing hGN cell-cluster-restricted induction (for example, *SHANK3*). **d**, Selected SCZ-associated genes having inducible expression in hGNs. The magnitude of fold change of mRNA levels in each cell cluster at different time points after membrane depolarization compared to unstimulated mRNA levels is indicated by dot color. Only time points with significantly induced mRNA levels (compared to unstimulated) are shown. The percentage of cells in each cluster from which expression was detected is represented by the size of the dot. For scRNAseq,  $n = 37,101$  cells examined over three independent experiments using independent differentiation lots of 1G neurons.



**Fig. 4 | hGN activity-dependent promoters and enhancers.** **a**, 1G H3K27ac ChIPseq peak size changes in 15-min or 2-h membrane-depolarized hGNs compared to unstimulated cultures represented by MA-plot. Peak regions with a significantly greater read count after membrane depolarization (inducible) are in red; those with a significantly lower read count are in blue. **b**, Summary of 1G H3K27ac ChIPseq regions, their inducibility and representation in the IVRL. Most inducible H3K27ac ChIPseq peaks are significantly increased at either 15 min (early) or 2 h (late) after membrane depolarization but not both. **c**, Compared to all 1G H3K27ac ChIPseq regions, 1G early-inducible regions are enriched for those that overlap with a TSS (promoter regions), and 1G late-inducible regions are enriched for those that do not overlap a TSS (enhancer regions). **d**, The CREB-binding sequence is the most enriched sequence motif within 15-min inducible H3K27ac regions, and the AP-1 motif is the most enriched within 2-h inducible H3K27ac regions. **e**, Aggregate plots of 1G H3K27ac ChIPseq reads within inducible H3K27ac ChIPseq regions before (unstimulated) and after 15 min or 2 h of membrane depolarization.

channel activation or inactivation<sup>37</sup>. Transcript-specific qPCR revealed that this induction is specific to the *CACNB2* transcript variant  $\beta$ -2e<sup>37</sup> (NM\_201570.2) and is dependent on the calcium-responsive CREB-regulated transcription coactivator

1 (CRTC1 or TORC1) (Fig. 2c,d). Further investigation of the cell-type-specific regulation of these disease-associated genes should provide insight into how different GN subtypes in the fore-brain contribute to neuropsychiatric disease risk.



**Fig. 5 | Genome-wide binding of the activity-dependent CREB complex.** **a**, Immunostaining of 1G hGN cultures for pCREB (red), CRTC1 (purple) and MAP2 (green) before and 15 min after membrane depolarization. After depolarization, nuclear pCREB and CRTC1 increase. Images are representative of three independent experiments. **b**, Genome browser tracks of all 1G total RNAseq, ChIPseq and ATACseq data generated at the ASD-associated inducible gene locus *DHCR7* (pooled and normalized reads from 2-3 biological replicates of ChIPseq and ATACseq, or 3-6 biological replicates of total RNAseq, for each condition). *DHCR7* is encoded on the negative strand of chromosome 11; therefore, transcript reads accumulate downward. CREB complex is bound at the *DHCR7* promoter in mid-gestational primary human fetal cortex. The 'half CRE' track indicates locations of the minimal CREB-binding motif in this region. The 'vert.cons.' track represents sequence conservation across 100 vertebrate species; sites predicted to be conserved are assigned positive scores (blue), whereas sites predicted to be fast-evolving are assigned negative scores (red). **c**, Overlap of CREB, pCREB and CRTC1 ChIPseq peaks used to define the CREB complex-bound regions. Diagram circles and overlap are proportional to the number of peaks for each TF ChIPseq. **d**, Genomic distribution plot of CREB complex ChIPseq peaks reveals that 85% of CREB complex peak summits are proximal (<1.7 kb; green) to a gene's TSS. **e**, Aggregate plots of CREB, pCREB and CRTC1 ChIPseq reads within 15-min inducible H3K27ac regions and 2-h inducible H3K27ac regions with and without stimulation. Both pCREB and CRTC1 enrichment increase after hGN membrane depolarization, and greater enrichment of CREB complex is seen in early-inducible H3K27ac regions than in late-inducible H3K27ac regions.



**Neuronal depolarization-responsive enhancers and promoters.**

To understand how non-coding disease-associated genomic variation affects gene expression in human neurons, we identified the gene regulatory elements in hGN cultures that undergo dynamic activation in neurons in response to acute membrane depolarization. We performed chromatin immunoprecipitation followed by sequencing (ChIPseq) to detect sites of enrichment of two histone modifications: acetylated lysine 27 of histone H3 (H3K27ac) and tri-methylated lysine 4 of histone H3 (H3K4me3). H3K4me3 is highly enriched at gene promoters, whereas H3K27ac marks active promoters as well as distal enhancer elements<sup>38</sup> and has been shown to rapidly accumulate at stimulus-responsive regulatory elements in neurons after membrane depolarization<sup>39</sup>.

We performed H3K27ac ChIPseq on multiple differentiation lots of a single hGN genotype (1G; Fig. 4) as well as across the four hGN cultures derived from distinct hiPSC genotypes (4G; Extended Data Fig. 5). In each case, ChIPseq ( $n=2-3$ ) was conducted on unstimulated cultures as well as 15 min and 2 h after KCl stimulation, when chromatin changes are underway that induce ERG and LRG expression, respectively. hGN H3K27ac-enriched regulatory regions were categorized as unchanged (constitutive), inducible (increasing) or decreasing in H3K27ac at each of the post-stimulation time points (Fig. 4a,e and Supplementary Data 11 and 12). We found a total of 72,724 H3K27ac-enriched, non-overlapping active promoter and enhancer regions in hGNs across all time points and datasets, with 48,049 and 80,638 from the 1G and 4G datasets, respectively. Of these, a subset of sequences significantly increased in H3K27ac within 15 min after membrane depolarization compared to the unstimulated cultures (1G=1,677 and 4G=5,622), and an additional set (1G=2,469 and 4G=2,401) showed increased in H3K27ac by 2 h (1G; Fig. 4b; 4G; Extended Data Fig. 5a,b). In total, across both datasets, 9,195 regulatory regions were found to significantly increase in H3K27ac after membrane depolarization.

To examine inter-donor variability in the data, we correlated fold changes in H3K27ac among each of the three independent genotypes and the fourth, 1G genotype at both stimulation time points, yielding Pearson's correlation coefficients of 0.74–0.81 (Extended Data Fig. 5f). Hierarchical clustering of fold changes in H3K27ac at these genomic regions revealed subsets with variability in induction levels across cells lines (Extended Data Fig. 5g). This variability might result from a combination of several factors, including genetic variation among individuals, technical variability among different hiPSC reprogramming methods and mutations that might have been enriched during primary donor cell outgrowth and reprogramming (Supplementary Table 1). Variability in H3K27ac fold changes across samples was greater at the 2-h time point than at the 15-min time point, consistent with late-wave inducible transcriptional programs being more cell type specific than early-wave programs (Fig. 3a)<sup>18</sup>. However, despite these sources of variation, most peaks of H3K27ac enrichment showed similar trends among all four donors.

We confirmed the ability of a subset of the identified activity-responsive gene regulatory elements to drive inducible reporter gene expression in response to membrane depolarization in mouse cortical cultures (Extended Data Fig. 5e). Consistent with their predicted function, we found that H3K27ac promoter regions were significantly more likely to be associated with inducible gene loci than constitutively active promoters (1G and 4G,  $P < 2.2 \times 10^{-16}$ ). To investigate the in vivo relevance of hGN H3K27ac regulatory elements, we compiled a merged list of H3K27ac-enriched genomic sites previously reported in three studies using human developing and adult brain tissues to create an in vivo reference list (IVRL) of human brain H3K27ac regions (Methods). Most active hGN H3K27ac regions (81% of 1G and 76% of 4G) were found to overlap with previously reported H3K27ac-enriched sites in vivo (Fig. 4b and Extended Data Fig. 5b), strongly suggesting that these

elements represent functional regulatory regions in the human brain. However, late-response elements had a lower (1G=72% and 4G=75%) representation in the IVRL compared to early-response elements (1G=98% and 4G=99%), likely reflecting the difficulty of detecting cell-type-restricted enhancers in the context of highly heterogeneous brain tissue.

Most inducible regulatory elements in our datasets showed an increase in H3K27ac at either 15 min or 2 h after stimulation but not both (Fig. 4b and Extended Data Fig. 5b). When comparing early- and late-inducible H3K27ac regions, we noted that early-inducible (15 min) regions were enriched for promoter elements (1G and 4G,  $P < 2.2 \times 10^{-16}$ ; Fig. 4c and Extended Data Fig. 5c), whereas late-inducible regions were enriched for enhancer elements (1G,  $P < 2.2 \times 10^{-16}$ ; Fig. 4c). To investigate the molecular mechanisms underlying these regulatory differences, we performed assay for transposase-accessible chromatin using sequencing (ATACseq) at these same time points after stimulation to narrow down the regions within the regulatory elements that are likely to bind TFs (1G: 81,240 ATACseq peaks and 4G: 115,081 ATACseq peaks) and then carried out TF-binding motif enrichment analysis on the 500-bp regions centered on the ATACseq peak summits. This analysis identified the full (TGACGTC) and half (TGACG) canonical cAMP/ $Ca^{2+}$ -responsive element (CRE) motifs, recognized by the TF CREB and its CRE modulator protein family members, as the top enriched motifs within early-inducible elements (Fig. 4d and Extended Data Fig. 5d). By contrast, the AP-1 motif (TGASTCA), recognized by heterodimers of the FOS and JUN transcription factor family proteins, was the top enriched motif within 2-h inducible H3K27ac elements (Fig. 4d and Extended Data Fig. 5d). In addition to these motif analyses, our parallel epigenomic datasets allowed us to employ a recently reported activity-by-contact (ABC) model to predict novel enhancer–gene interactions<sup>40</sup>. Using ATACseq and H3K27ac ChIPseq data across the three time points, the ABC model returned over 32,000 predicted constitutive and depolarization-induced regulatory interactions that provide an expanded view of hGN enhancer–gene regulation (Extended Data Fig. 8 and Supplementary Data 13).

**Genome-wide binding profile of the activated CREB complex in hGNs.**

Given CREB's important roles in calcium-dependent gene regulation, neuronal survival, brain development and plasticity<sup>41–43</sup>, it is a good candidate to be a central mediator of early-response promoter remodeling in response to membrane depolarization. To begin to characterize CREB activation patterns in hGNs, we immunostained hGN cultures using anti-CREB antibodies as well as antisera specific for the activated form of CREB phosphorylated at serine133 (pCREB)<sup>44</sup>. In addition, we also used immunostaining to examine the subcellular localization of CRTCI, an evolutionarily conserved co-factor for CREB-dependent transcription that has been reported to undergo regulated translocation into the nucleus in response to neuronal activity<sup>45</sup>. Together, these studies confirmed CREB's expected nuclear localization in hGN cultures and demonstrated increased CREB phosphorylation and CRTCI nuclear translocation within 15 min of depolarization (Fig. 5a and Extended Data Fig. 6).

We performed ChIPseq for CREB, pCREB and CRTCI using 1G hGNs before and 15 min after membrane depolarization ( $n=2-3$ ), as well as from mid-gestational human fetal cortex to serve as an in vivo comparison (Fig. 5b). CREB was typically observed pre-bound before stimulation, with increased phosphorylation of CREB Ser133 and the association of CRTCI observed upon membrane depolarization. This analysis identified 5,484 high-confidence CREB complex binding sites, defined as sites associated with CREB, pCREB and CRTCI after membrane depolarization, in hGNs (Fig. 5c). Virtually all (99%) of these CREB complex sites overlapped with active H3K27ac-enriched regulatory elements, with 95% associ-

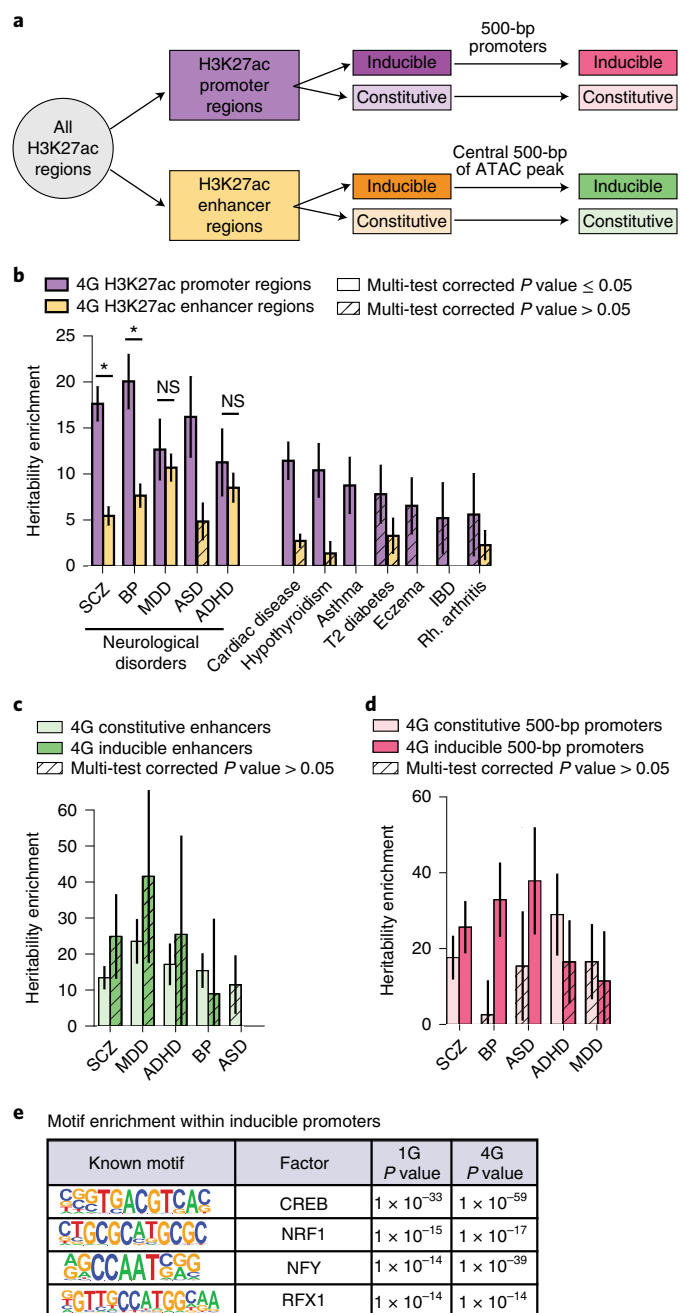
ated with a full or half canonical CRE motif sequence. Moreover, in agreement with the traditional view of CREB as a promoter-bound TF<sup>41</sup>, most (85%) CREB complex sites were located proximal to (<1.7 kb) a gene TSS (Fig. 5d).

Consistent with our previous TF-binding motif enrichment analysis, activated CREB complex was present at most (86.5%) early-inducible H3K27ac regions and at only a minority (22.3%) of late-inducible H3K27ac regions (Fig. 5e). Notably, 88% of early-inducible H3K27ac regions with a CRE were also associated with CREB complex as opposed to 11% at late-inducible CRE-containing regions, indicating that this binding enrichment ( $P < 2.2 \times 10^{-16}$ ) does not simply reflect enrichment of CRE sequence motifs in early-inducible regulatory elements. Indeed, the presence of a CRE-binding motif within an active regulatory element (36,559 H3K27ac regions with a full or half CRE) was far from predictive of CREB protein association because only 15% of these regions (5,331) were bound by the activated CREB complex and only 20% (7,244) by CREB/pCREB. These studies thus suggest that additional, as yet uncharacterized, genomic factors bias CREB binding to promoter regions and underscore the need for chromatin occupancy studies to identify bona fide CREB-binding sites in a given cell type.

**Neurological disease heritability enrichment in hGN activity-inducible regulatory regions.** GWASs have uncovered hundreds of loci associated with neuropsychiatric diseases. However, identifying causal variation within non-coding regions requires a better understanding of tissue-specific regulatory structure. Having defined active and stimulus-responsive promoters and enhancers and CREB complex binding sites in hGNs, we asked whether these regions were enriched for sequence variation associated with neuropsychiatric disease by applying stratified linkage disequilibrium score regression (LDSC) (Methods), a statistical model that is widely used to identify categories of genomic regions that have significantly more trait heritability than expected by chance. This model estimates disease heritability enrichment in genomic regions of interest based on the linkage disequilibrium score among single-nucleotide polymorphisms (SNPs), whose heritability effect sizes were inferred from previously reported GWASs. LDSC also controls for variation in genetic features, such as sequence conservation, as well as variation in epigenetic features shared across human tissues, such as DNA methylation and histone modification enrichment. In addition, we included previously annotated regions of H3K27ac enrichment from human brain, annotated brain regulatory elements from ENCODE and brain regulatory elements

as annotated by ChromHMM from the Roadmap Epigenomics Mapping Consortium as control categories to further account for the brain-like chromatin states that might be expected to be over-represented in any human neuronal culture system (Methods).

Subsets of hGN promoters and enhancers from both 1G and 4G datasets were tested for heritability enrichment across a panel of neurological conditions (SCZ, BP disorder, major depressive disorder (MDD), ASD and attention deficit hyperactivity disorder (ADHD)) and non-neurological conditions (type 2 diabetes, cardiac disease, hypothyroidism, asthma, eczema, irritable bowel disease and rheumatoid arthritis) for which large GWASs have been performed (Fig. 6a, Supplementary Table 3 and Extended Data Fig. 7). First, we analyzed the disease heritability enrichment of hGN H3K27ac-defined promoter regions and enhancer regions (Fig. 6b, Extended Data Fig. 7a,b and Supplementary Data 14). We found that hGN enhancer regions from both 1G and 4G H3K27ac ChIPseq experiments were



**Fig. 6 | Disease heritability enrichment in hGN promoter and enhancer regions.** **a**, Diagram summarizing the relationship between hGN H3K27ac ChIPseq peak regions and the subregions investigated for disease heritability enrichment. **b**, Heritability enrichment of 4G hGN H3K27ac promoter regions and enhancer regions across neurological and non-neurological diseases. **c**, Heritability enrichment of the 4G-inducible and constitutive 500-bp enhancers across neurological diseases, showing only constitutive enhancers having significant heritability enrichment. Pairwise two-sided  $t$ -test,  $*P < 0.05$ . **d**, Heritability enrichment of the 4G-inducible and constitutive 500-bp promoters across neurological diseases, showing significant enrichment of ASD and BP disorder heritability in inducible promoters but not constitutive promoters. SCZ heritability is similarly enriched in inducible and constitutive promoters. Pairwise two-sided  $t$ -test,  $*P < 0.05$ . **e**, Known sequence motifs enriched within inducible 500-bp promoters compared to constitutive 500-bp promoters from both 1G and 4G datasets. Each heritability enrichment value is provided in the bar plot ( $\pm$  s.e.). NS, not significantly different. Heritability enrichments with multitest corrected  $P$  values  $\geq 0.05$  are cross-hatched. Exact  $P$  values are provided in Supplementary Data 14. IBD, irritable bowel disease.

enriched for disease heritability across most neurological, but not non-neurological, disorders, consistent with enhancers representing the primary drivers of cell-type-specific gene expression<sup>46</sup>. By contrast, promoter regions were collectively enriched for heritability for some non-neurological disorders and were enriched in heritability for all tested neurological disorders, likely reflecting overlap among genes that are essential for healthy brain function and the health of other organs.

