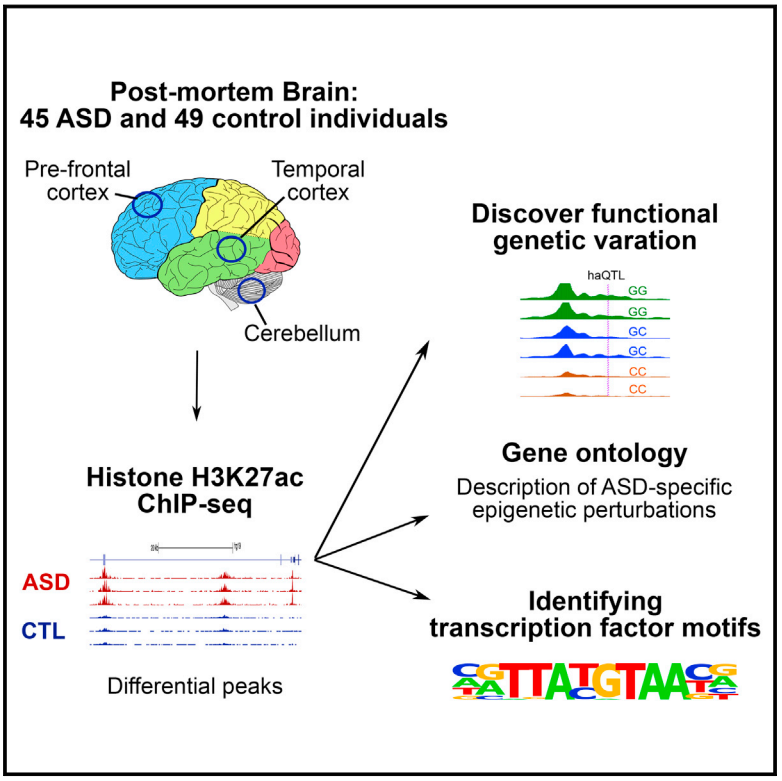
Resource

Cell

# Histone Acetylome-wide Association Study of Autism Spectrum Disorder

## Graphical Abstract




## Highlights

- Histone acetylation population study of ASD and control brain samples
- Discovery of ASD-specific epigenetic signature
- Similar epigenomic aberrations in syndromic and idiopathic ASD
- Thousands of QTLs are discovered

## Authors

Wenjie Sun, Jeremie Poschmann, Ricardo Cruz-Herrera del Rosario, ..., Jonathan Mill, Daniel H. Geschwind, Shyam Prabhakar

## Correspondence

dhg@mednet.ucla.edu (D.H.G.), prabhakars@gis.a-star.edu.sg (S.P.)

## In Brief

As part of the IHEC consortium, this study characterized histone acetylation patterns in brain samples from patients with autism spectrum disorder (ASD), uncovering a distinct epigenetic signature in ASD and providing a rich resource for future molecular analyses of ASD patients. Explore the Cell Press IHEC web portal at http://www.cell.com/consortium/IHEC.

Sun et al., 2016, Cell *167*, 1385–1397
November 17, 2016 
http://dx.doi.org/10.1016/j.cell.2016.10.031

CellPress

Resource

Cell

# Histone Acetylome-wide Association Study of Autism Spectrum Disorder

Wenjie Sun,[1,6] Jeremie Poschmann,[1,6] Ricardo Cruz-Herrera del Rosario,[2] Neelroop N. Parikshak,[3] Hajira Shreen Hajan,[1] Vibhor Kumar,[1] Ramalakshmi Ramasamy,[1] T. Grant Belgard,[3] Bavani Elanggovan,[1] Chloe Chung Yi Wong,[4] Jonathan Mill,[4,5] Daniel H. Geschwind,[3,*] and Shyam Prabhakar[1,7,*]

[1]Computational and Systems Biology, Genome Institute of Singapore, Singapore 138672, Singapore

[2]Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA

[3]Program in Neurogenetics, Department of Neurology, Center for Autism Research and Treatment, Semel Institute, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA 90095, USA

[4]Institute of Psychiatry, Psychology & Neuroscience, King's College London, London SE5 8AF, UK

[5]University of Exeter Medical School, University of Exeter, Exeter EX2 5DW, UK

[6]Co-first author

[7]Lead Contact

*Correspondence: dhg@mednet.ucla.edu (D.H.G.), prabhakars@gis.a-star.edu.sg (S.P.)

http://dx.doi.org/10.1016/j.cell.2016.10.031

## SUMMARY

**The association of histone modification changes with autism spectrum disorder (ASD) has not been systematically examined. We conducted a histone acetylome-wide association study (HAWAS) by performing H3K27ac chromatin immunoprecipitation sequencing (ChIP-seq) on 257 postmortem samples from ASD and matched control brains. Despite etiological heterogeneity, ≥68% of syndromic and idiopathic ASD cases shared a common acetylome signature at >5,000 *cis*-regulatory elements in prefrontal and temporal cortex. Similarly, multiple genes associated with rare genetic mutations in ASD showed common "epimutations." Acetylome aberrations in ASD were not attributable to genetic differentiation at *cis*-SNPs and highlighted genes involved in synaptic transmission, ion transport, epilepsy, behavioral abnormality, chemokinesis, histone deacetylation, and immunity. By correlating histone acetylation with genotype, we discovered >2,000 histone acetylation quantitative trait loci (haQTLs) in human brain regions, including four candidate causal variants for psychiatric diseases. Due to the relative stability of histone modifications postmortem, we anticipate that the HAWAS approach will be applicable to multiple diseases.**

## INTRODUCTION

Autism spectrum disorder (ASD) is a collection of neuro-developmental disorders characterized by deficits in social interaction and social communication, along with restricted and repetitive behavior patterns. DNA sequence variation affecting the function of several hundred genes has been implicated in the etiology of ASD at various levels of significance (Abrahams et al., 2013; de la Torre-Ubieta et al., 2016). These genetic changes include copy number variation, chromosomal rearrangements, and also rare single-base pair mutations in coding genes (Devlin and Scherer, 2012; de la Torre-Ubieta et al., 2016). In addition, environmental risk factors such as chemical toxins and maternal infection during gestation are thought to play a role in the occurrence of ASD (Grabrucker, 2013; Matelski and Van de Water, 2016). Thus, the etiology of ASD is complex and multifactorial, including both genetic and environmental components.