Using LSDC, we identified heritability enrichments of constitutive H3K27ac enhancer regions for all five neurological disorders (SCZ, BP disorder, MDD, ADHD and ASD), whereas the enrichment in inducible H3K27ac enhancer regions was not statistically significant (Extended Data Fig. 7c). Significant enrichments were also found for the central 500 bp of the ATACseq regions, presumed sites of TF binding, within the constitutive H3K27ac enhancer regions (constitutive enhancers) from both 1G and 4G experiments but not those of the inducible enhancers (Fig. 6c and Extended Data Fig. 7d). These analyses indicate that the neurological disorder heritability enrichment of hGN enhancers is limited to constitutive enhancers.

Further examining disease heritability enrichment within hGN promoter regions, we found that H3K27ac hGN promoter regions were significantly enriched for heritability across neurological disorders and had a significantly greater magnitude of heritability enrichment than enhancer regions for SCZ (1G,  $P=1.1 \times 10^{-8}$ ; 4G,  $P=2.1 \times 10^{-8}$ ), BP disorder (1G,  $P=2.9 \times 10^{-7}$ ; 4G,  $P=1.2 \times 10^{-4}$ ) and ASD (1G,  $P=0.046$ ; 4G,  $P=0.017$ ) (Fig. 6b, Extended Data Fig. 7a and Supplementary Data 14). To more carefully exclude adjacent non-promoter sequences within H3K27ac-enriched regions, we designated the 500-bp sequence upstream of every annotated TSS in the human genome as a promoter and tested the heritability enrichment of just the promoter within inducible and constitutive H3K27ac regions. Inducible promoters showed significant enrichment for ASD heritability in both 1G and 4G (1G,  $P=0.046$ ; 4G,  $P=0.017$ ) and for BP disorder and SCZ in 4G ( $P=6.0 \times 10^{-3}$  and  $P=2.0 \times 10^{-3}$ , respectively) (Fig. 6d and Extended Data Fig. 7e). Notably, we also observed an increase in ASD heritability enrichment magnitude from 13.75 (1G) and 18.26 (4G) to 66.00 (1G) and 37.87 (4G) (Fig. 6d and Extended Data Fig. 7e,f), despite having dramatically reduced the size of the genomic test regions. By contrast, constitutive promoters carried significant enrichment only for SCZ and ADHD heritability but not ASD or BP disorder (Fig. 6d and Extended Data Fig. 7e). Together, these results suggest that activity-inducible promoters in developing forebrain GNs are important components of ASD and BP disorder genetic architecture and might also contribute to the genetic architecture of SCZ.

To identify TFs that might regulate hGN-inducible promoters enriched for ASD and BP disorder GWAS heritability, we examined sequence motifs enriched within inducible promoters as compared to all constitutive promoters. Similarly to our previous analysis of 15-min inducible H3K27ac regions (Fig. 4d), the CREB-binding CRE motif was identified as the most enriched motif within all inducible 500-bp promoters (Fig. 6e), and CREB complex ChIPseq data confirmed that CREB binding was significantly enriched at 1G-inducible promoters compared to constitutive promoters ( $P<2.2 \times 10^{-16}$ ). In addition, the binding motif for NRF-1, a TF reported to contribute to calcium-dependent gene expression in muscle cells<sup>47</sup>, was also significantly enriched in 1G-inducible promoter elements. Together, these findings suggest that CREB, along with additional TFs such as NRF-1, might co-regulate an important set of calcium-responsive genes in developing human GNs, and that disruption of this regulatory network might contribute to ASD and/or BP disorder risk.

Finally, having established that depolarization-inducible hGN promoters are enriched for ASD risk, we returned to our transcriptional data to better understand the relationship between

hGN-inducible genes and ASD. We asked whether inducible genes specific to any individual scRNAseq cell cluster were enriched for SFARI ASD-associated genes. Progenitor cluster 8 and striatal pre-SPN cluster 10 inducible genes were significantly enriched for SFARI ASD-associated genes compared to all expressed genes within those clusters (8:  $P=9.27 \times 10^{-4}$ ; 10:  $P=4.11 \times 10^{-4}$ ). In addition, by scRNAseq, pre-SPN cluster 10 exhibited the highest proportion of its inducible genes to also have inducible promoters, compared to other clusters (Extended Data Fig. 7g). These findings suggest that ASD-associated sequence variation within inducible promoters and their cognate genes might affect ventral forebrain neuronal progenitor and developing SPN function.

## Discussion

Activity-dependent gene programs regulate nervous system development and function. However, deep characterization of these gene expression programs in human brain has proven difficult because the brain is a heterogeneous tissue, and the activity-dependent gene programs diverge significantly among neuronal subtypes. In this study, we comprehensively profiled depolarization-responsive changes in gene expression and gene regulatory element activity in an hiPSC-derived culture model of human forebrain GNs, identifying many loci not previously known to be regulated by neuronal activity.

We employed this resource to gain a deeper understanding of neuropsychiatric disease etiology. Using scRNAseq, we show that the ASD-associated genes *DHCR7* and *SHANK3*, as well as the SCZ-associated genes *CACNB2* and *ZSWIM6*, undergo inducible gene regulation in hGNs, suggesting that these disease genes might be subject to previously unappreciated activity-dependent regulation in human GNs. More broadly, we found that hGN neural progenitor cells and pre-SPN neurons induce a disproportionately large number of ASD-associated genes, suggesting that cells of the developing striatum might be vulnerable to the effects of ASD-associated genetic variation. Our results, thus, corroborate previous indications that immature GNs of the forebrain are a relevant cellular context for ASD-associated variations in gene expression<sup>48,49</sup> and specifically implicate activity-dependent gene regulation in this process.

Extending this analysis to activity-responsive gene regulatory elements, we found that inducible promoters, but not inducible enhancers, in hGN cultures carry a significant enrichment for BP disorder heritability as well as a strikingly large ASD heritability enrichment. In contrast, no such distinction could be made between inducible and constitutive promoters for SCZ or other neurological disorders. These findings suggest that the subset of gene promoters in hGNs that undergo significant increases in H3K27ac in response to neuronal depolarization contribute to ASD risk. Our results not only add to converging evidence that promoters are critical regions driving ASD risk<sup>50</sup> but also suggest that activity-induced promoters within hGNs might contribute significantly to this risk.

Additionally, binding motifs for CREB, NRF1, NFY and RFX1 family TFs were found to be enriched within activity-inducible promoters harboring elevated ASD heritability. Although, to date, we have only confirmed direct binding of the activated CREB complex at these promoters using ChIPseq, interactions between CREB and NRF1 proteins have previously been reported at stimulus-responsive loci<sup>47</sup>. Our findings raise the possibility that CREB and NRF1 binding at activity-dependent promoters might be disrupted in ASD, and that functional investigation of sequence variation affecting these and other critical transducers of neuronal calcium signaling should be prioritized.

Our finding that depolarization-induced genes of pre-SPN-like cells are enriched for ASD-associated genes, taken together with the neurological disease heritability enrichment observed in these activity-responsive gene regulatory elements, suggests a possible role for dysregulated calcium-dependent gene expression

within developing striatal neurons in ASD etiology. Moreover, our assembled data will enable improved fine mapping of neurological disease-associated sequence variants within neuronal activity-dependent regulatory elements and lay the foundation for future comparisons between the activity-dependent regulatory landscapes of different human neuronal subtypes, so as to uncover new aspects of human gene regulation during brain development and disease.

### Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41593-020-00786-1>.

Received: 22 July 2019; Accepted: 18 December 2020;

Published online: 04 February 2021

### References

- Geschwind, D. H. & Rakic, P. Cortical evolution: judge the brain by its cover. *Neuron* **80**, 633–647 (2013).
- Sousa, A. M. M., Meyer, K. A., Santpere, G., Gulden, F. O. & Sestan, N. Evolution of the human nervous system function, structure, and development. *Cell* **170**, 226–247 (2017).
- Lui, J. H., Hansen, D. V. & Kriegstein, A. R. Development and evolution of the human neocortex. *Cell* **146**, 18–36 (2011).
- Petanjek, Z. et al. Extraordinary neoteny of synaptic spines in the human prefrontal cortex. *Proc. Natl Acad. Sci. USA* **108**, 13281–13286 (2011).
- Hensch, T. K. Critical period regulation. *Annu. Rev. Neurosci.* **27**, 549–579 (2004).
- Ataman, B. et al. Evolution of Osteocrin as an activity-regulated factor in the primate brain. *Nature* **539**, 242 (2016).
- Lim, L., Mi, D., Llorca, A. & Marín, O. Development and functional diversification of cortical interneurons. *Neuron* **100**, 294–313 (2018).
- Medina, L., Abellán, A., Vicario, A. & Desfilis, E. Evolutionary and developmental contributions for understanding the organization of the basal ganglia. *Brain Behav. Evol.* **83**, 112–125 (2014).
- Shepherd, G. M. Corticostriatal connectivity and its role in disease. *Nat. Rev. Neurosci.* **14**, 278 (2013).
- Haythornthwaite, A. et al. Characterizing human ion channels in induced pluripotent stem cell-derived neurons. *J. Biomol. Screen.* **17**, 1264–1272 (2012).
- Berry, B. J. et al. Morphological and functional characterization of human induced pluripotent stem cell-derived neurons (iCell Neurons) in defined culture systems. *Biotechnol. Prog.* **31**, 1613–1622 (2015).
- Hochbaum, D. R. et al. All-optical electrophysiology in mammalian neurons using engineered microbial rhodopsins. *Nat. Methods* **11**, 825–833 (2014).
- Onorati, M. et al. Molecular and functional definition of the developing human striatum. *Nat. Neurosci.* **17**, 1804–1815 (2014).
- Mayer, C. et al. Developmental diversification of cortical inhibitory interneurons. *Nature* **555**, 457–462 (2018).
- Allaway, K. C. & Machold, R. Developmental specification of forebrain cholinergic neurons. *Dev. Biol.* **421**, 1–7 (2017).
- Wei, B. et al. The onion skin-like organization of the septum arises from multiple embryonic origins to form multiple adult neuronal fates. *Neuroscience* **222**, 110–123 (2012).
- Spiegel, I. et al. Npas4 regulates excitatory-inhibitory balance within neural circuits through cell-type-specific gene programs. *Cell* **157**, 1216–1229 (2014).
- Hrvatin, S. et al. Single-cell analysis of experience-dependent transcriptomic states in the mouse visual cortex. *Nat. Neurosci.* **21**, 120–129 (2018).
- Pruunsild, P., Bengtson, P. C. & Bading, H. Networks of cultured iPSC-derived neurons reveal the human synaptic activity-regulated adaptive gene program. *Cell Rep.* **18**, 122–135 (2017).
- Ingi, T. et al. Dynamic regulation of RGS2 suggests a novel mechanism in G-protein signaling and neuronal plasticity. *J. Neurosci.* **18**, 7178–7188 (1998).
- Nedivi, E., Hevroni, D., Naot, D., Israeli, D. & Citri, Y. Numerous candidate plasticity-related genes revealed by differential cDNA cloning. *Nature* **363**, 718–722 (1993).
- Fujioka, H., Dairyo, Y., Yasunaga, K.-I. & Emoto, K. Neural functions of matrix metalloproteinases: plasticity, neurogenesis, and disease. *Biochem. Res. Int.* **2012**, 789083 (2012).
- Zeisel, A. et al. Molecular architecture of the mouse nervous system. *Cell* **174**, 999–1014 (2018).
- Kang, H. et al. Spatio-temporal transcriptome of the human brain. *Nature* **478**, 483 (2011).
- Miller, J. A. et al. Transcriptional landscape of the prenatal human brain. *Nature* **508**, 199–206 (2014).
- Kapusta, A. et al. Transposable elements are major contributors to the origin, diversification, and regulation of vertebrate long noncoding RNAs. *PLoS Genet.* **9**, e1003470 (2013).
- Tadepally, H. D., Burger, G. & Aubry, M. Evolution of C2H2-zinc finger genes and subfamilies in mammals: species-specific duplication and loss of clusters, genes and effector domains. *BMC Evol. Biol.* **8**, 176 (2008).
- Ebert, D. H. & Greenberg, M. E. Activity-dependent neuronal signalling and autism spectrum disorder. *Nature* **493**, 327–337 (2013).
- Meur, L. N. et al. MEF2C haploinsufficiency caused by either microdeletion of the 5q14.3 region or mutation is responsible for severe mental retardation with stereotypic movements, epilepsy and/or cerebral malformations. *J. Med. Genet.* **47**, 22 (2010).
- Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427 (2014).
- Bipolar Disorder and Schizophrenia Working Group of the Psychiatric Genomics Consortium. Genomic dissection of bipolar disorder and schizophrenia, including 28 subphenotypes. *Cell* **173**, 1705–1715 (2018).
- Grove, J. et al. Identification of common genetic risk variants for autism spectrum disorder. *Nat. Genet.* **51**, 431–444 (2019).
- Francis, K. R. et al. Modeling Smith-Lemli-Opitz syndrome with induced pluripotent stem cells reveals a causal role for Wnt/ $\beta$ -catenin defects in neuronal cholesterol synthesis phenotypes. *Nat. Med.* **22**, 388–396 (2016).
- Riazuddin, S. et al. Alterations of the CIB2 calcium- and integrin-binding protein cause Usher syndrome type 1J and nonsyndromic deafness DFNB48. *Nat. Genet.* **44**, 1265–1271 (2012).
- Zhou, Y. et al. Atypical behaviour and connectivity in SHANK3-mutant macaques. *Nature* **570**, 326–331 (2019).
- Robertson, C. E. & Baron-Cohen, S. Sensory perception in autism. *Nat. Rev. Neurosci.* **18**, 671–684 (2017).
- Buraei, Z. & Yang, J. The  $\beta$  subunit of voltage-gated  $\text{Ca}^{2+}$  channels. *Physiol. Rev.* **90**, 1461–1506 (2010).
- Rada-Iglesias, A. et al. A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470**, 279 (2011).
- Malik, A. N. et al. Genome-wide identification and characterization of functional neuronal activity-dependent enhancers. *Nat. Neurosci.* **17**, nn.3808 (2014).
- Fulco, C. P. et al. Activity-by-contact model of enhancer–promoter regulation from thousands of CRISPR perturbations. *Nat. Genet.* **51**, 1664–1669 (2019).
- Lonze, B. E. & Ginty, D. D. Function and regulation of CREB family transcription factors in the nervous system. *Neuron* **35**, 605–623 (2002).
- Impey, S. et al. Defining the CREB regulon: a genome-wide analysis of transcription factor regulatory regions. *Cell* **119**, 1041–1054 (2004).
- Lalonde, J., Lachance, P. & Chaudhuri, A. Developmental and activity-dependent genomic occupancy profiles of CREB in monkey area V1. *Genes Brain Behav.* **8**, 149–160 (2009).
- Kornhauser, J. M. et al. CREB transcriptional activity in neurons is regulated by multiple, calcium-specific phosphorylation events. *Neuron* **34**, 221–233 (2002).
- Ch'ng, T. H. et al. Activity-dependent transport of the transcriptional coactivator CRTC1 from synapse to nucleus. *Cell* **150**, 207–221 (2012).
- Bulger, M. & Groudine, M. Functional and mechanistic diversity of distal transcription enhancers. *Cell* **144**, 327–339 (2011).
- Vercauteren, K., Pasko, R. A., Gleyzer, N., Marino, V. M. & Scarpulla, R. C. PGC-1-related coactivator: immediate early expression and characterization of a CREB/NRF-1 binding domain associated with cytochrome c promoter occupancy and respiratory growth. *Mol. Cell. Biol.* **26**, 7409–7419 (2006).
- Li, M. et al. Integrative functional genomic analysis of human brain development and neuropsychiatric risks. *Science* **362**, eaat7615 (2018).
- Satterstrom, F. K. et al. Large-scale exome sequencing study implicates both developmental and functional changes in the neurobiology of autism. *Cell* **180**, 568–584 (2020).
- An, J.-Y. Y. et al. Genome-wide de novo risk score implicates promoter variation in autism spectrum disorder. *Science* **362**, eaat6576 (2018).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2021

## Methods

**Statistical analyses.** Box plots depict median (center),  $\pm 1$  quartile (box) and  $\pm 2$  quartiles (whiskers), unless otherwise noted. *t*-tests are unpaired and without correction for multiple comparisons, unless otherwise noted. The *P* values of motif enrichments were obtained using Hypergeometric Optimization of Motif Enrichment (HOMER) findMotifsGenome.pl with default settings. No statistical methods were used to predetermine sample sizes, but our sample sizes are similar to those reported in previous publications<sup>18,39</sup>. For ChIPseq experiments, a minimum of two biological replicates (independent production lots of neuronal cultures) were used for each condition. For all other experiments, a minimum of two biological replicates were used but typically 3–4 whenever possible, except for GFAP immunostaining for which one experiment was obtained, and results were corroborated by scRNAseq and total RNAseq. This was deemed sufficient owing to the high reproducibility of the cultures as measured by RNAseq and immunostaining. Data collection and analysis were not performed blinded to the conditions of the experiments.