Large-scale gene expression studies on ASD postmortem brain regions, as well as the prenatal and postnatal developing brain, suggest that alterations in common molecular pathways such as transcriptional regulation, synaptic function, and immunity may occur during brain development and contribute to ASD pathophysiology (Voineagu et al., 2011; Parikshak et al., 2013; Willsey et al., 2013). How the genetic and environmental heterogeneity is translated into shared molecular pathways is not well understood. In addition, most of the efforts have so far focused on gene expression and genetic changes in coding regions. Many of these coding variants are extremely rare and account only for a small proportion of ASD cases (Stein et al., 2013; Geschwind and State, 2015). Therefore, it has been proposed that epigenetic changes caused by non-coding genetic variation or by environmental insults might contribute to ASD (Kubota et al., 2012). An attempt to characterize epigenomic changes in patients is thus likely to provide novel insights into the etiology of ASD (Akbarian et al., 2015). Thus far, epigenome-wide association studies (EWAS) of psychiatric and other diseases have mostly focused on DNA methylation (Kubota et al., 2012; Mill and Heijmans, 2013; Lunnon et al., 2014; Loke et al., 2015; Montano et al., 2016). In contrast, little is known about histone modification changes in psychiatric disease (Shulha et al., 2012) or the genetics of population variation in histone modification (del Rosario et al., 2015; Grubert et al., 2015; van de Geijn et al., 2015).

To address this lack of knowledge, we globally interrogated the histone acetylomes of enhancers in a large cohort of ASD and control samples by analyzing tissue from three brain regions postmortem: prefrontal cortex (PFC), temporal cortex (TC), and cerebellum (CB). These brain regions were chosen due to the role of frontal and temporal lobe in social cognition and the

cerebellar dysfunction observed in some animal models of ASD (Abrahams and Geschwind, 2010; de la Torre-Ubieta et al., 2016). H3K27ac was selected as the representative acetylation mark because it highlights active enhancers and promoters (Wang et al., 2008; Heintzman et al., 2009; Creyghton et al., 2010) and is also correlated with gene expression and transcription factor binding (Kumar et al., 2013). We used the data to define aberrantly acetylated enhancer and promoters in ASD brain and thereby characterize commonly altered pathways, upstream regulatory factors, and developmental dynamics of affected loci. In addition, we used chromatin immunoprecipitation sequencing (ChIP-seq) reads to call SNPs within enhancers and promoters. We then used the genotype-independent signal correlation and imbalance (G-SCI) test (del Rosario et al., 2015) to detect haQTLs in regulatory regions and assessed their relationship to known psychiatric disease-associated variants. This dataset from post-mortem human brains will provide a rich resource for future molecular analyses of ASD and serve as proof of principle for the HAWAS approach, which can be applied to a wide variety of human diseases.

## RESULTS

### Data Generation, Processing, and Differential Acetylation Analysis

In total, we performed 257 H3K27ac ChIP-seq assays on tissue samples from PFC, TC, and CB, in 94 individuals aged 10 years and above (45 ASD, 49 control; Figure 1A; Table S1). Forty-eight ChIP-seq profiles were discarded based on data quality, resulting in a final acetylome set comprising 209 profiles (STAR Methods; Table S2): 81 from PFC (41 ASD, 40 control), 66 from TC (30 ASD, 36 control), and 62 from CB (31 ASD, 31 control). We used DFilter (Kumar et al., 2013) to call peaks in each of the 209 ChIP-seq profiles and then defined two consensus peak sets: 56,503 cortical peaks (union of PFC and TC) and 38,069 CB peaks. Each consensus ChIP-seq peak defined a region of focal histone acetylation and thus represented a putative promoter or enhancer region.

The heights (aggregate read counts) of consensus peaks represent acetylation levels of cortical and cerebellar regulatory regions in each sample. We normalized these peak heights for GC content (Figure S1) and distributional skews and then controlled for confounders by regressing out multiple biological covariates such as age, sex, and proportion of neurons and also multiple technical covariates (STAR Methods; Figure S2). Corrected peak heights were used to define an initial set of differentially acetylated (DA) loci between ASD and control in each brain region (Wilcoxon rank sum test; $Q \leq 0.05$, fold change $\geq 1.3$). Based on acetylation levels at these DA loci, we measured inter-sample divergence and found that a small number of atypical ASD samples showed greater similarity to control and vice versa (Figure S3). This was not surprising, given the tremendous etiological heterogeneity of ASD and previous findings from transcriptomic analysis (Voineagu et al., 2011). Nevertheless, in the majority of cases, ASD acetylomes resembled each other more than they resembled control and vice versa (Figure S3).