**hGN cultures.** 1G hiPSC-derived GNs were obtained commercially (Cellular Dynamics International iCell Neurons, NRC-100-010-001—now sold as iCell GABANeurons, R1013). Other hGN genotypes (1501, 1505 and CW20049) were purchased as iCell GABANeurons derived from independent hiPSC lines (Supplementary Table 1). Samples were received on dry ice and stored in liquid nitrogen until use. Cell culture plates were pre-coated overnight with a solution of 20  $\mu\text{g ml}^{-1}$  of poly-L-ornithine (Sigma-Aldrich) and 20  $\mu\text{g ml}^{-1}$  of laminin (Invitrogen) in sterile water. Before plating, the cell culture plates were washed twice with sterile water and washed once with iCell Neurons Complete Maintenance Medium (CMM). iCell Neurons were thawed, plated and maintained according to the iCell Neurons User's Guide (version 1.4). Briefly, they were thawed in a 37 °C water bath for 3 min, transferred to a 50-ml centrifuge tube, diluted in warm CMM and plated at 150,000–200,000 cells per  $\text{cm}^2$ . A complete media change was done 24 h after plating. Every 3–4 d thereafter, cells were fed by removing 50–70% of media and replacing with fresh 37 °C media. Neurons were grown until the fourteenth to seventeenth day in vitro (DIV 14–17).

Independent production lots of 1G iCell GABANeurons were obtained from independent differentiation instances by Cellular Dynamics International. Therefore, different production lots were treated as 'biological replicates' within one genotype (1434) for statistical purposes. When assessing genomic patterns across genotypes, one production lot of each genotype (1434, CW20049, 1501 and 1505) was obtained, grown and assayed in parallel, and each genotype was, instead, treated as a biological replicate. Technical replicates were obtained for statistical purposes when cells from the same production lot were grown, harvested and assayed more than once and on separate experimental days.

**Human tissue specimens.** Research performed on samples of human origin was conducted according to protocols approved by the institutional review boards of Beth Israel Deaconess Medical Center, Boston Children's Hospital and Harvard University Faculty of Medicine. Fetal brain tissue was received after release from clinical pathology, with a maximum postmortem interval of 4 h. Tissue was transported in HBSS medium (Life Technologies) on ice to the laboratory for research processing and stored at –80 °C. Cases with known anomalies were excluded. Two human fetal cortical samples were used for CREB and CRT1 ChIPseq experiments: FB08 (19 gestational weeks) and FB25 (23 gestational weeks). Gestational ages were determined using fetal foot length.

**KCl depolarization of cultured neurons.** DIV 14–17 hGNs or DIV 7 E16.5 mouse cortical cultures (for luciferase assays) were silenced overnight with 1  $\mu\text{M}$  TTX and 100  $\mu\text{M}$  APV for 12–16 h. Then, the next day, neurons were left in the silenced condition (unstimulated) or depolarized for various times with KCl depolarization buffer (170 mM KCl, 2 mM  $\text{CaCl}_2$ , 1 mM  $\text{MgCl}_2$ , 10 mM HEPES, solution pH 7.4) to a final concentration of 31% in the neuronal culture medium, achieving a final KCl concentration of 53 mM.

**siRNA treatment of hGN cultures.** Accell siRNA SMARTpool 50 nM (human CRT1 23373 or non-targeting (NT) pool, GE Healthcare Dharmacon) was first resuspended in 1 $\times$  siRNA buffer (Thermo Fisher Scientific, B2000UB100) to a final concentration of 100  $\mu\text{M}$ . siRNA was then added to neuronal cultures at a 1:100 dilution for a final concentration of 1  $\mu\text{M}$  in media every 72 h starting at DIV 7 until the cultures were harvested. During each siRNA addition, neuronal media were replaced with 50% hGN-conditioned media and 50% new media. CRT1 siRNA treatment for 6 d resulted in an average knockdown of 61.8% (s.d. = 18.7%); treatment for 9 d resulted in 64.1% knockdown (s.d. = 4%).

**Immunocytochemistry, imaging and quantification.** DIV 14–17 hGN cultures, grown on sterile poly-L-ornithine and laminin-coated no. 1 glass coverslips (Thermo Fisher Scientific, 501215159), were fixed with a solution of 4% paraformaldehyde and 4% sucrose in 1 $\times$  PBS pH 7.4 for 8 min at room temperature, blocked for 1 h at 4 °C with 0.1% (wt/vol) gelatin and 0.3% (vol/vol) Triton X-100 in 1 $\times$  PBS pH 7.4 (GDB) and incubated overnight at 4 °C with the following primary antibodies diluted in GDB buffer: anti-NESTIN (mouse, 1:500,

R&D Systems, MAB1259), anti GFAP (rabbit, 1:500, Dako, Z033429-2), anti-MAP2 (chicken, 1:1,000, Lifespan Biosciences, LS-C61805), anti-neurofilament (SMI-312) (1:1,000, Abcam, ab24574), anti-CREB (1:500, EMD Millipore, 06-863), anti-CRTC1 (1:500, Bethyl Laboratories, 300-769) and anti-Ser133 pCREB (1:500, Cell Signaling Technology, 87G3, 9198S). All secondary antibodies were AlexaFluor-conjugated (Life Technologies/Invitrogen) and incubated at 1:1,000 dilution in GDB buffer for 1 h at 4 °C. Coverslips were mounted with DAPI Fluoromount G (SouthernBiotech). Immunostained samples were imaged on either an AxioVision Imager Z1 (Zeiss) or a Leica SP8X confocal microscope (Neurobiology Imaging Facility). Live cultures were imaged on a Nikon Eclipse TS100. Immunostained culture images from 1–2 separate coverslips (experimental replicates) of 1–3 production lots of hGNs (biological replicates) were manually counted to quantify the number of cells expressing a given protein. Proportions were calculated by dividing by the total number of DAPI + nuclei in the same images, obtained by custom Fiji macros (available upon reasonable request).

**qPCR.** Isolated total RNA was extracted using Trizol (Invitrogen) followed by RNeasy Micro Kit (Qiagen) including on-column DNaseI digestion. Complementary DNA (cDNA) libraries were synthesized using a High Capacity cDNA Reverse Transcription Kit (Applied Biosystems) or the SuperScript IV VIL0 Master Mix with ezDNase Kit (Invitrogen). The cDNA was the source of input for qPCR with SYBR Green reagents (Applied Biosystems) or TaqMan reagents (see below), using a Step One Plus Real-Time PCR Instrument for *FOs* and *NPAS4* transcript detection or a Quant Studio 3 System (Thermo Fisher Scientific) for all other transcripts. Using Microsoft Excel for Mac version 16, the *NPAS4* and *FOs* relative expression plot was constructed using concentration values that were normalized to corresponding *GAPDH* concentrations, whereas qPCR relative expression plots in Extended Data Fig. 3c were constructed using concentration values normalized to *PGK1*. The following primer sets were used:

*DHCR7* F-5'-TCCACAGCCATGTGACCAATGC-3', R-5'-CGAAGTGGT CATGGCAGATGTC-3'. *DHCR24* F-5'-CAGGAGAACCCTTCGTGGAAG-3', R-5'-CCACATGCTTAAAGAACCACGGC-3'. *FOs* F-5'-AGGTCGTGCAGAG ATGCTC-3', R-5'-AGGTCGTGCAGAGTCT-3'. *GPR3* F-5'-CCTTCCAGTCTACTGCCTGCT-3', R-5'-TCTGCACATCCTGTGCGGAA-3'. *GAPDH* F-5'-GTCTCCTCTGACTTCAACAGCG-3', R-5'-ACCACCTGTTGCTGTAGC CAA-3'. *KCNQ3* F-5'-CGTCTGATTGCCACCTTTT-3', R-5'-TTCTGACGG TGTGTCTCCTGCA-3'. *MMP1* F-5'-ATGAAGCAGCCAGATGTGGAG-3', R-5'-TGGTCCACATCTGCTCTTGGCA-3'. *NPAS4* F-5'-TGGTCTTACTGA TGAGTTGCAT-3', R-5'-TCCCCTCCACTTCCATCTT-3'. *PGK1* F-5'-CCGCTT CATGTGGAGGAAGAAG-3', R-5'-CTCTGTGAGCAGTGCCAAAAG-3'. *PTCHD1* F-5'-CGGATTGGTGACCAATAGCCTG-3', R-5'-CGGATTGGTG ACCATAAGCCTG-3'. *SHANK3* F-5'-AGGATCACACCCGCCGAGATTA-3', R-5'-CTACAGACTTGGTCCGTGAATC-3'. *XIRP1* F-5'-AACTCCAGG CAGCAGTGGGAAT-3', R-5'-AGAGAGTAGGCAGTCAGTCCGA-3'.

TaqMan reagents (Thermo Fisher Scientific) were used for *CACNB2* transcript measurements. For TaqMan qPCR of *CACNB2* transcripts, 9-d treatment NT or CRT1 siRNA cultures were used that were unstimulated or KCl depolarized for 2 h. The following Taqman probes were used: (1) custom probe for long transcripts NM\_201597.2, NM\_201593.2 and NM\_201596.2; (2) custom probe for long transcripts NM\_201572.3, NM\_201571.3 and NM\_001167945.1; (3) Hs\_01100744\_m1 probe for pan-*CACNB2* transcripts; and (4) Hs01110867\_m1 probe for short transcript NM\_201570.2.

**Electrophysiology.** Two biological replicates of hGNs were cultured separately, as described above, on sterile poly-L-ornithine and laminin-coated no. 1 glass coverslips (Thermo Fisher Scientific, 501215159) until DIV 14–15. Coverslips were transferred to a recording chamber mounted on an upright microscope (Olympus BX51WI) and imaged using infrared-differential interference contrast with an  $\times 40$  water immersion Olympus objective. Cells were maintained with perfusion of 34 °C artificial cerebral spinal fluid (aCSF) containing (in mM) 125 NaCl, 2.5 KCl, 1.25  $\text{NaH}_2\text{PO}_4$ , 25  $\text{NaHCO}_3$ , 11 glucose, 2  $\text{CaCl}_2$  and 1  $\text{MgCl}_2$ . The aCSF was continuously equilibrated by bubbling with 95%  $\text{O}_2$ /5%  $\text{CO}_2$ . Whole-cell voltage-clamp and current-clamp recordings were obtained by forming intracellular seals with target neurons with patch pipettes pulled from borosilicate glass (BF150-86-7.5, Sutter). Pipettes (2–4-M $\Omega$  pipette resistance) were pulled with a P-97 flaming micropipette puller (Sutter). Pipettes were filled with either a Cs<sup>+</sup>-based internal recording solution containing (in mM) 135 CsMeSO<sub>4</sub>, 10 HEPES, 1 EGTA, 4 Mg-ATP, 0.3 Na-GTP, 8 Na<sub>2</sub>-phosphocreatine, 3.3 QX-314 (Cl-salt), pH adjusted to 7.3 with CsOH and diluted to 290–295 mOsm  $\text{kg}^{-1}$  for voltage-clamp recordings or a K<sup>+</sup>-based internal recording solution containing (in mM) 120 KMeSO<sub>4</sub>, 10 HEPES, 0.2 EGTA, 8 NaCl, 10 KCl, 4 Mg-ATP, 0.3 Na-GTP, pH adjusted to 7.3 with CsOH and diluted to 290–295 mOsm  $\text{kg}^{-1}$  for current-clamp recordings. Putative GABA-mediated inhibitory post-synaptic currents were isolated by voltage clamping at 0 mV, the reversal potential for excitatory currents, and confirmed with 10  $\mu\text{M}$  gabazine (SR-95531, Tocris).

Voltage- and current-clamp recordings were amplified and filtered at 3 kHz using a Multiclamp 200B (Axon Instruments) and digitized at 10 kHz with a National Instruments acquisition board. Data were saved with a custom version of ScanImage written in MATLAB (Mathworks; <https://github.com/>

bernardosabatini/SabalabSoftware\_Nov2009). All data were analyzed on MATLAB, and spontaneous excitatory post-synaptic currents and inhibitory post-synaptic currents were analyzed using a modified Peaker Analysis Toolbox (Andrew Penn (2020)—Peaker Analysis Toolbox (<https://www.github.com/acp29/Peaker>), retrieved March 27, 2020).

**Western blot.** hGN lysates were collected and boiled for 3 min in SDS sample buffer, resolved by SDS-PAGE using 8–10% polyacrylamide gels and Precision Plus protein ladder (Bio-Rad). Samples were then transferred to nitrocellulose and immunoblotted using primary antibodies (see antibodies above) at 1:1,000 dilution and then secondary 800-nm DyLight-conjugated antibodies (Rockland Immunochemical, 75934-472) at 1:50,000 dilution, before visualization and quantification with a Licor Odyssey system. Original blot images are provided as Supplementary Information.

**High-throughput DNA sequencing.** All cDNA libraries for total RNAseq, scRNAseq, ChIPseq and ATACseq were sequenced on a Nextseq 500 Illumina sequencer.

**Total RNAseq library preparation.** 1G total RNA was isolated from hGN cultures by using Trizol (unstimulated,  $n = 6$ ; 15 min,  $n = 3$ ; 1 h,  $n = 3$ ; 2 h,  $n = 6$ ; 4 h,  $n = 3$ ). RNA was then purified using RNeasy Micro Kit (Qiagen) including an on-column DNaseI digestion, and strand-specific and single-end cDNA libraries were generated using the SE RNA-seq Library Kit (Illumina) or the NEBNext Ultra Directional RNA Library Prep Kit for Illumina. Ribosomal RNA depletion was performed using the RiboMinus Eukaryote Kit for RNAseq (Thermo Fisher Scientific) and verified using a Bioanalyzer RNANano kit (Agilent) before library preparation and high-throughput sequencing. Initially, three different lots of hGNs were assayed for unstimulated, 15-min and 2-h depolarization, and, for logistical reasons associated with sample processing, additional replicates were assayed unstimulated and 1 h, 2 h and 4 h after depolarization. Given that all Spearman correlation coefficients across pairwise comparisons of six replicates at any time point remained greater than 0.9522, we opted to include data from the additional replicates.

4G total RNA was isolated from a single culture and production lot of hGN cultures derived from four independent hiPSC lines by using Trizol (unstimulated,  $n = 4$ ; 15 min,  $n = 4$ ; 1 h,  $n = 3$ ; 2 h,  $n = 4$ ; 4 h,  $n = 4$ ). RNA was then purified using RNeasy Micro Kit (Qiagen, 74004), including an on-column DNaseI digestion, and strand-specific and single-end cDNA libraries were generated using 500 ng of total RNA with the NEBNext rRNA Depletion Kit (E6310) and NEBNext UltraDirectional RNA Library Prep Kit from Illumina (E7420) following the manufacturer's protocol, with the omission of actinomycin D during first-strand synthesis. This resulted in up to 20% of reads mapping to the opposite strand from the original template, but it does not affect our downstream analysis as we are only quantifying sense reads mapping to non-overlapping gene annotations.

**Total RNAseq data processing and differential gene expression analysis.** Raw high-throughput sequencing reads in FASTQ format were processed using our in-house software MAPToFeatures to quantify normalized gene exonic expression levels, as previously described<sup>6</sup>, using hg38 (GRCh38.p2) for total RNAseq gene expression analysis in Figs. 1 and 2 or hg19.3 (GRCh37.p5) for total RNAseq gene expression analysis to be compared with ChIPseq data annotations. Differential gene expression analysis using EdgeR was performed to identify gene transcripts whose levels significantly changed after KCl depolarization. EdgeR was used to compare the transcript levels at each KCl depolarization time point (15 min, 1 h, 2 h or 4 h) to the unstimulated condition, after quantile normalization of all replicates for the two time points. A gene's expression was considered activity regulated if the following thresholds were met: (1) at least a 1.5-fold change (either increasing or decreasing); (2) a Benjamin–Hochberg-corrected  $q$  value consistent with FDR = 0.05; and (3) a minimal expression level requiring at least five reads in all reps at both times as well as reads per kilobase of transcript, per million mapped reads (RPKM)  $\geq 0.05$  for at least one of the two time points compared.

**Gene set enrichments.** Gene set enrichments were tested using Pearson's chi-squared test with Yates' continuity correction in R.

**Determining evolutionary conservation of gene sequences.** Evolutionary lineage analyses of the expressed LINC-RNAs and ZNFs were performed using multiple resources, including ENSEMBL's Comparative Genomics Tool (<https://www.ensembl.org/info/genome/compara/index.html>), GenTree (<http://gentree.ioz.ac.cn/index.php>) and HomoloGene (<https://www.ncbi.nlm.nih.gov/homologene>). A gene was considered 'conserved' if it is conserved in multiple lineages outside the primate lineage.

**scRNAseq (InDrops) sample and library preparation.** Two sets of biological replicates, each of approximately 5,000 cells, were obtained from unstimulated and 1-, 2- and 4-h KCl depolarized cultures, and a third set of biological replicates, each of approximately 5,000 cells, was obtained from unstimulated and 1- and 2-h KCl depolarized cultures. Cultures were dissociated to single-cell suspension

and prepared for InDrops encapsulation at staggered times to allow for minimal differences in wait time before reverse transcription. Dissociation was performed similarly to methods described in Hrvatin et al.<sup>18</sup> hGN cultures at each time point were washed once with CMF-PBS and incubated for 20 min at 37 °C with papain and DNase (Worthington) in the presence of a five-inhibitor cocktail to preserve the transcriptional state of the neurons: 1 mM tetrodotoxin citrate, 100 mM D-AP5, 5 mg ml<sup>-1</sup> of actinomycin D, 20 mM Triptolide and 10 mg ml<sup>-1</sup> of anisomycin (Sigma-Aldrich). Minimal trituration was performed to resuspend dissociated cells, and single-cell suspensions contained very little cellular debris after being passed through a 0.02- $\mu$ m cell strainer and resuspended in dissociation solution containing 0.04% BSA and 15% Optiprep (Sigma-Aldrich) and placed on ice. Cellular suspensions were not gradient centrifuged. Single cells were encapsulated with 1CellBio v2 barcoded hydrogel beads using the InDrops platform followed by reverse transcriptase reaction at the Single Cell Core, Department of Systems Biology, Harvard Medical School. Libraries were subsequently generated according to the 1CellBio v2 manufacturer's protocol. After parallel library preparation, all libraries generated from each replicate were pooled and sequenced together.

**scRNAseq data processing.** Raw reads in FASTQ format were converted into counts tables of transcripts per gene for every cell using a previously published pipeline<sup>51</sup>. We built a reference transcriptome index from the ENSEMBL annotation of the *Homo sapiens* genome GRCh37.75 (soft-masked) using Bowtie version 1.2.2 (ref. 52). Note that the ENSEMBL annotation does not include all gene names existing in the annotation used for total RNAseq analysis, and, therefore, certain transcripts (for example, *LINC00602*) were not quantified by this analysis. All quality control steps and mapping to the transcriptome were performed using default parameters unless explicitly specified.

**scRNAseq Seurat-based analysis.** Counts tables were loaded into R (version 3.4.1) and analyzed using the Seurat package (versions 2.3.4 and 3.0.0)<sup>53,54</sup>. In total, we sequenced 244,155 cells. Cells were filtered by excluding cells with unique molecular identifier (UMI) counts less than 500, UMI counts greater than 10,000, mitochondrial gene expression greater than 30% or ribosomal gene expression greater than 15% using Seurat's FilterCells function. The data were then log-normalized to 10,000 transcripts per cell and scaled using a negative binomial model as well as by UMI, mitochondrial gene expression and ribosomal gene expression using NormalizeData and ScaleData. Variable genes within each sample were found using the FindVariableGenes function with parameters  $x.low.cutoff = 0.1$ ,  $x.high.cutoff = 3$  and  $y.cutoff = 1$ . Data from all biological replicates and time points were merged with RunMultiCCA, using the top 1,000 scaled variable genes from each sample to calculate 30 canonical vectors, and AlignSubspace, using 15 dimensions. Clustering analysis was performed with FindClusters, using 15 dimensions and a resolution of 0.6. Dimensionality reduction was performed with RunTSNE, using 15 dimensions. We discovered that one cluster had a considerably lower number of UMI counts, so we excluded this cluster from analysis. We re-ran our filtering, normalization and scaling steps for each sample excluding the apoptotic cluster cells.

Our final analysis used Seurat version 3.0.0, in which RunMultiCCA had been deprecated. For each of our filtered, normalized and scaled samples, we ran the functions UpdateSeuratObject, NormalizeData and FindVariableFeatures (selection.method='vst', nfeatures = 2,000). We merged our samples with the functions FindIntegrationAnchors and IntegrateData, using 30 dimensions, and ran ScaleData on the combined Seurat object. We then performed principal component analysis (RunPCA) to compute 30 components and dimensionality reduction with Uniform Manifold Approximation and Projection (UMAP) (RunUMAP) for 30 dimensions. A shared nearest neighbors graph was constructed with FindNeighbors using 20 dimensions, and cells were clustered with FindClusters at 0.5 resolution. Our final Seurat object contains 37,101 cells and 15 clusters.

**scRNAseq differential gene expression analysis.** Single-cell RNA sequencing differential gene expression analysis was performed with Monocle<sup>55</sup> (version 2.6.4) for each cluster, comparing gene expression at 1 h, 2 h and 4 h after KCl depolarization against gene expression in unstimulated cells. For a gene to be called 'inducible', we set the thresholds that the gene had to be (1) expressed in at least 1% of all cells in the comparison being performed (for example, cluster 1 4-h depolarized cells and cluster 1 unstimulated cells) and (2) have a fold change greater than 1.5 with a  $q$  value smaller than 0.05.

**Cumulative distribution plot.** A list of human TFs was downloaded from [http://humantfs.ccb.utoronto.ca/download/v\\_1.01/TF\\_names\\_v\\_1.01.txt](http://humantfs.ccb.utoronto.ca/download/v_1.01/TF_names_v_1.01.txt) and intersected with our single-cell early-response and late-response genes. The number of clusters each TF was expressed in was calculated to generate the cumulative distribution plot.

**Human BrainSpan data representation.** Expression levels (RNA-Seq\_Gencode\_v10\_summarized\_to\_genes) from the publicly available BrainSpan atlas (<http://www.brainspan.org/static/download.html>) are derived from RNAseq for 19,364 genes consistent with RefSeq and 524 human samples, including 42 individuals, 31

different ages ranging from embryonic to adulthood and 26 specific brain regions. The data shown in Extended Data Fig. 2g–j cover six separate brain regions that we define as neocortex (by combining data labeled as AIC, DFC, IPC, ITC, M1C, M1C-S1C, MFC, Ocx, OFC, PCx, S1C, STC, TCx, V1C and VFC), striatum (data labeled as STR), hippocampus (data labeled as HIP), amygdala (data labeled as AMY), thalamus (by combining data labeled as DTH and MD) and cerebellum (by combining data labeled as CB and CBC). For each gene, data for all samples at all available time points in each brain region were fit via a local polynomial regression (the Loess function in R version 3.3.0) and shown as mean Loess curves interpolated across the whole age range. The width of the one-standard-error side bands was similarly calculated via a Loess fit to the standard errors deduced at whatever ages data were available in each region for each gene. Spearman correlation coefficients were determined for all pairwise comparisons between normalized average expression data from available time points among the eight developmental time points across the 19 brain regions separately (as represented in Supplementary Table 2) and normalized average expression data from the unstimulated conditions of our 1G and 4G total RNAseq datasets. For comparison, total RNAseq data from five primary human fetal brain cultures (Ataman et al.<sup>6</sup>) were re-analyzed using exactly the same parameters and thresholds as the hGN total RNAseq in this study for subsequent equivalent correlation analysis.

**ChIPseq.** Typically, 2–6 million nuclei were used for a single ChIP experiment. 1G hGNs from independent production lots (independent neuronal differentiations) served as biological replicates. For 4G hGN H4K27ac, for each of the four genotypes, two independent cultures derived from the same neuronal differentiation lot (technical replicates) of cells were stimulated and harvested, and ChIPseq libraries were prepared on two different experimental days from 500,000 to 900,000 nuclei for each library. Data from each of the four genotypes were considered biological replicates.

Frozen (–80 °C) primary human tissue was weighed in a 15-ml tube and placed on ice for 15 min. Tissue was homogenized using a Polytron Homogenizer in 5× (at least 5 ml) ice-cold crosslinking solution (see below) for 5 s or just until tissue is homogenized. Homogenized samples were incubated for 10 min at room temperature while rocking before quenching with 125 mM glycine for 5 min at room temperature. Homogenized tissue was then pelleted at 830g for 5 min at 4 °C. The supernatant was aspirated, and the pellet was resuspended in 10 ml of ice-cold sterile CMF-PBS. The pellet was then washed a second time with CMF-PBS before resuspending in 10 ml of Buffer I (see below) and transfer to pre-chilled a 15-ml dounce homogenizer. Samples were dounced on ice 10× with a tight pestle and then transferred back to a clean 15-ml tube. Nuclei were pelleted for 10 min at 4 °C at 845 relative centrifugal force (RCF) and then rinsed in Buffer II (see below), and the remaining protocol was followed as described below.

1G hGN cultures were KCl depolarized for different amounts of time just before crosslinking. For CREB, pCREB and CRT1 ChIPseq experiments used to define CREB complex peaks, cultures were treated with NT control siRNA (GE Healthcare Dharmacon) for 9 d and then were unstimulated or 15-min KCl depolarized before crosslinking. To crosslink protein–DNA complexes, we removed media from cell cultures and incubated with crosslinking solution (1% formaldehyde, 100 mM NaCl, 1 mM EDTA pH 8.0, 0.5 mM EGTA pH 8.0, 50 mM HEPES pH 7.9) for 10 min at room temperature. Crosslinking was quenched with 125 mM glycine for 5 min at room temperature. Cells were rinsed twice with ice-cold PBS and scraped in PBS containing complete protease inhibitor cocktail tablets (Roche). Cells were then pelleted and lysed in 2 ml of Buffer I (50 mM HEPES KOH, pH 7.5, 140 mM NaCl, 1 mM EDTA pH 8.0, 10% glycerol, 0.5% NP-40, 0.25% Triton X-100, 10 mM β-glycerophosphate, 10 mM sodium fluoride, 1× protease inhibitor cocktail, 100 nM okadaic acid) and incubated for 10 min on a rotating platform at 4 °C. Nuclei were then pelleted for 10 min at 4 °C at 845 RCF. The isolated nuclei were then rinsed in 2 ml of Buffer II (200 mM NaCl, 1 mM EDTA pH 8.0, 0.5 mM EGTA pH 8.0, 10 mM Tris pH 8.0, 10 mM β-glycerophosphate, 10 mM sodium fluoride, 1× protease inhibitor cocktail, 100 nM okadaic acid). They were again incubated for 10 min on a rotating platform at 4 °C and pelleted for 10 min at 4 °C at 845 RCF. Finally, 1.5 ml of Buffer III was added to nuclei (10 mM Tris-HCl pH 8.0, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% Na-deoxycholate, 0.5% N-laurylsarcosine, 10 mM NaF, 1× protease inhibitor cocktail, 100 nM okadaic acid). 4G ChIPseq experiments were harvested in exactly the same way but without NaF or okadaic acid in the buffers.

1G nuclear extracts were then placed in 15-ml polyallomer tubes (Thermo Fisher Scientific, NC9530407), and sonication was carried out using a Bioruptor sonicator on high power with a 30-s ‘on’ interval and a 45-s ‘off’ interval per cycle, for a total of 32 cycles at 4 °C. 4G nuclear extracts were biorupted on a new Bioruptor sonicator that achieved a similar level of chromatin fragmentation using, instead, ten cycles, as determined by agarose gel electrophoresis. After sonication, tubes were centrifuged at 376 RCF briefly, and material was moved to 2-ml DNA LoBind Eppendorf tubes and spun down at 16,000 RCF for 10 min to remove insoluble materials. Supernatant was transferred to a new tube, and Triton-X 100 was added to make 10% of the total solution. Inputs between different samples were normalized by obtaining an 80-μl aliquot from each sample and incubating with 120 μl of TE/1% SDS for 15–30 min at 95 °C. These samples were then purified with Qiagen PCR purification kit and eluted into 60 μl of EB buffer. Concentrations

of samples were then determined via NanoDrop. Using these concentrations, antibody–bead coupling reactions were performed so that the antibody:sample ratio would remain consistent across all samples.

Samples were then pre-cleared by adding 15 μl of pre-rinsed protein A Dynabeads. Beads were washed in 200 μl of TBSTPb (0.01% BSA and 0.2 mM PMSF in 1× TBST) for every 15 μl of beads. They were washed three times for 5 minutes each, and then lysate was pre-cleared for 2–4 h at 4 °C while rotating. Simultaneously, beads were incubated with antibody for the same period of time by aliquoting 15 μl of beads per immunoprecipitation after washes and incubating with the relevant antibody: H3K27Ac (Abcam, ab4729), H3K4me3 (Millipore, 07-473), CREB EMD (Millipore, 06-863), CRT1 (Bethyl Laboratories, 300–769) or Ser133 pCREB (Cell Signaling Technology, 87G3, 9198 S). Note that 87G3 rabbit monoclonal antibody also detects the phosphorylated form of the CREB-related protein ATF-1, which is similarly phosphorylated in response to stimulation and can dimerize with CREB<sup>56</sup>. For H3K27Ac, we diluted the antibody 1:100 and then used 15 μl of diluted antibody for 40 μg of chromatin. For CREB, we used 4.5 μl of 1 mg ml<sup>–1</sup> antibody per ChIP, and, for CRT1, we used 22.5 μl of 0.2 mg ml<sup>–1</sup> antibody. For pCREB, we used 26 μl of antibody per ChIP. Thus, for TF ChIP, we kept a constant 4.5 μg of antibody per ChIP.

After coupling antibody to beads, the beads were then washed again in TBSTPb 3 × 5 min at room temperature, and pre-cleared lysates were incubated with antibody-coupled beads overnight at 4 °C with rotation. Before incubating with antibody, 1/20th of the volume of pre-cleared lysates was saved for use as input material.

The next day, the beads bound to immune complexes were placed on a magnet to remove supernatant, which was saved as flow-through. They were then washed twice with each of the following buffers, with which they were incubated for 5 min at 4 °C with rotation. The buffers used were the following: low-salt buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl, pH 8.1, 150 mM NaCl), high-salt buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl, pH 8.1, 500 mM NaCl) and LiCl buffer (0.25 M LiCl, 1% IGEPAL CA630, 1% deoxycholic acid (sodium salt), 2 mM EDTA, 20 mM Tris, pH 8.1). The washed beads were then rinsed once with 1× TE (10 mM Tris-HCl, pH 8.0, 1 mM EDTA). The immunoprecipitated materials were eluted from the beads twice by adding 100 μl of elution buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA, pH 8.0, 1% SDS) to each ChIP reaction and incubating the sample at 65 °C for 30 min with brief vortexing every 10 min. Elution buffer was also added to the input material (for a total of 200 μl), and these samples were processed with the ChIP samples for the remainder of the experiment. All eluates were reverse crosslinked at 65 °C overnight for 12–16 h.

To each eluate, we then added 10 μg of RNase A per sample and incubated for 1 h at 37 °C on beads. Then, 7 μl of proteinase K was added per tube (20 mg ml<sup>–1</sup> of stock), and samples were incubated for 2–3 h at 55 °C with shaking. DNA was then isolated via phenol–chloroform extraction where an equal volume of phenol–chloroform (pH 7.9) was added to each tube, and tubes were vortexed 2 × 30 s. Samples were then spun down at 17,115 RCF for 3 min, and the top aqueous phase was transferred to a new tube. Samples were purified via QIAquick PCR Purification Kit (Qiagen) with 5 M sodium acetate (pH 5.5), and DNA fragments were eluted in 60 μl of EB buffer.

1G libraries for ChIPseq were prepared using Nugen Ovation Ultralow Library System V1 using 2 ng of starting DNA and performing 18 cycles of amplification. 4G ChIPseq libraries were prepared using Nugen Ovation Ultralow Library System V2 with amplification cycles adjusted for each library as measured by parallel SYBR Green qPCR reaction. Library DNA sizes were then confirmed to be majority 300–1,000 bp in length using 2100 Agilent Bioanalyzer for high-sensitivity DNA. Single-end 75-bp reads were obtained from all ChIPseq libraries using a NextSeq 500 (Illumina).

**ChIPseq data processing.** 1G ChIPseq library sequencing data were obtained as FASTQ files from four separate lanes for each sample. FASTQ files were concatenated for each sample, and then reads were truncated to 70-bp lengths and aligned to the hg19 assembly of the human genome using the Burrows–Wheeler Aligner<sup>57</sup> with default settings. The SAM files produced by the aligner were converted to BAM files, and duplicates were removed using the samtools package. BAM files were then converted to BED files and used for peak calling using Model-based Analysis of ChIP-Seq (MACS) using an input control. For TF ChIP, we used the default settings, and, for H3K27Ac ChIPseq, we used the following parameters: –nomodel –shiftsize = 150.

To visualize data on the UCSC Genome Browser, genome-aligned BED files from all biological replicates were concatenated (pooled) and reads were extended to 200-bp fragments and then normalized to 10 M reads. The resulting bedgraph files were converted to bigwig format. Data were visualized along with UCSC Genome Browser-provided ‘vert.cons’ track displaying sequence conservation across 100 vertebrate species using PhyloP with default settings. ‘half CRE’ track was generated by locating all hg19 coordinates for the sequence motif CGTCA<sup>58</sup>.

For each ChIPseq experiment, peaks were called separately for biological replicates and then intersected. 1G H3K27ac reproducible peak coordinates were those from one biological replicate that intersected with peaks in at least a second biological replicate. The global 1G H3K27ac peak lists were obtained by merging all reproducible peaks present in unstimulated and stimulated conditions.

4G H3K27ac reproducible peaks from two technical replicates of each of the four genotypes were merged across genotypes and time points to create a global peak list. To define the activated CREB complex peaks, we called peaks for each replicate of CREB, pCREB and CRT1 ChIP and then defined reproducible peaks as those present in two or more biological replicates of each ChIP. We intersected these reproducible peaks for each mark to obtain the high-confidence set of 5,484 CREB complex peaks. Note that this method will not include many of the lower-confidence peaks or regions of CREB/pCREB binding without CRT1, which could still be biologically relevant in certain neuronal subpopulations.

To determine differentially 1G H3K27ac ChIPseq peaks, we used merged reproducible peak coordinates from the all replicates across all time points to generate a 1G global peak list. The number of raw reads under this set of peaks in each biological replicate was determined using the HOMER annotatePeaks function (default parameters). These read counts were used for edgeR differential expression analysis software<sup>59</sup> to determine differentially acetylated peaks using tagwise dispersion and an FDR cutoff of 0.05.

To generate aggregate plots, tag directories of ChIPseq BED files were created using the HOMER (version 4.9) command MakeTagDirectory. These tag directories were then used as input into the HOMER command annotatePeaks.pl with parameters -size 5000 -hist 10.

For 4G H3K27ac ChIPseq analysis, FASTQ files for each sample were concatenated and then trimmed using CutAdapt V2.5. The adapter sequences (Nugen) 'AGATCGGAAGAGCACACGTCTGAACTCCAGTCA' and 'AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT' were removed. Reads were mapped to the hg19 genome using Bowtie version 1.2.2 and the following parameters: bowtie -p 1 -S -n 2 -e 70 -m 1 -k 1 -l 70 -best -chunkmbs 200. We called peaks using MACS2 version 2.1.1 with the following parameters: -nomodel -shiftsize = 150. The corresponding input sample for each ChIP sample was used as the control sample for peak calling. Peaks were called for each of two technical replicates at each time point, and then, only peaks that were found in both technical replicates for a given donor cell line were kept. Peaks from all four donors were concatenated and merged and used for differential peak calling (80,638).

To call inducible H3K27ac peaks, we used DESeq2. We used each genotype as a biological replicate in this analysis and compared four replicates (four donors) at 15-min KCl or 2-h KCl stimulated cultures to those from unstimulated cultures. To call inducible peaks for each donor separately, we used both (2) technical replicates from each donor stimulation time point compared to those from the unstimulated condition. Peaks that were called as differential with an adjusted *P* value of less than 0.05 were considered significantly induced or decreased with KCl stimulation.

The IVRL of H3K27ac-enriched genomic sites previously reported in human developing and adult brain tissues was created by merging all reported regions across all developmental stages from Reilly et al.<sup>60</sup>, Vermunt et al.<sup>61</sup> and Vermunt et al.<sup>62</sup>.

**1G ATACseq.** Three different production lots (biological replicates) of 1G hGNs were grown in culture until DIV 17. For each biological replicate, cells were with left silenced (unstimulated), KCl depolarized for 15 min or KCl depolarized for 2 h. To assess regions of open chromatin, ATACseq library preparation was performed as previously described<sup>63</sup>, and libraries were selected for fragments ranging from 200 to 800 bp by gel electrophoresis before high-throughput sequencing. FASTQ files were concatenated for each sample, and then reads were trimmed by quality using Trimmomatic<sup>64</sup> (version 0.33) with the option SLIDINGWINDOW:5:30. Reads were mapped to hg19 with Bowtie version 1.1.2 using the following options: -p 8 -S -n 2 -e 70 -m 1 -k 1 -l 70 -best hg19 -chunkmbs 200. ATACseq peaks were called using MACS2 version 2.1.1, without input control, and using the options -nomodel -extsize 200. Replicates were analyzed with the irreproducible discovery rate (IDR) pipeline, using a threshold of 0.05 for true replicate analysis and a threshold of 0.01 for pooled-consistency analysis. From our three biological replicates at each stimulation condition, the final set of 81,240 ATACseq peaks merged across all conditions are those that pass threshold in all three pairwise comparisons and that also intersect with peaks appearing in the MACS2 peak call. For motif enrichment and heritability enrichment analysis, we used the 500-bp regions centered on MACS2 peak summits within IDR-confirmed ATACseq peaks.