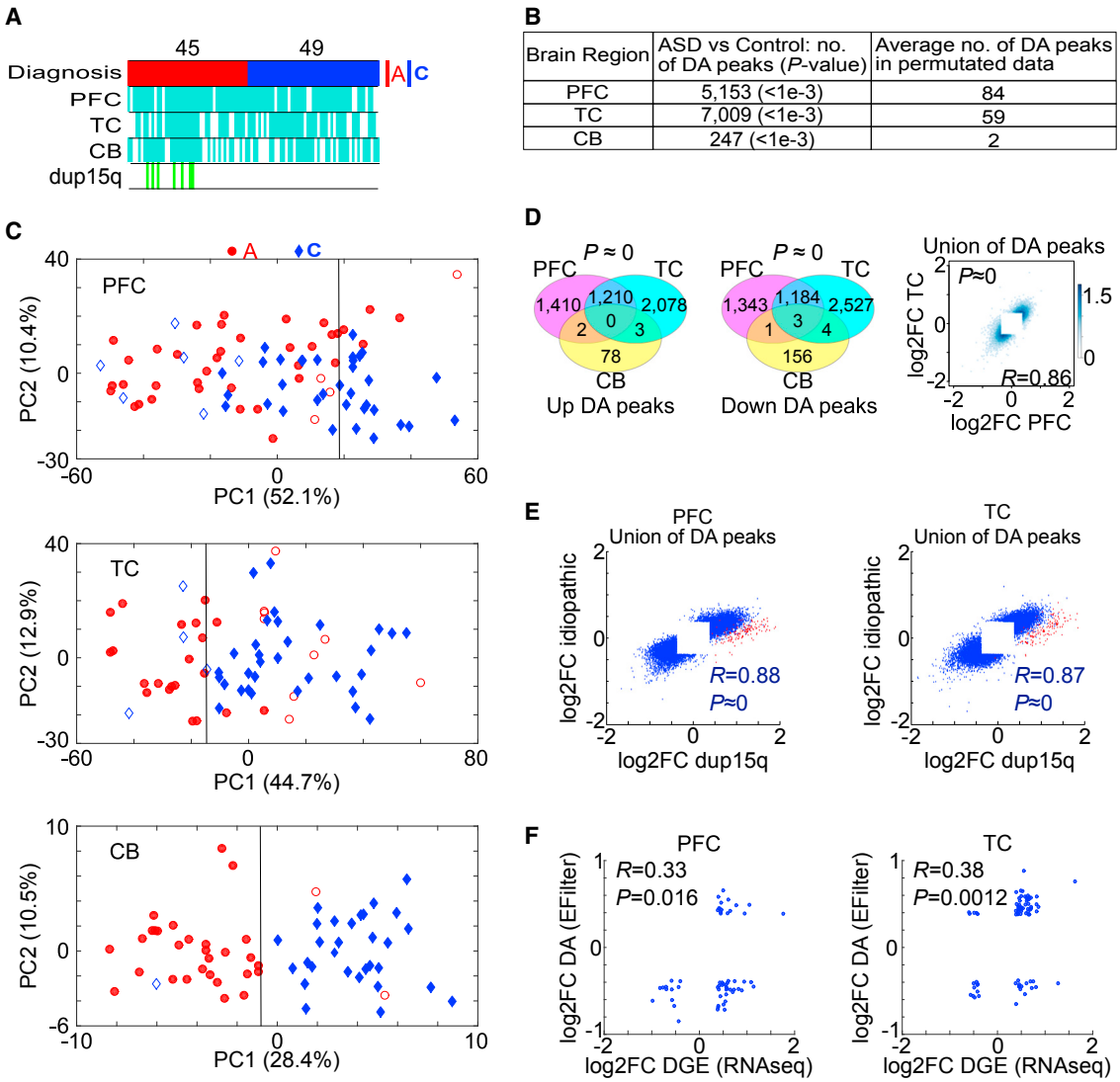
In order to define the core set of chromatin aberrations in typical ASD cases, we employed a systematic mathematical criterion to exclude atypical samples (STAR Methods). Because the number of excluded samples was relatively small (5%–20% of total cases and 14%–20% of dup15q), acetylation fold changes were not substantially altered (PFC: R = 0.94, TC: R = 0.90, CB: R = 0.98; Figure S4). We then used the remaining (typical) samples to define the final set of DA peaks for each brain region. Strikingly, we detected 5,153 DA peaks in PFC and 7,009 in TC, indicating widespread, systematic histone acetylation changes in ASD cerebral cortex (Figures 1B and 1C). In contrast, only 247 DA peaks were detected in CB. The limited molecular pathology of ASD cerebellum is consistent with results from transcriptomic studies (Voineagu et al., 2011; Parikshak et al., 2016).

To evaluate the likelihood of false-positive DA peaks, we repeated the entire procedure (initial DA peaks, discarding atypical samples, final DA peaks) after randomly permuting ASD and control labels. At the same false discovery rate (FDR) threshold ($Q \leq 0.05$), permuted datasets generated fewer than 100 DA peaks, on average. Moreover, after 1,000 tries, none of the permutated datasets generated as many DA peaks as the true dataset (Figure 1B). Thus, the chromatin changes we detected in ASD samples were far in excess of what would be expected by chance.

To further characterize the overall consistency of the DA peak sets, we examined their overlaps. Over 45% of ASD-upregulated regions in PFC overlapped ASD-Up peaks in TC ($p \approx 0$; Figure 1D). The same was true of ASD-downregulated peaks. Moreover, the ASD-versus-control acetylation fold change was highly correlated between PFC and TC ($R = 0.86$; $p \approx 0$). Thus, the chromatin dysregulation signature of ASD was highly consistent between the two cortical regions. Cerebellar DA peaks, on the other hand, showed only ~5% overlap with same-direction cortical DA peaks.

Of the 45 cases, 7 had a monogenic form of ASD, duplication 15q syndrome (dup15q; Figure 1A), while the others had no detectable structural variants and were therefore idiopathic (Parikshak et al., 2016). It is possible that individuals with syndromic ASD could have unique chromatin aberrations. We therefore defined DA peaks separately for syndromic and idiopathic ASD, relative to the same set of controls (STAR Methods). Remarkably, acetylation changes were highly concordant genome-wide between the two forms of ASD (Figure 1E, R = 0.88 in PFC, R = 0.87 in TC). To maximize statistical power, we therefore retained the original set of DA peaks based on all ASD samples (syndromic plus idiopathic).