**4G ATACseq.** For each of the four donor hGN cell lines, hGNs were grown in culture until DIV 17. Cells were left silenced (unstimulated), KCl depolarized for 15 min or KCl depolarized for 2 h. To assess regions of open chromatin, ATACseq library preparation was performed as previously described<sup>65</sup>, and libraries were selected for fragments ranging from 200 to 800 bp by gel electrophoresis before high-throughput sequencing. For each of the four donor lines, two technical replicate libraries were prepared for one biological replicate. FASTQ files were concatenated for each sample and trimmed using CutAdapt V2.5. The adapter sequences (Nextera) 'CTGTCTCTTATACACATCTCCGAGCCACGAGA' and 'CTGTCTCTTATACACATCTGACGCTGCCGACGA' were trimmed. Reads were then mapped using Bowtie version 1.1.2 using the following options: -p 8 -S -n 2 -e 70 -m 1 -k 1 -l 70 -best hg19 -chunkmbs 200. Peaks were called for each technical replicate using MACS2 version 2.1.1, without input control, and using the options -nomodel -extsize 200. For each genotype, only peaks called in both technical replicates were kept, and then the peaks from each genotype were concatenated and merged to produce the global 4G ATACseq peak list.

**Motif enrichment and genomic distribution analysis.** Motif enrichment analyses were done with HOMER<sup>65</sup> (version 4.9) using the findMotifsGenome.pl command using human genome hg19 or hg38 when the genomic regions was translated to hg38 coordinates using LiftOver (<https://genome.ucsc.edu/cgi-bin/hgLiftOver>) The search sequence was restricted to the central 500 bp of the ATACseq peaks within the region sets being compared (see 'ATACseq data processing') or the 500 bp upstream of every TSS. The results included redundant motifs, of which only the motif with the smallest *P* value is reported. For each motif enrichment test, a set of H3K27Ac peaks unchanged with KCl stimulation or 500-bp regions upstream of TSSs within H3K27Ac peaks unchanged with KCl stimulation were used as background.

Distance of ChIPseq peaks from TSS was determined by annotating peaks to nearest genes using HOMER annotatePeaks.pl. ChIPseq peaks were defined as 'promoter regions' if determined to intersect a TSS (Refseq hg19) using bedtools intersect and defined as 'enhancer regions' if they did not intersect a TSS.

**ABC enhancer-gene interaction modeling.** We used the ABC model (<https://github.com/broadinstitute/ABC-Enhancer-Gene-Prediction>) to predict enhancer-gene connections in hGNs in each of three stimulation conditions: unstimulated, 15-min depolarization and 2-h depolarization. These predictions were made based on measurements of 1G chromatin accessibility (ATACseq) and histone modifications (H3K27ac ChIPseq), as previously described by Fulco et al.<sup>40</sup>. In each condition, the ABC model reports an 'ABC score' for each element-gene pair, where the element is within 5 Mb of the TSS of the gene. Briefly, for each condition, we did the following. (1) We called peaks on the chromatin accessibility dataset using MACS2 with a lenient *P* value cutoff of 0.1. (2) We counted chromatin accessibility reads in each peak and retained the top 150,000 peaks with the most read counts. We then resized each of these peaks to be 500-bp centered on the peak summit. To this list, we added 500-bp regions centered on all gene TSSs and removed any peaks overlapping blacklisted regions. Any resulting overlapping peaks were merged. (3) We calculated element Activity as the geometric mean of quantile normalized chromatin accessibility and H3K27ac ChIPseq counts in each candidate element region. Chromatin accessibility and H3K27ac ChIPseq signals in each candidate element were quantile normalized. (4) We calculated element-promoter Contact using the average Hi-C signal across ten human Hi-C datasets as described previously. (5) We computed the ABC score for each element-gene pair as the product of Activity and Contact, normalized by the product of Activity and Contact for all other elements within 5 Mb of that gene.

**Luciferase assay.** The use of animals was approved by the Animal Care and Use Committee of Harvard Medical School. Luciferase assays were performed as previously described<sup>6</sup>. Pregnant E15 C57BL/6 female mice (Charles River Laboratories) were used to derive primary mouse embryonic cortical cultures employed in luciferase assay experiments. Mice were used on arrival and not housed. Primary mouse embryonic cortical cultures were derived from over ten mice for these studies. C57BL/6 mice were used to derive the primary embryonic cortical cultures employed.

Cloning of plasmids. Genomic DNA (gDNA) was extracted from iCell Neurons (cat. no. NRC-100-010-001, lot no. 1366825) using DNeasy Blood & Tissue Kit (Qiagen) according to their protocol starting from Step 1d for cultured cells. We varied from the protocol at Step 1d by thawing the vial of iCell Neurons, diluting in 10 ml of CMF-PBS (Gibco, 10010-023) and spinning down at 8,000g to remove freezing medium. At Step 8, we eluted with TE buffer. Stock gDNA was stored at -20°C, and working gDNA was kept at 4°C.

Sequences of interest were cloned from iCell Neurons gDNA using custom primers and Q5 Hot Start Hi-Fidelity MasterMix (2×) (New England Biolabs) and run on a Bio-Rad T100 Thermal Cycler with the following program: 98°C 30 s, 98°C 10 s, variable from 50°C to 70°C 30 s, 72°C 65 s, 72°C 120 s, 4°C hold. PCR products were run on a 1% agarose gel, and bands of interest were cut out. DNA was then purified using GeneJET Gel Extraction Kit (Thermo Fisher Scientific).

Promoter sequences were cloned into pGL4.11[luc2P] (backbone created through SacI and XhoI double digest), and enhancer sequences were cloned into pGL4.11[luc2P] containing a NUCED2 promoter sequence (Addgene, 59744) (backbone created through SbfI and AscI double digest) using Gibson Assembly Master Mix (New England Biolabs). Competent cells (XL10-Gold Ultracompetent Cells, Agilent, 200315) were then transformed with the ligation product and plated on agarose plates with carbenicillin (Sigma-Aldrich). Individual plasmid clones were grown and then purified using GeneJET Plasmid Miniprep Kit (Thermo Fisher Scientific) followed by Sanger sequencing for verification. Plasmids and sequences are available upon reasonable request.

**Disease-associated genes.** SFARI ASD-associated genes (11-21-2018 release) were downloaded on December 18, 2018, from the now archived SFARI site (<https://gene-archive.sfari.org/>), and only genes that were labeled as 'syndromic' or had a gene score of 1, 2 or 3 were used for analysis (389 gene names). All statistical tests for enrichment of/within ASD-associated genes yielded similar conclusions when performed using the updated SFARI gene scoring system (03-04-2020 release) with 472 gene names that meet the criteria of gene score = 1, 2 or syndromic. SCZ-associated genes (349 gene names) were extracted from Supplementary



Table 3 of the publication by the Schizophrenia Working Group of the Psychiatric Genomics Consortium<sup>30</sup> at <https://www.med.unc.edu/pgc/>.

**Partitioning heritability with stratified LDSC.** To perform heritability enrichment analysis, we used stratified LDSC as previously described in Finucane et al.<sup>66</sup>. Given a functional partition of the genome, the analysis determines whether an annotation carries more genetic heritability than expected based on GWAS summary statistics. We obtained a baseline model of 54 annotations from Finucane et al.<sup>66</sup> and augmented the model with the annotations defined from our ChIPseq and ATACseq processing as described above. Specifically, we augmented the model with the following pairs of annotations: H3K27ac regions overlapping RefSeq TSSs (H3K27ac promoter regions) and H3K27ac regions not overlapping RefSeq TSSs (H3K27ac enhancer regions); H3K27ac enhancer regions that are unchanged (constitutive) and H3K27ac enhancer regions that are inducible; 500-bp ATACseq regions within H3K27ac enhancer regions (enhancers) that are constitutive and inducible; and 500-bp regions upstream of a TSS (promoters) that are constitutive and inducible by H3K27ac. We additionally included, in each model, three brain-associated control categories obtained from Rizzardi et al.<sup>67</sup>: (1) a set of regions marked by H3K27ac in human brain; (2) a union of regulatory regions active in brain; and (3) a union of regulatory regions found by ChromHMM using Roadmap Epigenomics data. For each model, we excluded regions in the three control categories that overlapped our annotations included in the model. Following the recommendations of Finucane et al., we excluded the major histocompatibility complex region from analysis and restricted to HapMap3 SNPs (International HapMap3 Consortium) when fitting the model. When comparing enrichment of an annotation across traits, we corrected for multiple hypothesis tests using the Holm step-down procedure.

LDSC provides a statistical test with associated *P* value that any category in the model carries non-zero enrichment for heritability. Given the natural pairing of our additional categories in the model, we extended the LDSC model to test whether one category is significantly more enriched than another category. Let  $C_i$  and  $C_j$  be two categories in the model. LDSC estimates the heritability explained by these categories as  $\hat{h}^2(C_i)$  and  $\hat{h}^2(C_j)$  that are normally distributed as  $N(\mu_i, \sigma_i^2)$  and  $N(\mu_j, \sigma_j^2)$ , respectively, under the assumptions of the model (Finucane et al.<sup>66</sup>). Additionally, these two estimates are correlated because individual SNPs might belong to multiple categories. LDSC estimates the parameters of the normal distributions using ordinary least-squares regression and estimates the covariance,  $\sigma_{ij}^2$ , using a jackknife procedure. Suppose we want to test whether  $C_i$  is significantly more enriched for heritability than  $C_j$ . That is equivalent to testing whether  $\frac{\hat{h}^2(C_i)}{|C_i|} > \frac{\hat{h}^2(C_j)}{|C_j|}$ , where  $|C_i|$  denotes the number of SNPs in category  $i$ . The left- and right-hand terms are both normally distributed because the numerators are normal random variables, and the denominators are constant scalars. Thus,  $\frac{\hat{h}^2(C_i)}{|C_i|} - \frac{\hat{h}^2(C_j)}{|C_j|} \sim N\left(\frac{\mu_i}{|C_i|} - \frac{\mu_j}{|C_j|}, \frac{\sigma_i^2}{|C_i|^2} + \frac{\sigma_j^2}{|C_j|^2} - 2\frac{\sigma_{ij}}{|C_i||C_j|}\right)$ . We, thus, explicitly test whether this quantity is significantly different than zero using the parameter estimates provided by LDSC to determine whether category  $i$  is more enriched than category  $j$ .

Additional comparisons of LDSC analyses shown in Extended Data Fig. 7h, when using summary statistics from two alternative SCZ GWASs, and in Extended Data Fig. 7i–j, when including Alzheimer's disease. The studies from which summary statistics were obtained for each tested trait are provided in Supplementary Table 3.

**Reporting Summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

The sequencing data that support the findings of this study are available under Gene Expression Omnibus accession number GSE136656. Publicly available datasets include ENSEMBL's Comparative Genomics Tool (<https://www.ensembl.org/info/genome/compara/index.html>), GenTree (<http://gentree.ioz.ac.cn/index.php>), HomoloGene (<https://www.ncbi.nlm.nih.gov/homologene>), BrainSpan atlas (<http://www.brainspan.org/static/download.html>), UCSC Genome Browser (<https://genome.ucsc.edu/>), SFARI ASD-associated genes 11-21-2018 release (<https://gene-archive.sfari.org/>) and the Schizophrenia Working Group of the Psychiatric Genomics Consortium (<https://www.med.unc.edu/pgc/>). Source data are provided with this paper.

## References

- Klein, A. M. et al. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* **161**, 1187–1201 (2015).
- Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
- Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).

- Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **33**, 495–502 (2015).
- Qiu, X. et al. Single-cell mRNA quantification and differential analysis with Census. *Nat. Methods* **14**, 309–315 (2017).
- Liu, F., Thompson, M., Wagner, S., Greenberg, M. & Green, M. Activating transcription factor-1 can mediate Ca<sup>2+</sup>- and cAMP-inducible transcriptional activation. *J. Biol. Chem.* **268**, 6714–6720 (1993).
- Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
- Zhang, X. et al. Genome-wide analysis of cAMP-response element binding protein occupancy, phosphorylation, and target gene activation in human tissues. *Proc. Natl Acad. Sci. USA* **102**, 4459–4464 (2005).
- Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
- Reilly, S. K. et al. Evolutionary changes in promoter and enhancer activity during human corticogenesis. *Science* **347**, 1155–1159 (2015).
- Vermunt, M. W. et al. Large-scale identification of coregulated enhancer networks in the adult human brain. *Cell Rep.* **9**, 767–779 (2014).
- Vermunt, M. W. et al. Epigenomic annotation of gene regulatory alterations during evolution of the primate brain. *Nat. Neurosci.* **19**, 494–503 (2016).
- Sharma, N. et al. ARNT2 tunes activity-dependent gene expression through NCoR2-mediated repression and NPAS4-mediated activation. *Neuron* **102**, 390–406 (2019).
- Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
- Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
- Finucane, H. K. et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).
- Rizzardi, L. F. et al. Neuronal brain-region-specific DNA methylation and chromatin accessibility are associated with neuropsychiatric trait heritability. *Nat. Neurosci.* **22**, 307–316 (2019).

## Acknowledgements

We thank M. Chahrouh, J. L. Hecht, J. Partlow and C. A. Walsh for assistance with human tissue collection; C. C. Harwell, C. Mayer and M. T. Garcia for guidance on single-cell sequencing and cell-type identification; K. D. Hansen and A. P. Feinberg for providing additional LDSC brain-associated control category annotations; and A. N. Chang, E. E. Duffy, G. Fishell, D. Malhotra and D. Reich for helpful discussions. We acknowledge the large body of previous work that informed this study and regret omission of relevant citations due to space constraints. This work was supported by the National Institutes of Health: P50MH106933 and R01NS028829 (to M.E.G.), F32NS086270 (to G.L.B.) and T32GM007753 (to E.D.); the ROADS Program funded by F. Hoffmann–La Roche (to M.E.G.); and the Paul G. Allen Frontiers Group (to M.E.G.). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of General Medical Sciences or the National Institutes of Health.

## Author contributions

G.L.B., E.D., B.A. and M.E.G. conceived of the experiments. G.L.B., E.D., A.C.C., D.A.H., B.A. and K.M. performed total RNAseq. G.L.B., K.M., D.R.H. and S.H. performed scRNAseq. E.D. performed 1G ChIPseq and initial disease enrichment analysis. G.L.B. and A.C.C. performed 4G ChIPseq. G.L.B., A.C.C. and K.M. performed ATACseq. M.A.S. performed LDSC heritability enrichment analysis. K.M. performed luciferase assays, TaqMan assays and statistical analyses. B.A. performed evolutionary conservation analysis. G.L.B. and B.A. performed qPCR and immunocytochemistry with assistance from M.R.B. D.A.H. performed human BrainSpan data visualizations and correlations and statistical analyses. A.J.G. performed electrophysiology. J.M.E. performed ABC modeling. M.G.Y. assisted with IVRL comparisons. G.L.B. performed, directed or assisted with all experiments and analyses. G.L.B., M.E.G. and E.C.G. wrote the manuscript with input from all authors.

## Competing interests

The authors declare no competing interests.

## Additional information

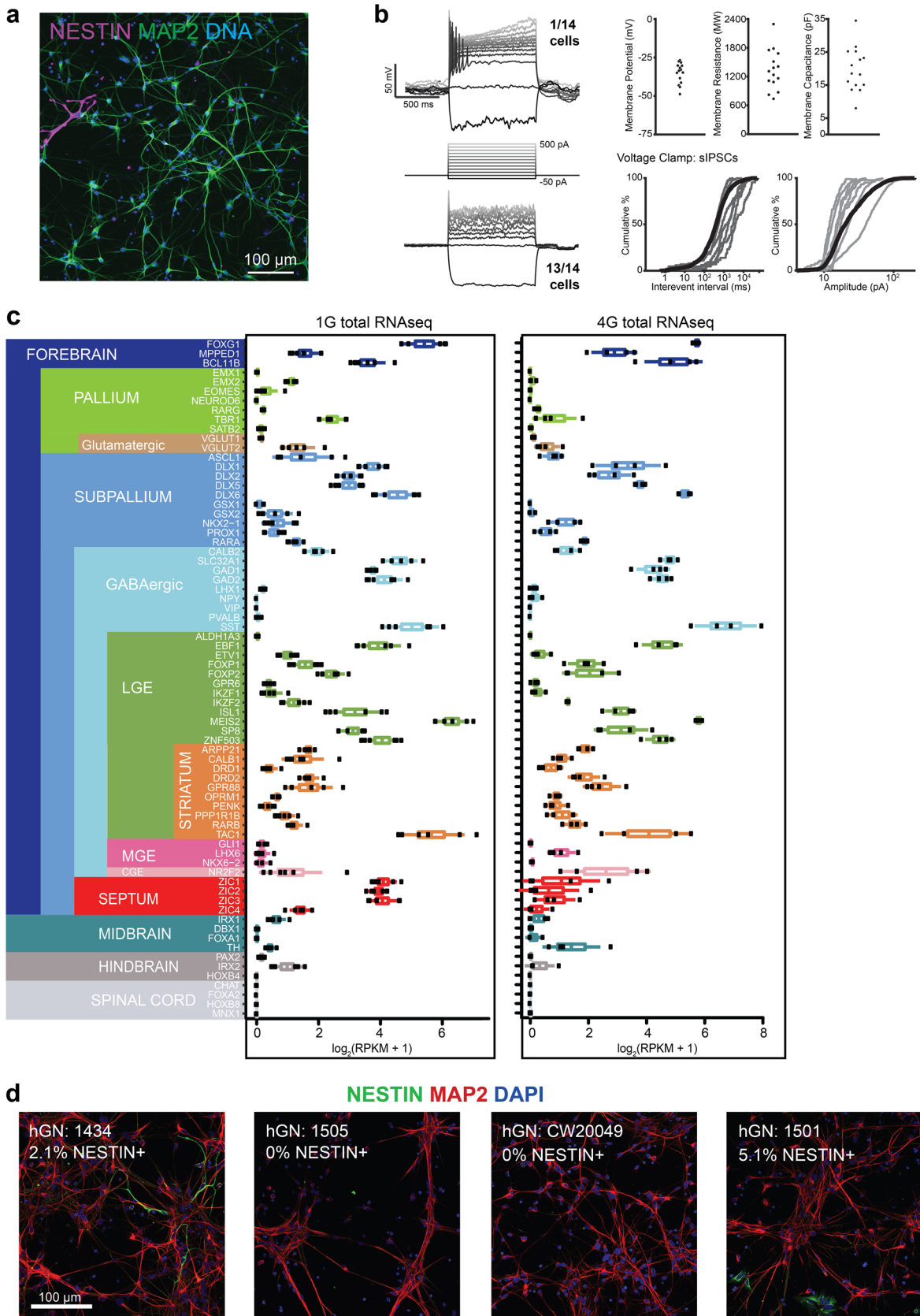
**Extended data** is available for this paper at <https://doi.org/10.1038/s41593-020-00786-1>.

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41593-020-00786-1>.

**Correspondence and requests for materials** should be addressed to G.L.B. or M.E.G.

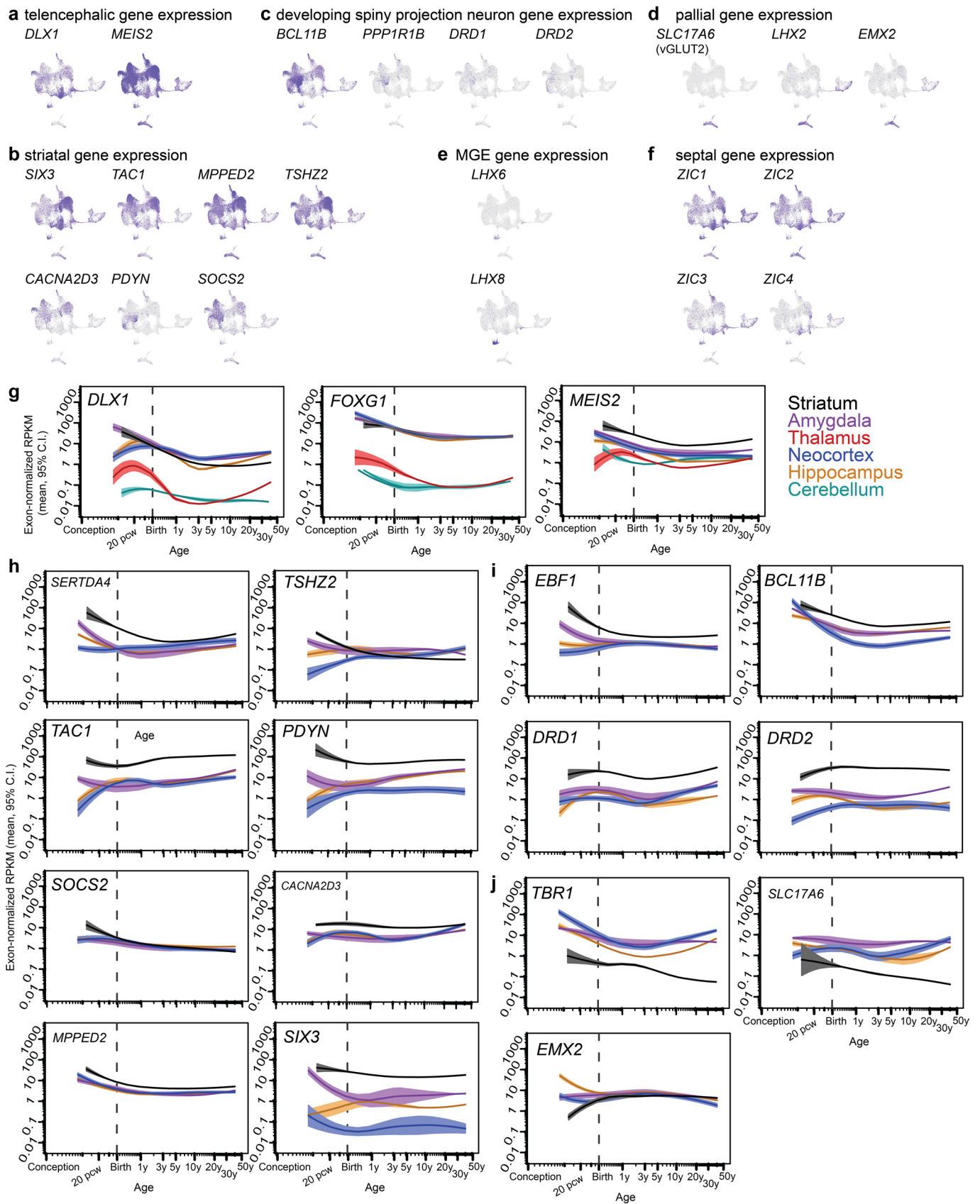
**Peer review information** *Nature Neuroscience* thanks Andrew Jaffe and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

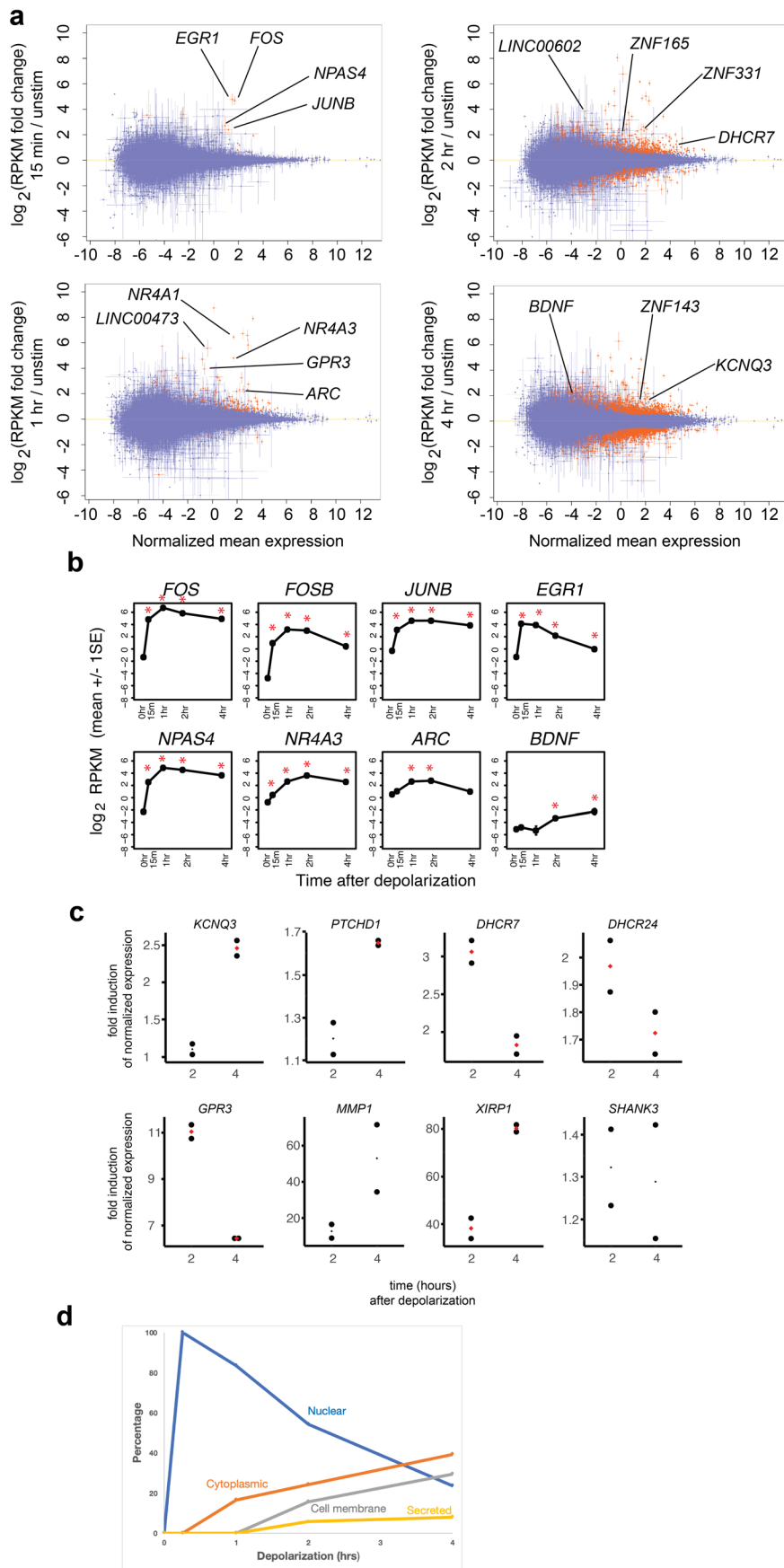


Extended Data Fig. 1 | See next page for caption.

**Extended Data Fig. 1 | The gene expression profile of hGNs is similar to that of developing human telencephalic GABAergic neurons.** **a.** Image from DIV14 1G hGN culture immunostained for neural progenitor marker protein NESTIN, post-mitotic neuronal marker protein MAP2 showing the majority of hGN cells expressing MAP2 and very few express NESTIN. Representative image of 4 independent experiments. **b.** Whole-cell patch clamp measurements of DIV14-15 hGN cells reveals immature neuronal firing and membrane properties. Frequency and amplitude of the spontaneous inhibitory post-synaptic currents (sIPSCs), which confirm that hGNs form synapses in these culture conditions. **c.** mRNA expression levels of regional and developmental marker genes in unstimulated 1G and 4G hGN cultures by total RNAseq (1G:  $n=6$ ; 4G:  $n=4$ ; both: box center = mean; box minima/maxima =  $\text{mean} \pm 1 \text{ SE}$ ; whiskers minima/maxima =  $\text{mean} \pm 1 \text{ SD}$ ). Adapted from Straccia M., et al., 2015. **d.** DIV14 4G hGN cultures immunostained for NESTIN and MAP2. The majority of hGN cells express MAP2 and very few express NESTIN in hGNs from genotypes 1434 and 1501. No NESTIN-positive cells were detected in hGN cultures from genotypes 1505 and CW20049. For 1434: representative image of 4 independent experiments. For other three genotypes: images from one independent experiment.

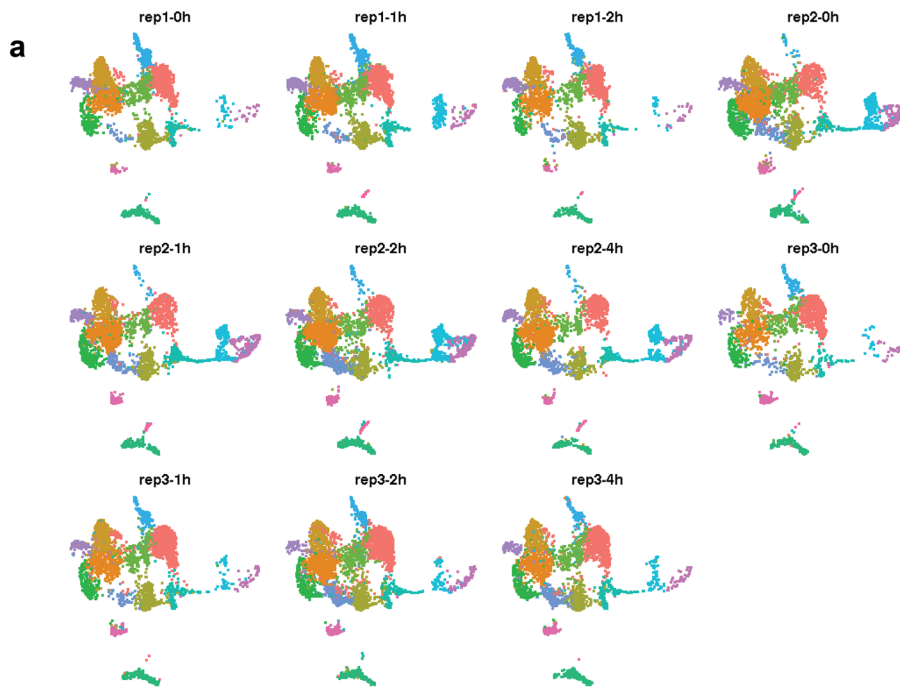


**Extended Data Fig. 2 | Single-cell RNA-sequencing of hGNs reveals developing ventral forebrain cell types. a - f.** UMAP visualizations of the entire scRNAseq data set with individual cells colored purple if expression of a given marker gene was detected. **g - j.** Brainspan human developmental brain expression data for marker genes within different brain regions. Loess curves interpolated through the means of available data at each age, with error bands on either side interpolated through values, at each age, 1.96 standard errors above or below the mean.



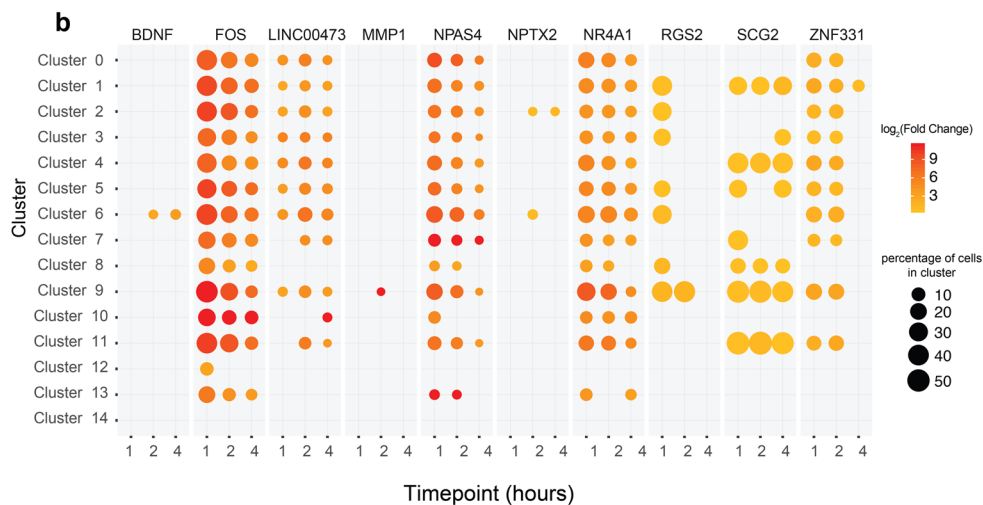
Extended Data Fig. 3 | See next page for caption.

**Extended Data Fig. 3 | Activity-dependent gene expression patterns detected by total RNAseq.** **a.** mRNA expression level changes measured by total RNAseq of 4 G hGNs at 15 minutes (n = 4), 1 hour (n = 3), 2 hours (n = 4), and 4 hours (n = 4) after membrane depolarization compared to expression in unstimulated cultures (n = 4) represented by MA-plot. Genes with a significantly different gene expression level and a minimum fold-change magnitude of 1.5 after depolarization are marked in red. Genes having inducible expression fall above  $y=0$  and example gene names are labeled. **b.** Time courses of previously known activity-inducible gene transcript levels in unstimulated and stimulated 1G hGNs. Means  $\pm$  s.e.m. Timepoints at which the transcript was significantly induced compared to unstimulated cultures are marked with an asterisk. **c.** qPCR Fold-induction of mRNA transcripts in 1G hGNs cultures of genes detected to be inducible by total and single cell RNAseq. Sensitivity limitations of qPCR prevent detection of significant differences when a gene's expression is modestly induced in a minority of the cells (for example *MMP1* and *SHANK3*). **d.** The percentage of 1G and 4G hGN inducible genes whose gene products have known subcellular localizations. Early inducible genes (15 min. - 1 hr.) have a greater percentage of nuclear localized products than gene transcripts induced at later stimulation timepoints, when the percentage of induced genes in the membrane-bound or secreted category increases.