PFC and TC gene expression levels have been comprehensively measured using RNA sequencing (RNA)-seq in a parallel study on the same cohort (Parikshak et al., 2016). To investigate the consistency between chromatin aberrations and gene expression changes in ASD, we focused on promoter regions of differentially expressed (DE) genes (FDR $\leq 0.05$; linear mixed model) (Parikshak et al., 2016). We used the EFilter tool (Kumar et al., 2013) to convert promoter histone acetylation profiles into expression estimates and then identified the subset of DE genes whose acetylation-based expression estimates were significantly divergent between ASD and control ($Q \leq 0.05$; Wilcoxon test; Benjamini-Hochberg correction), after controlling for covariates as before. At these gene loci, promoter




| Brain Region | ASD vs Control: no. of DA peaks (*P*-value) | Average no. of DA peaks in permutated data |
|---|---|---|
| PFC | 5,153 (<1e-3) | 84 |
| TC | 7,009 (<1e-3) | 59 |
| CB | 247 (<1e-3) | 2 |

**Figure 1. Histone Acetylome-wide Association Study of ASD**

(A) Overview of post mortem tissues in three brain regions used in this study. H3K27ac ChIP-seq was performed on prefrontal cortex (PFC), temporal cortex (TC), and cerebellum (CB) samples from 45 ASD (A) and 49 control (C) individuals. ASD subjects with 15q duplication syndrome are highlighted in green.

(B) Number of DA peaks detected in the three brain regions. For comparison, the average number of DA peaks called in 1,000 randomized datasets (permutation of case-control labels) is shown. The p value was computed as the fraction of randomized datasets yielding at least as many DA peaks as the true dataset.

(C) PCA of ChIP-seq peak heights (DA peaks only) in the three brain regions. Red dots, ASD samples; blue diamonds, control samples. Unfilled dots and diamonds indicate atypical samples. Variance explained by PC1 and PC2 are shown in parentheses. The vertical line, the threshold on ASD-specific global acetylome signature (AGAS) score.

(D) Venn diagrams showing Overlap between DA peak sets from the three brain regions. The hypergeometric test was used to calculate p values, with the set of all peaks as background. The density plot shows the log2 fold change in TC versus PFC in the union of DA peaks. The corresponding p value was calculated assuming a t-distributed Pearson correlation coefficient.

(E) Correlation between log2 fold change of dup15q and idiopathic samples in PFC and TC. The correlation coefficient and its p value were calculated as in (D). ChIP-seq peaks within the 15q duplication region are highlighted in red.

(F) Correlation between fold change of differential acetylation in the promoter region (DA) and differential gene expression (DGE) in PFC and TC. Only significantly differential loci are shown (PFC:58 genes, TC:79 genes; STAR Methods). Statistical significance of concordance in fold change direction was calculated using the hypergeometric test.

See also Figures S1, S2, S3, S4, and S5 and Tables S1, S2, and S4.

acetylation changes were significantly correlated with expression fold change, both in PFC (R = 0.33; p = 0.016) and in TC (R = 0.38; p = 0.0012) (Figure 1F). This analysis was not extended to CB, due to the lack of detectable DE genes in that tissue. Thus, while measurement noise and biological differences between chromatin variation and expression variation may have

attenuated the similarity between the two, we nevertheless observed evidence of overall consistency between acetylation aberrations and expression dysregulation in ASD.

Having confirmed the robustness and consistency of the DA peak set, we constructed an ASD-specific global acetylome signature (AGAS), defined as the first principal component (PC1) of the corresponding peak height matrix. The strength of this signature in each brain sample was thus given by its PC1 score (X coordinate in Figure 1C). In all three brain regions, ASD samples had significantly lower AGAS scores, and disease status explained 12%–63% of the score variance (Figures 1C and S5). Conversely, a simple threshold on the AGAS score could be used to predict disease status for 68%–95% of samples in the three brain regions (Figure 1C). Thus, ASD was associated with a coherent global shift in the histone acetylome.

## Functional Properties of ASD Chromatin Aberrations

Although ASD is known to be etiologically highly heterogeneous, we hypothesized that its diverse causal genetic and environmental perturbations could potentially converge on a small set of downstream pathways (Voineagu et al., 2011; Parikshak et al., 2015). We therefore used the GREAT tool (McLean et al., 2010) to test for significantly enriched (fold change $\geq$ 1.5, Q value $\leq$ 0.01) gene categories in DA loci. GREAT maps regulatory elements to flanking genes based on proximity and uses the hypergeometric test to determine if the fraction of DA peaks near genes from any particular functional category is greater than expected by chance. Overall, DA peaks in PFC and TC showed very similar functional profiles. In both cortical regions, upregulated DA peaks were significantly enriched in neuronal functions known to be perturbed in ASD, including synaptic transmission, metal cation transport, epilepsy, and the glutamate receptor pathway (Voineagu et al., 2011) (Figure 2A; Table S3). Known ASD genes from these categories include *CACNA1C* and *GRIN2B* (Splawski et al., 2004; O'Roak et al., 2011), which flank five and seven DA peaks, respectively (Figure 2B). In light of the observation that zinc deficiency is common in ASD (Yasuda et al., 2011), it is intriguing that upregulated DA peaks were also enriched in loci related to zinc ion homeostasis. Genes contributing to this result include the *SLC30A5* zinc transporter gene, which flanks multiple DA peaks and has been shown to harbor rare single nucleotide variants in ASD (O'Roak et al., 2011; Sanders et al., 2012).