Replicate Timepoint Contribution to Each Single Cell RNAseq Cluster

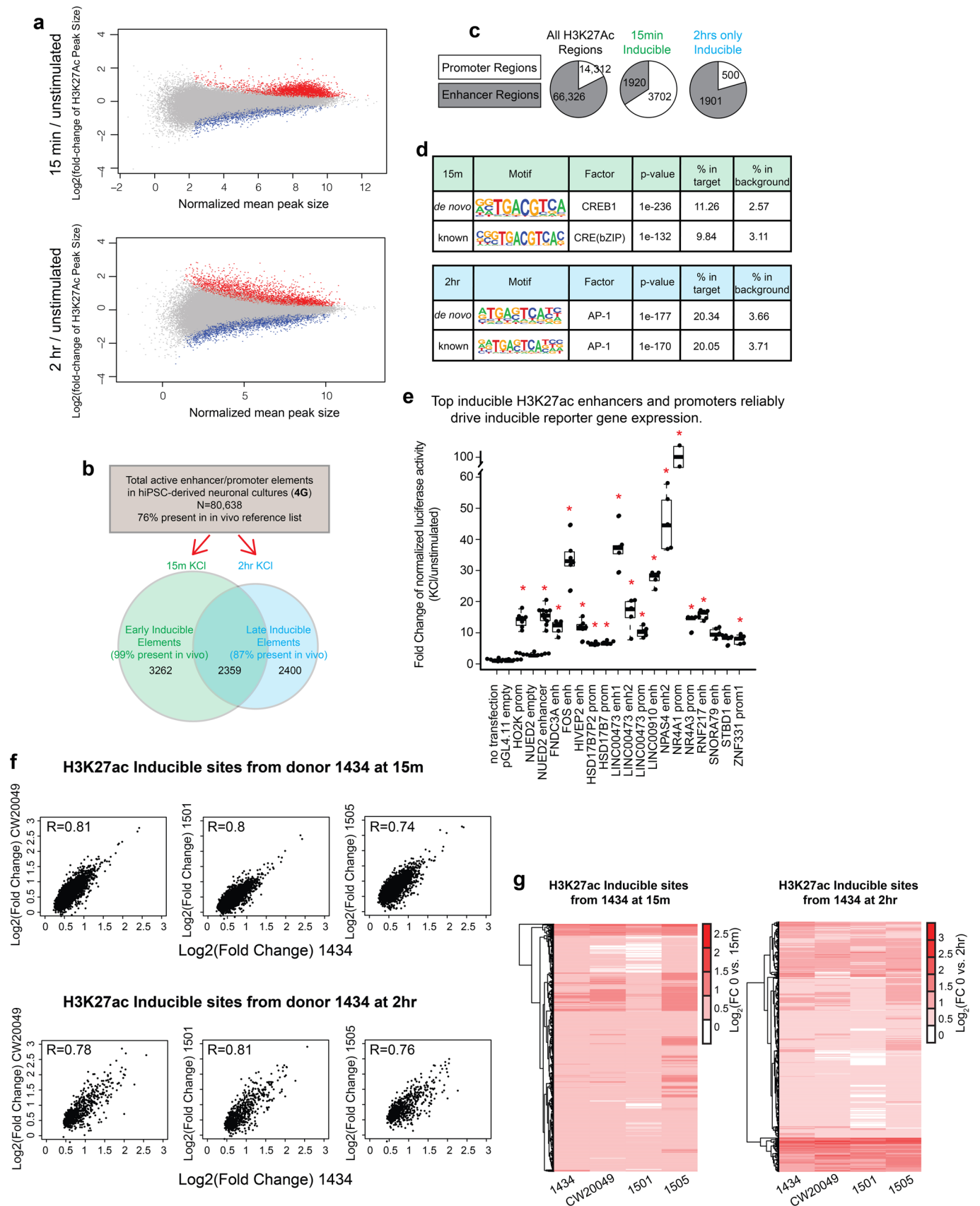
cluster #	total cells	rep1 - unstim	rep1 - 1hr KCl	rep1 - 2hr KCl	rep1 STDEV	rep2 - unstim	rep2 - 1hr KCl	rep2 - 2hr KCl	rep2 - 4hr KCl	rep2 STDEV	rep3 - unstim	rep3 - 1hr KCl	rep3 - 2hr KCl	rep3 - 4hr KCl	rep3 STDEV
0	6504	0.10	0.09	0.07	0.01	0.06	0.06	0.07	0.06	0.01	0.09	0.14	0.14	0.13	0.03
1	5654	0.07	0.07	0.05	0.01	0.15	0.13	0.15	0.11	0.02	0.05	0.07	0.09	0.07	0.02
2	3468	0.12	0.11	0.09	0.01	0.09	0.09	0.09	0.08	0.01	0.06	0.10	0.09	0.09	0.02
3	3381	0.11	0.11	0.09	0.01	0.08	0.09	0.10	0.07	0.02	0.06	0.10	0.10	0.09	0.02
4	3009	0.11	0.13	0.09	0.02	0.10	0.06	0.09	0.05	0.02	0.06	0.09	0.12	0.09	0.03
5	2559	0.07	0.07	0.06	0.01	0.11	0.12	0.14	0.11	0.01	0.06	0.08	0.10	0.09	0.02
6	1964	0.10	0.12	0.09	0.02	0.10	0.09	0.09	0.07	0.01	0.07	0.09	0.09	0.09	0.01
7	1889	0.09	0.09	0.07	0.01	0.08	0.11	0.13	0.09	0.02	0.05	0.09	0.10	0.11	0.03
8	1755	0.03	0.09	0.01	0.04	0.17	0.16	0.25	0.17	0.04	0.01	0.03	0.05	0.04	0.02
9	1560	0.20	0.22	0.18	0.02	0.01	0.02	0.03	0.02	0.00	0.06	0.10	0.08	0.09	0.01
10	1549	0.02	0.03	0.03	0.01	0.08	0.09	0.23	0.12	0.07	0.03	0.04	0.17	0.17	0.08
11	1504	0.14	0.16	0.15	0.01	0.08	0.09	0.08	0.06	0.01	0.03	0.07	0.06	0.07	0.02
12	1295	0.02	0.04	0.02	0.01	0.19	0.20	0.23	0.15	0.04	0.03	0.03	0.05	0.04	0.01
13	864	0.03	0.05	0.03	0.01	0.09	0.11	0.10	0.08	0.01	0.11	0.13	0.14	0.14	0.02
14	146	0.01	0.08	0.03	0.03	0.19	0.27	0.23	0.16	0.05	0.01	0.01	0.00	0.01	0.01



Extended Data Fig. 4 | See next page for caption.

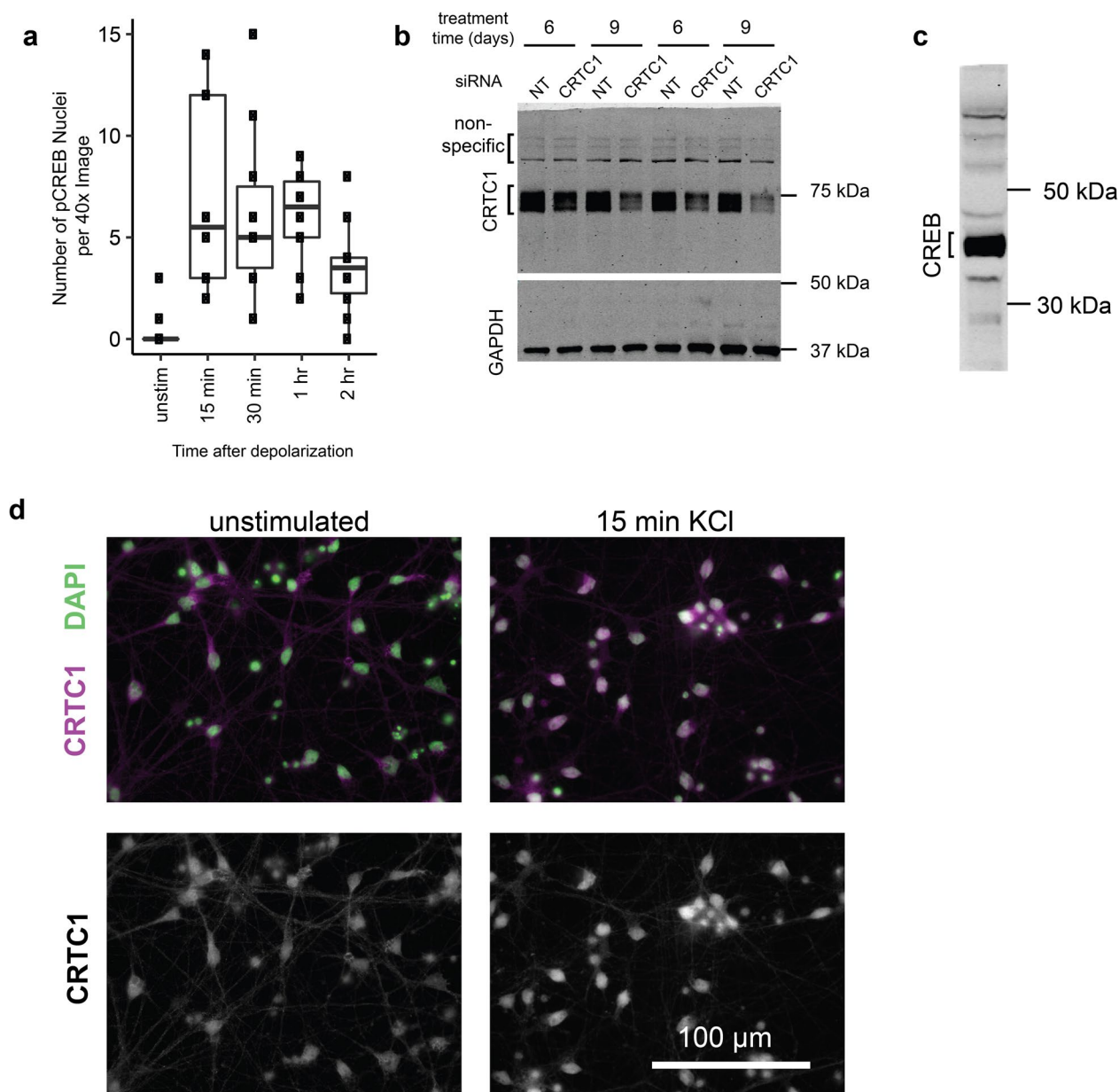
**Extended Data Fig. 4 | Activity-dependent gene expression patterns detected by single-cell RNAseq. a.** One biological replicate (rep1) of unstimulated (0 h), 1 hour, and 2 hour membrane depolarized hGN cultures, and two biological replicates (rep2 and 3) of unstimulated, 1 hour, 2 hour, and 4 hour membrane depolarized hGN cultures contributed to all cell clusters identified by single cell RNA sequencing. Replicate – timepoint contribution is calculated for each cluster in the table below **b.** Selected genes previously known to have inducible expression are also inducible in hGNs after membrane depolarization, as measured by single-cell RNAseq. This includes genes with induction in many hGN cell types (for example *FOS*, *LINC00473*, *NPAS4*) and genes showing hGN cell type-restricted induction (for example *BDNF*, *NPTX2*). *MMP1* expression was identified to be activity-inducible in this study by total RNAseq, and single-cell RNAseq shows that it is only induced in cluster 9 striatal-like neurons of hGN cultures. The magnitude of fold-change of mRNA levels in each cell cluster at different timepoints after membrane depolarization compared to unstimulated mRNA levels is indicated by dot color. The percentage of cells in each cluster from which expression was detected is represented by the size of the dot. Only timepoints with significantly induced mRNA levels (compared to unstimulated) are shown.



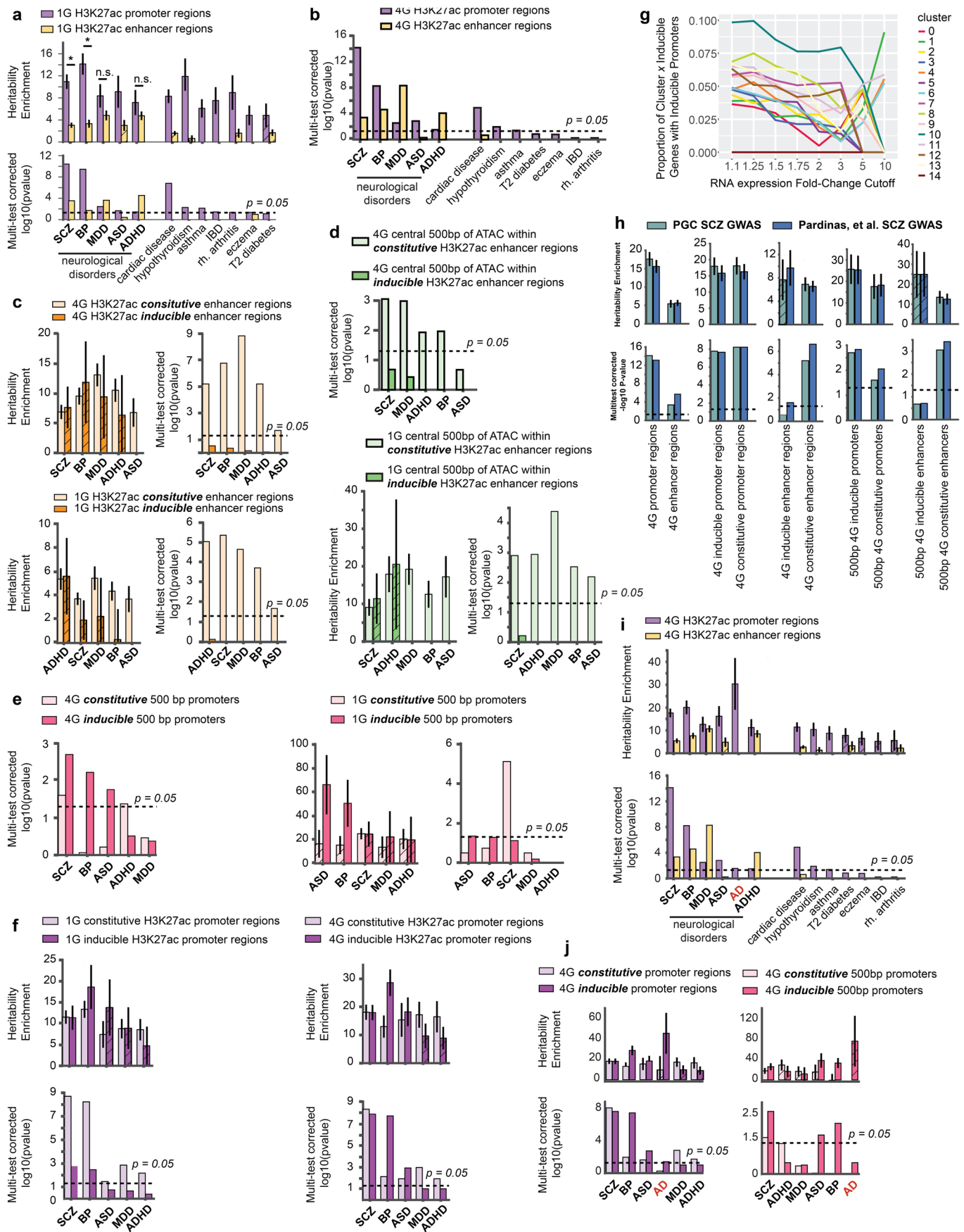


Extended Data Fig. 5 | See next page for caption.

**Extended Data Fig. 5 | hGN activity-dependent enhancers and promoters.** hGNs from four independent iPSC donors (4G) were cultured in parallel and H3K27ac ChIPseq was performed in unstimulated cultures, at 15 minutes, and two hours after membrane depolarization for each of them in duplicate (technical replicates). Using each of the four genotypes as biological replicates, differential H3K27ac enrichment analysis was performed to identify promoters and enhancers that underwent histone modification changes in response to neuronal depolarization. **a.** MA-plots representing all 4G H3K27ac peaks identified with those in red having increased significantly in H3K27ac and those in blue having decreased significantly at 15 minutes or 2 hours. **b.** The total number of reproducible 4G H3K27ac peaks that were called across the four genotypes of hGNs and the subset that were significantly inducible. As with the 1G H3K27ac peaks, 4G H3K27ac peaks are highly represented in the in vivo reference list, and most inducible peaks were inducible at either 15 min or 2 hours but not at both time points. **c.** Proportions of H3K27ac peaks that intersect a TSS (promoter regions) or do not intersect a TSS (enhancer regions). As with the 1G H3K27ac peaks, 4G 15 minute inducible H3K27ac peaks are significantly enriched for promoter regions. **d.** As with the 1G H3K27ac peaks, the CREB binding sequence is the most enriched sequence motif within 15 min 4G inducible H3K27ac regions, and the AP-1 motif is the most enriched within 2 hour 4G inducible H3K27ac regions. **e.** 1G inducible H3K27ac promoter and enhancer regions were cloned into the pGL4.11 (promoter regions) or NUED2 (enhancer regions) plasmids and transfected into embryonic mouse cortical cultures. Luciferase assays were performed to test for these sequences' ability to drive luciferase reporter gene expression in response to membrane depolarization. 15/15 constructs that show significant reporter expression induction compared to the corresponding empty control plasmid condition are indicated with an asterisk.  $n = 2-10$  independent replicates (neuronal dissection, plasmid transfection, and luminometer reading); box and whisker plot showing median and interquartile range. **f.** Pearson correlations showing similar levels of H3K27ac induction between hGNs from independent genetic donors of the 4G data set. **g.** Hierarchical clustering of H3K27ac induction across all 4G samples.



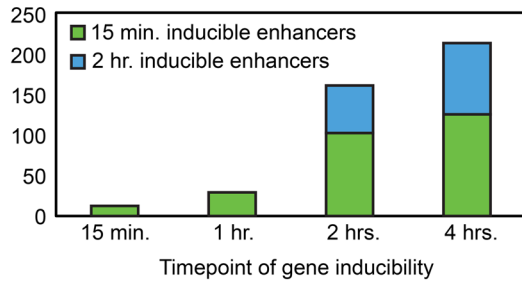
**Extended Data Fig. 6 | Detection of CREB, pCREB, and CRTCl.** **a**, 15 minutes after membrane depolarization is the timepoint at which the greatest number of hGNs immunostain positive using the phosphorylated ser133 CREB antibody used for ChIPseq. Median and interquartile range of counts from six images for each condition from a single immunostaining event of one differentiation lot of 1G neurons. **b**, Western blot for CRTCl and GAPDH from hGN cell lysates after 6 or 9 days of non-targeting (NT) or CRTCl-targeting siRNA blotted with the antibody that was used for CRTCl ChIPseq. Blot displays the two independent experiments performed for this quantification. 9 day treatments were used for siRNA treatments in this study, which resulted in an average 64.1% protein knockdown. **c**, Western blot of hGN cell lysates using the CREB antibody used for CREB ChIPseq. Image is representative of two independent experiments. **d**, Immunostaining for CRTCl in hGN cultures before (unstimulated) and 15 minutes after membrane depolarization. Membrane depolarization causes CRTCl to translocate to the nucleus. Scale bar = 100  $\mu$ m. Images are representative of 3 independent experiments.



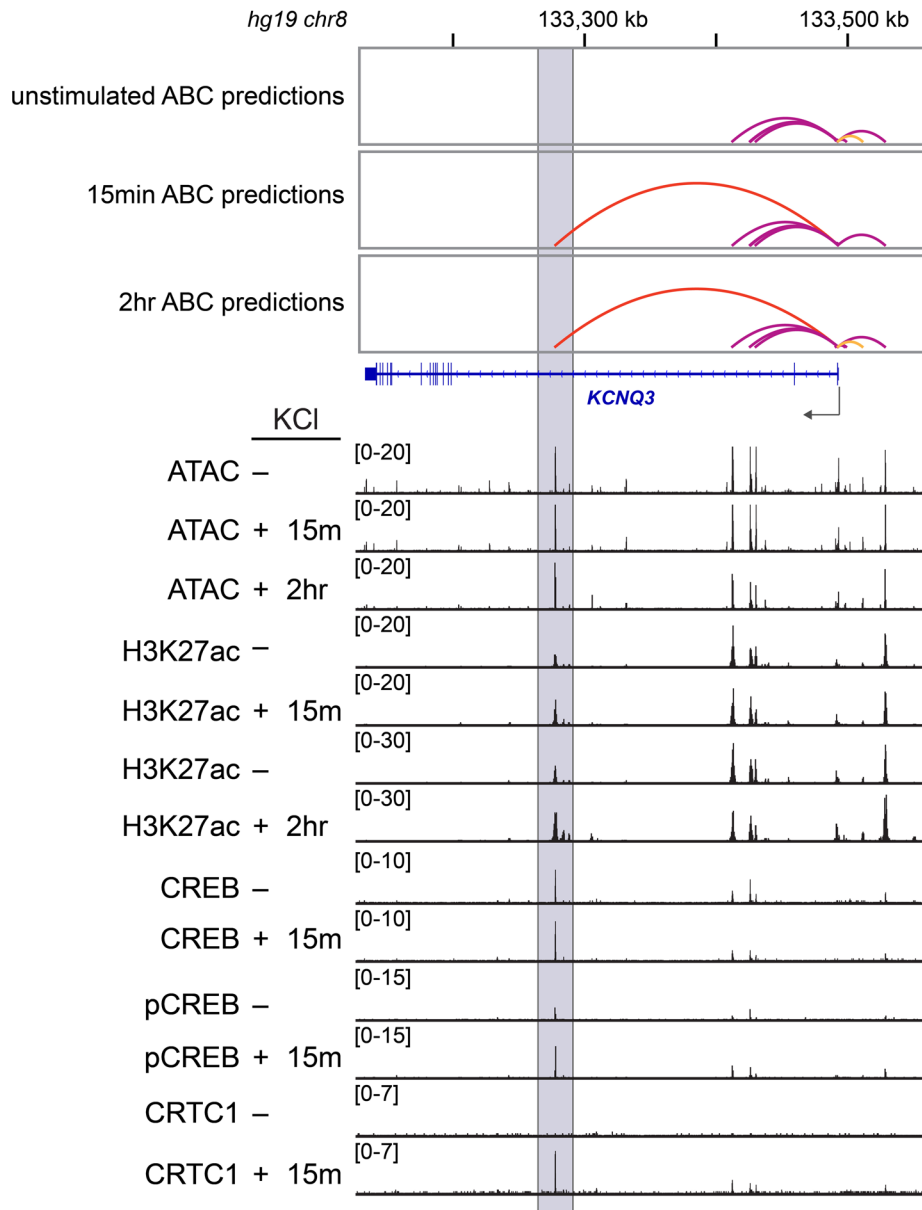
Extended Data Fig. 7 | See next page for caption.

**Extended Data Fig. 7 | Disease heritability enrichment in hGN promoter and enhancer regions.** **a.** Heritability enrichments (top) and p-values of heritability enrichment (bottom) of the 1G hGN H3K27ac promoter and enhancer regions across neurological and non-neurological disorders. **b.** p-values of heritability enrichment of 4G hGN H3K27ac promoter and enhancer regions across neurological and non-neurological disorders. **c.** Heritability enrichments (left) and p-values of heritability enrichment (right) of the inducible and constitutive H3K27ac enhancer regions across neurological diseases for both 4G (top) and 1G (bottom), showing significant enrichment for only constitutive enhancer regions. **d.** p-values of heritability enrichment of 4G (top) hGN inducible and constitutive enhancers (central 500 bp of 4G ATACseq peaks within 4G H3K27ac enhancer regions) across neurological disorders, and (bottom) 1G heritability enrichments (left) and p-values of heritability enrichment (right), showing significant enrichment for only constitutive enhancers. **e.** p-values of heritability enrichment of 4G (left) hGN inducible and constitutive promoters across neurological disorders, and 1G heritability enrichments (center) and p-values of heritability enrichment (right), showing significant ASD and BP heritability enrichment in inducible promoters but not constitutive promoters. **f.** Heritability enrichment (top) p-values of heritability enrichment (bottom) of the inducible and constitutive 1G (left) and 4G (right) H3K27ac promoter regions across neurological diseases. **g.** Proportions of each hGN scRNAseq cluster's inducible gene list that have inducible promoters by 1G H3K27ac ChIPseq. Proportions are calculated at multiple gene expression fold-change cut-off thresholds. Cluster 10 (striatal spiny projection neuron-like cells) inducible genes have the highest proportion of SFARI ASD-associated genes, regardless of gene expression fold-change cut-off threshold. **h.** LDSC heritability enrichments were calculated using summary statistics from two schizophrenia (SCZ) GWASs: Psychiatric Genomics Consortium (PGC), 2014 and Pardiñas, *et al.*, 2018. The latter incorporates 40.5 K cases compared to 37 K in the 2014 PGC study and uses fewer controls (63 K vs 113 K). **i.** LDSC heritability enrichment calculations were repeated for all neurological disorders in parallel with Alzheimer's Disease (AD), using the summary statistics from Lambert, J., Ibrahim-Verbaas, C., Harold, D. *et al.*, Nature Genetics, 2013. Please note that the GWASs for the other neurological disorders included in our analyses were uniformly processed by the PGC whereas the AD GWAS was processed differently, which may contribute to the relatively large error bars associated with the AD analysis. All LDSC Plots are color coded according to Fig. 6a. Each heritability enrichment value is provided in bar plot (+/- std. error). p-values of heritability enrichment are multi-test corrected and heritability enrichment bars with  $p =$  or  $> 0.05$  are cross-hatched. Exact p-values are provided in Supplementary Data 14.

**a** Number of inducible genes (at each time point) predicted by ABC model to associate with inducible enhancers



**b**



Extended Data Fig. 8 | See next page for caption.

**Extended Data Fig. 8 | ABC model predictions of inducible enhancer-gene associations. a.** The number of inducible genes and H3K27ac enhancer regions, as defined by this study (1G), predicted to interact by the Activity-by-Contact (ABC) model described in Fulco et al., Nat. Genetics, 2019. **b.** An example of one ASD-associated and neuronal activity-dependent gene locus, *KCNQ3*, as visualized in the IGV genome browser, with chromatin data presented in this resource (1G tracks below) and ABC model predictions of enhancer associations with the *KCNQ3* promoter (arcs above). These include enhancer associations that remain constitutively predicted across all stimulation conditions (purple), those that are potentially transient but not necessarily activity-dependent (yellow), and those that are likely to be activity-dependent (red). In this case, at 15 minutes and 2 hours after neuronal depolarization, the *KCNQ3* promoter is predicted to interact with an activity-dependent intronic enhancer (shaded), which increases in H3K27ac enrichment and becomes associated with the activated CREB-complex after depolarization. See Supplementary Data 13 for the full list of ABC model predicted enhancer-gene associations.

## Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided  
*Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g.  $F$ ,  $t$ ,  $r$ ) with confidence intervals, effect sizes, degrees of freedom and  $P$  value noted  
*Give  $P$  values as exact values whenever suitable.*
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's  $d$ , Pearson's  $r$ ), indicating how they were calculated

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

**Data collection** All software packages used in this study are previously published and open-source: Electrophysiology data Data were saved with a custom version of ScanImage written in Matlab (Mathworks; [https://github.com/bernardosabatini/SabalabSoftware\\_Nov2009](https://github.com/bernardosabatini/SabalabSoftware_Nov2009)).

**Data analysis** All software packages used in this study are previously published and commercially available or open source: Microsoft Excel for Mac version 16; Electrophysiology data was analyzed on Matlab and spontaneous EPSCs and IPSCs were analyzed using a modified Peaker Analysis Toolbox (Andrew Penn (2020). Peaker Analysis Toolbox (<https://www.github.com/acp29/Peaker>), GitHub. Retrieved March 27, 2020); Activity-by-Contact (ABC) model (<https://github.com/broadinstitute/ABC-Enhancer-Gene-Prediction>), custom software MAPtoFeatures for total RNAseq mapping and EdgeR for differential gene expression analysis was previously reported (6); Single cell RNAsequencing analysis used R (version 3.4.1), and Bowtie (version 1.2.2)(52) and analyzed using the Seurat package (Version 2.3.4 and 3.0.0) (53,54) and Monocle (Version 2.6.4) (55); ChIPseq and ATACseq analysis performed using CutAdapt V2.5, Burrows-Wheeler Aligner, Samtools, Bedtools, MACS, MACS2 V2.1.1, DESeq2, HOMER (version 4.9), and EdgeR. Custom Fiji macros for immunostaining image quantification, and custom code for generating mean Loess curves from brainsapn data (as previously reported in 6) are available upon request.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.



## Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All raw data (fastq files) and final processed data (gene expression tables, ChIPseq bigwig and bed files, etc...) are available on GEO: GSE136656. Main figures have raw data associated with them: Figure 1 – total RNAseq; Figure 2 – total RNAseq; Figure 3 – single cell RNAseq; Figure 4 – H3K27ac and H3K4me3 ChIPseq; Figure 5 – CREB, pCREB, CRTCL1, ChIPseq. Figure 5 is associated with LDSC disease heritability enrichment calculations (Supp. Data. 13). There are no restrictions on these data's availability. Publicly available data sets: ENSEMBL's Comparative Genomics Tool (<https://www.ensembl.org/info/genome/compara/index.html>), GenTree (<http://gentree.ioz.ac.cn/index.php>), Homologene (<https://www.ncbi.nlm.nih.gov/homologene>), BrainSpan atlas (<http://www.brainspan.org/static/download.html>), UCSC genome browser (<https://genome.ucsc.edu/>), SFARI ASD-associated genes 11-21-2018 release (<https://gene-archive.sfari.org/>), Schizophrenia Working Group of the Psychiatric Genomics Consortium (<https://www.med.unc.edu/pgc/>).

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences       Behavioural & social sciences       Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No statistical methods were used to pre-determine sample sizes but our sample sizes are similar to those reported in previous publications (18, 39). For ChIPseq experiments, a minimum of 2 biological replicates (independent production lots of neuronal cultures) were used for each condition, which Landt et al. (Genome Research, 2012) describe as being sufficient for binding site discovery and which resulted in the discovery of typical numbers of protein enrichment sites. For all other experiments, a minimum of two biological replicates were used, but typically 3-4 whenever possible, except for GFAP immunostaining for which one experiment was obtained, and results were corroborated by single cell RNAsequencing and total RNA sequencing. This was deemed sufficient due to the high reproducibility of the cultures as measured by RNAseq and immunostaining.
Data exclusions	Single-cell RNA sequencing data exclusions were pre-established: Cells were filtered by excluding cells with UMI counts less than 500, UMI counts greater than 10000, mitochondrial gene expression greater than 30%, or ribosomal gene expression greater than 15%. For all other assays, it was pre-established that experiments were to be excluded if positive controls failed.
Replication	Four independent replicates of ATACseq experiments were performed and one replicate failed (did not produce sequencing libraries). The IDR pipeline was applied to ATACseq data to verify reproducibility of the three remaining experimental data sets. Reproducible CREB Complex ChIPseq peaks were defined by ChIP using three different antibodies recognizing epitopes in the same protein complex. RNAseq was performed multiple times (3-6 biological replicates) with highly correlated results. All attempts to reproduce data were successful unless the neuronal cultures were visibly unhealthy or positive controls for gene induction failed, in which case the data was excluded.
Randomization	Randomization is not relevant to our study as all comparative analyses were performed using an unbiased computer software package.
Blinding	Blinding is not relevant to our study as all comparative analyses were performed using an unbiased computer software package.

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input type="checkbox"/>	<input checked="" type="checkbox"/> Animals and other organisms
<input type="checkbox"/>	<input checked="" type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

### Methods

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

## Antibodies

Antibodies used	anti-NESTIN (mouse, 1:500, R&D, MAB1259), anti-GFAP (rabbit, 1:500, Dako, Z033429-2), anti-MAP2 (chicken, 1:1,000, Lifespan Biosciences, LS-C61805), anti-neurofilament [SMI-312] (1:1000, Abcam, ab24574), anti-CREB (1:500, EMD Millipore #06-863), anti-CRTC1 (1:500, Bethyl Laboratories 300-769), anti-Ser133 pCREB (1:500, Cell Signaling 87G3 #9198S), anti-H3K27Ac (Abcam ab4729), anti-H3K4me3 (Millipore 07-473), AlexaFluor conjugated secondary antibodies for immunostaining and imaging cultured cells (Invitrogen A-31570, A-21202, A-21206, A-31572, A32795, A32933), Dylight 800nm Conjugated secondary antibodies for western blot (Rockland Immunochemical 75934-472).
Validation	anti-NESTIN R&D, MAB1259 for immunostaining– manufacturer validation: Nestin was detected in immersion fixed human fetal neural progenitor cells, anti-GFAP, 1:500, Dako, Z033429-2- manufacturer validation: In indirect ELISA, the antibody shows no reaction with human plasma and cow serum. GFAP shows 90-95% homology between species (5), and as demonstrated by immunocytochemistry, the antibody reacts strongly with human GFAP., anti-MAP2 Lifespan Biosciences, LS-C61805 – manufacturer validation: western blot shows specific for the ~280kDa MAP2 protein. In-study validation: saw expected, specific immunocytochemistry staining pattern; anti-neurofilament, SMI-312, Abcam, ab24574 – manufacturer validation: Western blot shows specificity for predicted molecular weight: 61, 102, 112 kDa. In-study validation: saw expected, specific immunocytochemistry staining pattern; anti-CREB, EMD Millipore #06-863 – manufacturer validation: tested for specificity in Chromatin Immunoprecipitation (ChIP), ChIP-seq, Electrophoretic Mobility Shift Assay (EMSA), Immunoprecipitation, and Western Blotting. In-study validation: saw expected, specific immunocytochemistry and western blot staining pattern; anti-CRTC1, Bethyl Laboratories 300-769 – manufacturer validation: western blot shows specific for the 70-100 kDa CRTC1 protein bands. In-study validation: saw expected, specific immunocytochemistry and western blot staining pattern, and significant decrease of mRNA and protein signal detection upon siRNA treatment ; anti-Ser133 pCREB, Cell Signaling 87G3 #9198S – manufacturer validation: based on western blot, stimulated mouse brain samples, and ChIPseq this antibody detects endogenous levels of CREB only when phosphorylated at serine 133. The antibody also detects the phosphorylated form of the CREB-related protein, ATF-1. In-study validation: saw expected, specific immunocytochemistry and western blot staining pattern; anti-H3K27Ac, Abcam, ab4729 – manufacturer validation: ChIP-qPCR, western blot, and immunostaining showed expected enrichments and staining patterns; anti-H3K4me3 (Millipore 07-473) – manufacturer validation: ChIPseq and western blot showed expected enrichments and staining patterns.

## Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s)	hGNs, hiPSC-derived GABAergic neurons, Cellular Dynamics : <a href="https://fujifilmcdi.com/products-services/icell-products/icell-gabaneurons/">https://fujifilmcdi.com/products-services/icell-products/icell-gabaneurons/</a>
Authentication	Confirmed nuclear immunoreactivity for human nuclear antigen.
Mycoplasma contamination	Manufacturer validation: Each production lot is tested for sterility and mycoplasma. In-study validation: cultures were tested with LookOut Mycoplasma PCR Detection Kit, Sigma MP0035, and always tested negative (no PCR product for mycoplasma DNA).
Commonly misidentified lines (See <a href="#">ICLAC</a> register)	No commonly misidentified cell lines were used in the study.

## Animals and other organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research

Laboratory animals	Pregnant E15 C57BL/6 female mice (Charles River) were used to derive primary mouse embryonic cortical cultures employed in luciferase assay experiments. Mice were used on arrival and not housed. Primary mouse embryonic cortical cultures were derived from over 10 does for these studies.
Wild animals	The study did not involve wild animals.
Field-collected samples	The study did not involve samples collected from the field.
Ethics oversight	The use of animals was approved by the Animal Care and Use Committee of Harvard Medical School.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics	Two human fetal cortical samples were used: FB08 (19 gestational weeks) and FB25 (23 gestational weeks). Gestational ages were determined using fetal foot length.
Recruitment	Fetal brain tissue was received after release from clinical pathology, with a maximum post-mortem interval of 4 h. Cases with known anomalies were excluded.

## Ethics oversight

Research performed on samples of human origin was conducted according to protocols approved by the institutional review boards of Beth Israel Deaconess Medical Center, Boston Children's Hospital and Harvard University Faculty of Medicine.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## ChIP-seq

## Data deposition

- Confirm that both raw and final processed data have been deposited in a public database such as [GEO](#).
- Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

## Data access links

*May remain private before publication.*

The sequencing data that support the findings of this study are available on GEO accession # GSE136656. Publicly available data sets: ENSEMBL's Comparative Genomics Tool (<https://www.ensembl.org/info/genome/compara/index.html>), GenTree (<http://gentree.ioz.ac.cn/index.php>), Homologene (<https://www.ncbi.nlm.nih.gov/homologene>), BrainSpan atlas (<http://www.brainspan.org/static/download.html>), UCSC genome browser (<https://genome.ucsc.edu/>), SFARI ASD-associated genes 11-21-2018 release (<https://gene-archive.sfari.org/>), Schizophrenia Working Group of the Psychiatric Genomics Consortium (<https://www.med.unc.edu/pgc/>).

## Files in database submission

All raw and processed sequencing data files are available on GEO accession # GSE136656.

## Genome browser session

(e.g. [UCSC](#))

N/A

## Methodology

## Replicates

A minimum of 2 biological replicates (using independent production lots of neuronal cultures) were used for each condition, which Landt et al. (Genome Research, 2012) describe as being sufficient for binding site discovery, and which resulted in the discovery of typical numbers of protein binding sites.

## Sequencing depth

ChIPseq libraries were sequenced to a minimum depth of 20M reads per experiment, with a minimum of ~50% of reads uniquely mapped to the genome. Sequencing was performed via single-end 75bp sequencing.

## Antibodies

anti-CREB (1:500, EMD Millipore #06-863), anti-CRTC1 (1:500, Bethyl Laboratories 300-769), anti-Ser133 pCREB (1:500, Cell Signaling 87G3 #9198S), anti-H3K27Ac (Abcam ab4729), anti-H3K4me3 (Millipore 07-473).

## Peak calling parameters

Mapping of ChIP-seq reads was performed by Burrows-Wheeler Aligner (Li and Durbin, 2009) with default settings by truncating all reads to 70bp and aligning to hg19 assembly of the human genome. Peak calling was performed by MACS software (Model-based Analysis of ChIP-seq). We used input controls as control files for peak calling. For transcription factor ChIP-seq the default parameters were used, and for H3K27Ac and H3K4Me3 the following parameters were used: --no model --shiftsize = 150.

## Data quality

As per MACS2 (Feng J et al, 2012) methods, peak quality is ensured by comparing regions of the genome to control that have enrichment that is at least 10x more than expected but smaller than 30x as default. Quality was also ensured by checking extremes of peaks (smallest and largest peaks called) by visualizing on UCSC genome browser compared to control. Finally, quality was also ensured by performing several independent biological replicates for each experiment and defining final peaks as those that have been called in at least two biological replicates.

## Software

BWA (Burrows-Wheeler Aligner) or Bowtie V1.2.2 was used for mapping of ChIP-seq data to hg19. MACS or MACS2 (Model-based Analysis of ChIP-seq) V2.1.1 was used for peak calling, and EdgeR in R (version 3.4.1) or DESeq2 for differential peak enrichment. HOMER (Hypergeometric Optimization of Motif EnRichment) was used for ChIP-seq analysis, including determining the number of counts for each peak.