Downregulated DA peaks also showed highly significant enrichment for specific functions (Figure 2C). Immune-related terms such as abnormal immune serum protein physiology and lymphatic system disease were most prominently enriched in this peak set in PFC, perhaps reflecting unique microglial (Rodriguez and Kern, 2011; Zhan et al., 2014) or lymphoid cell states (Louveau et al., 2015) in ASD cortex. Downregulated peaks in TC showed similar immune-related enrichments (Table S3). DA-Down peaks in the two cortical regions were also enriched near histone deacetylase genes, including *HDAC2* and *HDAC4* (Pazin and Kadonaga, 1997). In particular, the syndromic autism gene *HDAC4* (Williams et al., 2010) neighbored 16 downregulated DA peaks in TC and 4 in PFC (Figure 2D; Table S4; note that some of the DA peaks in Table S4 have been annotated with the names of smaller genes within the introns of *HDAC4*). Chemokine gene loci were also enriched for DA-Down peaks in both cortical regions. In addition, when all genes in the genome were individually scored for enrichment near DA-Down peaks, the chemokine receptor *CX3CR1* ranked among the top five in both PFC and TC (Table 1). Multiple gut-related gene groups such as embryonic digestive tract morphogenesis and digestive/alimentary phenotype showed significantly reduced histone acetylation in ASD cortex. These gene sets included multifunctional morphogenetic genes such as *FGFR2*, chemokine ligand and receptor genes (including *CX3CR1*), and the HDAC *SALL3* (Table S3). Among the individually enriched genes near downregulated peaks in TC, the behavior-related gene *GRB10* (Garfield et al., 2011) was the most statistically significant ($p < 1e\text{-}9$).

We hypothesized that DA peaks might act as *cis*-regulatory elements for some of the genes thought to be causal for ASD. We therefore analyzed a curated list of 296 ASD genes for proximity to DA peaks (SFARI database, gene score $\leq$4) (Abrahams et al., 2013) (Table S5). In order to control for biases in gene length and intergenic size, this analysis was performed using the same statistical procedure as in the GREAT tool (McLean et al., 2010) (hypergeometric test). In both PFC ($p = 0.017$) and TC ($p = 0.025$), peaks showing increased acetylation in ASD were significantly enriched near the ASD gene set (Table S5). ASD-downregulated peaks in TC were also enriched, though not significantly ($p = 0.055$) and downregulated peaks in PFC showed no enrichment for known ASD loci. Cerebellar DA peaks were too few in number to analyze in this manner. Cortical DA peaks as a whole (union of the four DA peak sets) were clearly overrepresented near ASD genes ($p = 7.6e\text{-}4$; fold enrichment = 1.1).

To identify transcription factors (TFs) that potentially mediate aberrant histone acetylation in ASD, we used the HOMER tool (Heinz et al., 2010) to scan for TF-binding motifs within DA peaks. Most notably, we found strong enrichment of RFX motifs in ASD-upregulated peaks, both in TC and in PFC (Figure 3; DA Up). *RFX2* has a DA peak at its promoter and *RFX3* contains an intronic DA peak (Table S4). These two TFs are therefore the most likely candidates for driving increased acetylation in ASD. Three other TFs or TF families were enriched in DA Up peaks across both cortical regions: PAR bZIP, AP-1 and MEF2. Among the PAR bZIP candidate TFs, E4BP4 and HLF are the most promising, because their promoters are acetylated in cerebral cortex. Among the MEF2 factors, MEF2C is clearly the most prominent candidate, because the corresponding gene locus hosts five DA peaks from the downregulated list in PFC ($p = 1.1e\text{-}4$; Table S6). The nuclear receptor motif enriched in ASD-upregulated peaks in TC could relate to the glucocorticoid or mineralocorticoid receptor (GR or MR), because the corresponding gene promoters are marked by H3K27ac peaks. In contrast to the five to six motifs overrepresented in DA Up peaks in each cortical region, SOX was the only motif enriched in DA Down peaks in TC and ETS the only motif in PFC. These binding site enrichment results potentially indicate the presence of master TFs that drive dysregulation of groups of regulatory elements across the ASD genome.

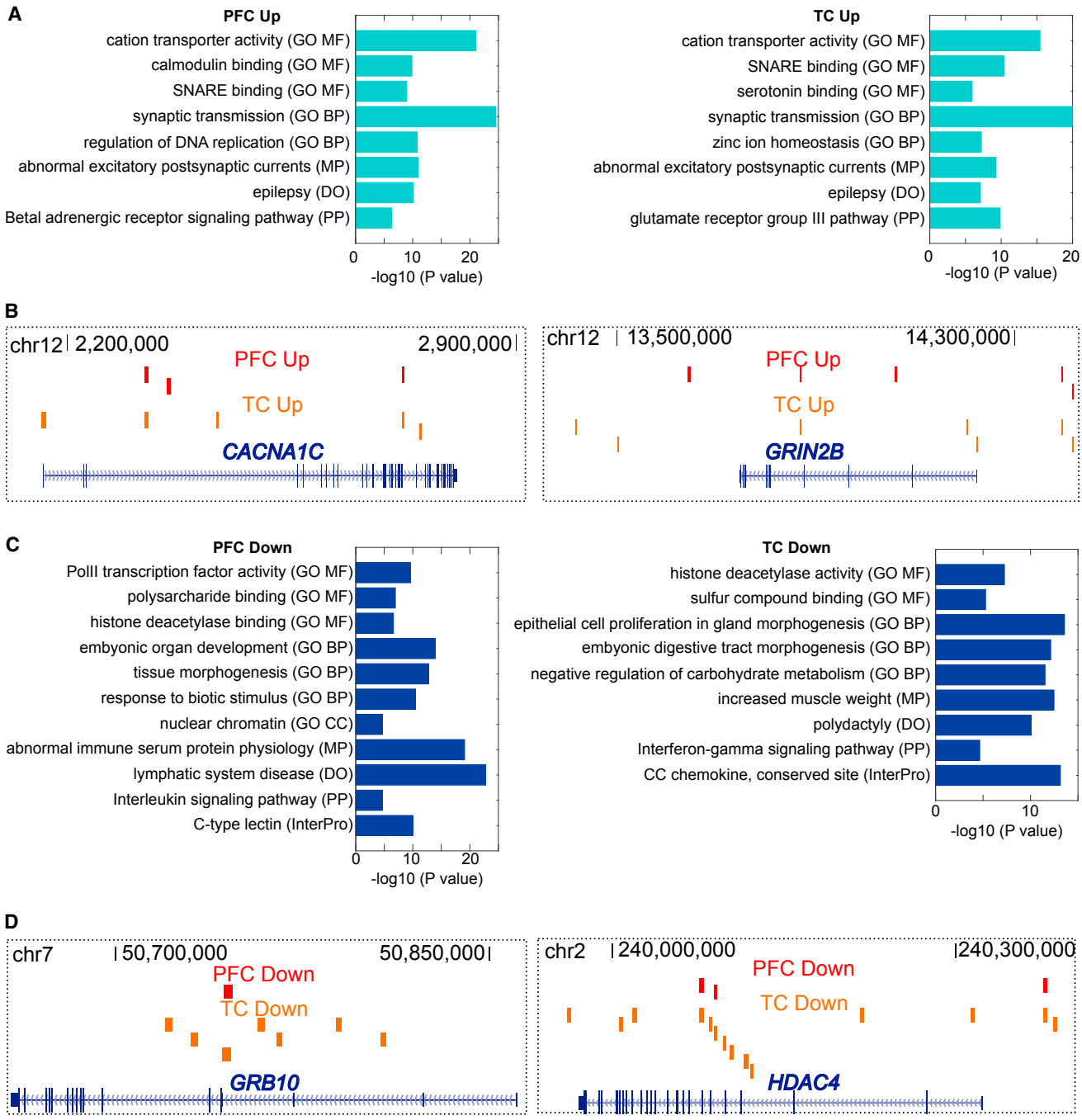



**Figure 2. Enrichment Analysis of Up and Down DA Peaks**

(A) Functional enrichment analysis of ASD-Up DA peaks in PFC and TC (GREAT tool). MF, molecular function; BP, biological process; MP, mouse phenotype; DO, disease ontology; PP, PATHER pathway.

(B) ASD-Up DA peaks in the *CACNA1C* (chr12: 2,161,809-2,900,900) and *GRIN2B* (chr12: 13,427,172-14,303,010) gene loci. Only the DA peaks closest to *CACNA1C* and *GRIN2B* are visible in these genomic windows.

(C) Functional enrichment analysis of ASD-Down DA peaks in PFC and TC (GREAT tool). CC, cellular component.

(D) ASD-Down DA peaks in the *GRB10* (chr7: 50,657,760-50,861,159) and *HDAC4* (chr2: 239,960,131-240,388,294) gene loci. Only the DA peaks closest to *GRB10* and *HDAC4* are visible in these genomic windows.

See also Table S3.

**Table 1. Top Five Genes Enriched in DA Peaks**

| Gene | p Value Rank | Raw p Value | FDR Q Value | Observed | Expected | Enrichment |
|---|---|---|---|---|---|---|
| PFC up | | | | | | |
| *LOC100996286* (*FBXW7* intron) | 1 | 6.1e−13 | 2.1e−9 | 13 | 0.97 | 13 |
| *DEAR* (*FBXW7* intron) | 2 | 2.1e−11 | 3.6e−8 | 8 | 0.37 | 22 |
| *MSRA* | 3 | 7.7e−9 | 9.1e−6 | 13 | 1.8 | 7.4 |
| *METTL24* | 4 | 4.7e−8 | 4.1e−5 | 7 | 0.46 | 15 |
| *SLC39A14* | 5 | 6.5e−8 | 4.6e−5 | 6 | 0.32 | 19 |
| PFC down | | | | | | |
| *LMOD3* | 1 | 5.7e−11 | 1.9e−7 | 12 | 1.0 | 12 |
| *NR2F2* | 2 | 3.7e−10 | 6.2e−7 | 7 | 0.31 | 22 |
| *ABCC4* | 3 | 1.2e−9 | 1.3e−6 | 9 | 0.63 | 14 |
| *FRMD4B* | 4 | 1.5e−8 | 1.3e−5 | 14 | 2.2 | 6.5 |
| *CX3CR1* | 5 | 3.9e−8 | 2.2e−5 | 7 | 0.45 | 16 |
| TC up | | | | | | |
| *SNAP25-AS1* | 1 | 1.0e−8 | 4.8e−5 | 9 | 0.81 | 11 |
| *GPM6A* | 2 | 2.7e−6 | 4.1e−3 | 9 | 1.3 | 6.8 |
| *SHANK2* | 3 | 3.1e−6 | 4.1e−3 | 8 | 1.0 | 7.7 |
| *LINC01616* | 4 | 3.6e−6 | 4.1e−3 | 5 | 0.35 | 14 |
| *NRG3-AS1* | 5 | 5.1e−6 | 4.6e−3 | 8 | 1.1 | 7.3 |
| TC down | | | | | | |
| *GRB10* | 1 | 4.3e−10 | 1.5e−6 | 16 | 2.4 | 6.5 |
| *FGFR2* | 2 | 6.7e−10 | 1.5e−6 | 17 | 2.8 | 6.0 |
| *CCL3L3/CCL3L1* | 3 | 3.1e−9 | 4.6e−6 | 8 | 0.59 | 13 |
| *LOC105375556* (*CNTNAP2* intron) | 4 | 1.4e−8 | 1.1e−5 | 8 | 0.66 | 12 |
| *CX3CR1* | 5 | 1.4e−8 | 1.1e−5 | 8 | 0.66 | 12 |

See also Table S6.

## Developmental Stage Specificity of Epigenetically Dysregulated Loci

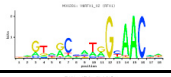
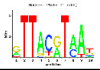
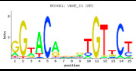
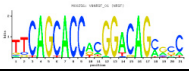
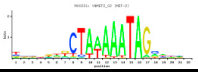
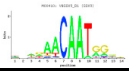
It has been shown that genes upregulated during early postnatal development are often differentially expressed in adolescent and adult ASD brain (Parikshak et al., 2013). We therefore asked whether early postnatal genes might also be enriched for the ASD-related acetylation changes we detected in older subjects ($\geq$10 years old). Using a database of human RNA-seq profiles (BrainSpan, 2015), we defined the 2,000 genes most upregulated at each developmental stage (fold change relative to median expression across all stages). We then tested for enrichment of DA peaks near each such gene set. This analysis was performed separately for PFC and TC, using expression profiles from the corresponding regions of the developing human brain. As expected, ASD-Up DA peaks in the adult (more precisely, $\geq$10 year) brain were significantly overrepresented near adult-upregulated genes (Figure 4). Surprisingly, however, we found even greater enrichment of ASD-Up DA peaks near genes upregulated at 10–12 months after birth, which corresponds to the stage of synapse formation, and neuronal maturation. In contrast, ASD-Down DA peaks did not show stage-specificity. Thus, although chromatin aberrations in ASD affect genes with a broad variety of developmental specificities, genes upregulated at or near 12 months after birth are particularly strongly associated with increased acetylation in ASD cortex.

## Histone Acetylation QTLs in Human Brain Regions

Noncoding genetic variants that affect disease susceptibility potentially act via a gene regulatory mechanism (Boyle et al., 2012; Maurano et al., 2012). Because histone acetylation serves as a measure of gene regulatory function, such variants are also likely to influence acetylation levels. It is therefore instructive to identify histone acetylation QTLs (haQTLs), which are defined as genetic variants that correlate with population variation in histone acetylation (del Rosario et al., 2015). As we and others have previously shown (del Rosario et al., 2015; Grubert et al., 2015), haQTLs can be used to prioritize causal variants within disease-associated loci.

To identify haQTLs in the three human brain regions, we used the G-SCI pipeline that was previously validated on lymphoblastoid cell lines (del Rosario et al., 2015). The pipeline uses ChIP-seq reads to call DNA sequence variants in active regulatory regions, followed by filtering to remove low-confidence variants (STAR Methods). By analogy to exome sequencing, this stage of the pipeline can be termed "regulome sequencing." A unique aspect of the G-SCI method is that called variants need not be explicitly genotyped. Rather, counts and base qualities of reference- and alternative-allele ChIP-seq reads are used to infer genotype likelihoods. These likelihoods are then used to compute the haQTL p value of the variant using the G-SCI test, which maximizes statistical power by combining information from peak

A

| Peaks | Motif name | Motif class | Protein | Motif logo | *P*-value | Q-value | Fold enrichment |
|---|---|---|---|---|---|---|---|
| TC up | V$RFX1_02 | RFX | RFX |  | 8e-31 | <1e-4 | 2.03 |
| | V$VBP_01 | PAR-BZIP | E4BP4, HLF, TEF, DBP |  | 1e-15 | <1e-4 | 1.73 |
| | V$FRA1_Q5 | BZIP | AP-1 family |  | 3e-15 | <1e-4 | 1.40 |
| | V$AR_01 | ZFC4-NR | GR, MR, AR, PR |  | 1e-8 | <1e-4 | 1.60 |
| | V$NRSF_01 | ZFC2H2 | NRSF |  | 4e-8 | <1e-4 | 4.25 |
| | V$MEF2_02 | MADS | MEF2 |  | 2e-7 | <1e-4 | 1.36 |
| TC down | V$SOX9_B1 | HMG | SOX |  | 9e-14 | <1e-4 | 1.39 |

B

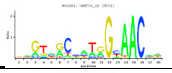
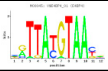
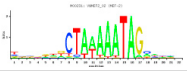
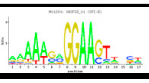
| Peaks | Motif name | Motif class | Protein | Motif logo | *P*-value | Q-value | Fold enrichment |
|---|---|---|---|---|---|---|---|
| PFC up | V$RFX1_02 | RFX | RFX |  | 4e-19 | <1e-4 | 1.86 |
| | V$FRA1_Q5 | BZIP | AP-1 family |  | 2e-10 | <1e-4 | 1.34 |
| | V$E4BP4_01 | PAR-BZIP | E4BP4, HLF, TEF, DBP |  | 9e-8 | <1e-4 | 1.70 |
| | V$HSF2_01 | HSF | HSF |  | 5e-7 | 1e-4 | 1.95 |
| | V$MEF2_02 | MADS | MEF2 |  | 1e-6 | 1e-4 | 1.37 |
| PFC down | V$SPIB_01 | ETS | ETS family |  | 2e-6 | 7e-4 | 1.34 |

**Figure 3. Enrichment of Transcription Factor-Binding Motifs in DA Peaks**

(A) Motifs significantly enriched in ASD-Up or ASD-Down DA peaks in TC (HOMER tool).

(B) Similar table, PFC.

height variability and allelic imbalance across all individuals within the cohort. In order to separate the *cis* effect of regulatory SNPs from more general disease effects, we first adjusted ChIP-seq peak heights by regressing out the diagnosis variable (ASD versus control). We then applied the G-SCI test to called SNPs and identified ~2,000 haQTLs in each of the three brain regions (Figure 5A; Table S7). Note that these haQTLs are not specific to ASD. Rather, they represent region-specific regulatory variation in the general population.

GWAS analyses have not so far uncovered statistically significant ASD-associated variants that have been replicated in independent analysis at a genome-wide level. We therefore intersected the haQTL set with genome-wide significant ($p \leq 5e\text{-}8$) variants known to be associated with shared aspects of five psychiatric disorders: schizophrenia, bipolar disorder, major depressive disorder, ASD, and attention-deficit/hyperactivity disorder (Cross-Disorder Group of the Psychiatric Genomics Consortium, 2013). While this GWAS set was too small to test for statistical enrichment near haQTLs, we did uncover two instances where brain haQTLs were strongly linked ($R^2 \geq 0.8$) to disease-associated variants (Table S7). Most notably, an haQTL (rs4765905) in an intron of the syndromic ASD gene *CACNA1C*

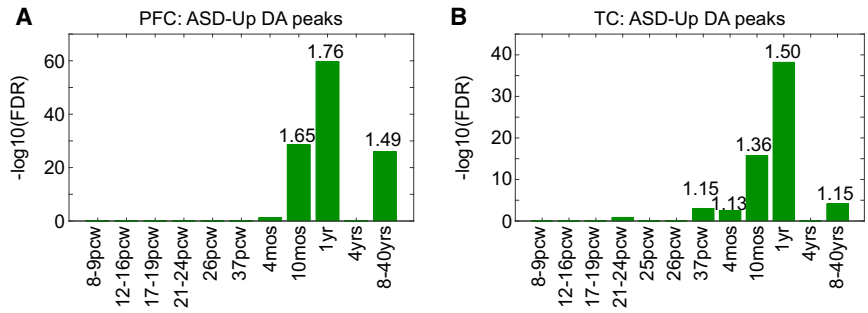



**Figure 4. Enrichment of DA Peaks near Genes Upregulated at Specific Stages of Brain Development**

(A) ASD-Up DA peaks in PFC are most significantly enriched near genes upregulated 1 year after birth. Bar height indicates enrichment Q value (FDR). Numbers above bars indicate fold enrichment ($Q \leq 0.05$).

(B) Similar plot, TC.

was linked to multiple psychiatric disease-associated SNPs within the locus (Figure 5B). Based on Hi-C data from GM12878 cells (Jin et al., 2013), the putative enhancer containing this haQTL SNP was predicted to form a long-range loop to the *CACNA1C* promoter, suggesting that it could exert its influence on psychiatric disease by modulating the chromatin state of *CACNA1C*. In addition, we intersected haQTLs with 128 SNPs associated with schizophrenia in a recent large-scale meta-analysis of schizophrenia (Ripke et al., 2014). This analysis revealed two additional haQTLs strongly linked to psychiatric disease-associated variants (Table S7). For example, we found that the haQTL SNP rs8054791 was linked to the schizophrenia-associated variant rs9922678 in an intron of *GRIN2A*, a glutamate receptor gene that has also been associated with ASD (Figure 5C).

## DISCUSSION

Despite etiological heterogeneity, our results indicate that shared aberrations in histone acetylation are widespread in ASD cerebral cortex: over 5,000 enhancer or promoter loci were systematically shifted up or down (Figure 1B). The fact that histone acetylation changes were broadly similar between PFC and TC indicates similarity in ASD mechanisms across cortical regions and also suggests that our results on differential acetylation are unlikely to represent methodological artifacts. Note that, as expected for a complex disorder with highly heterogeneous etiology, this global signature of chromatin alteration is not shared by all ASD samples (Figure 1C). An earlier transcriptomic study revealed a similar pattern of changes shared by many, but not all, ASD cases (Voineagu et al., 2011). Nevertheless, the fact that the majority of patients conform to a single global epigenomic pattern indicates that the diverse causal mechanisms of ASD have shared downstream effects on the acetylome. In contrast to cerebral cortex, only 247 loci were found to be perturbed in cerebellum, indicating that the former is affected to a much greater degree. This disparity between ASD cerebrum and cerebellum has also been observed at the transcriptomic level (Voineagu et al., 2011). Syndromic dup15q cases showed acetylome alterations that were highly correlated with those observed in idiopathic ASD ($R \geq 0.87$ in cerebral cortex), suggesting that most chromatin aberrations are shared between idiopathic ASD and this syndromic form.

To examine the genetic basis of the epigenomic aberrations detected in ASD, we tested all high-coverage SNPs within DA peaks for genetic differentiation between patients and controls (chi-square test). The distribution of genetic differentiation p values was close to uniform (data not shown), suggesting that genetic variation in *cis* SNPs is not a major contributor to case-control acetylation differences at DA peaks. It is thus likely that ASD-specific differential acetylation is driven mostly by other factors such as environmental influences, SNPs in *trans* (at a different locus), indels, and larger chromosomal variants (Krumm et al., 2015).

Overall, acetylation changes in ASD cerebral cortex were significantly correlated with differential gene expression, consistent with the known functional consequences of these alterations in chromatin structure (Figure 1F). However, the majority of DA peaks did not lie next to DE genes. This is consistent with previous studies; we and others have shown that differences in chromatin state between two sample types are only moderately correlated with differential expression (Kumar et al., 2013; Yen and Kellis, 2015). Differences in the sensitivity of ChIP-seq and RNA-seq at various loci could provide one explanation for this phenomenon. For example, post-mortem RNA degradation or low steady-state mRNA levels could reduce the detectability of DE genes in some cases, while low read mappability or occlusion of the acetylated epitope (for example) could limit the sensitivity of DA peak analysis at other loci. Moreover, noise levels could vary between the mRNA and chromatin readouts at individual loci, resulting in differential statistical power. Finally, although histone acetylation and gene expression are correlated in general, post-transcriptional regulation, other histone modifications, DNA methylation status, and the influence of additional regulatory elements within the same locus could all contribute to genuine biological differences between mRNA fold change and acetylation shifts. Thus, case-control chromatin profiling could serve as a valuable complement to the more common strategy of transcriptomic profiling by highlighting novel disease mechanisms.

We found evidence for shared pathways and functional themes among DA loci in ASD cerebral cortex (Figure 2). Among loci with increased H3K27ac, there was strong enrichment for genes related to ion channels, synaptic function, and epilepsy/neuronal excitability, all of which have previously been shown to be dysregulated in this disorder (Voineagu et al., 2011; Bourgeron, 2015). Moreover, these adult DA loci were strongly enriched for genes developmentally upregulated at or around 12 months of life (Figure 4), which coincides with the peak of early experience-dependent synaptogenesis. A similar temporal enrichment has also been observed for cerebral DE genes in

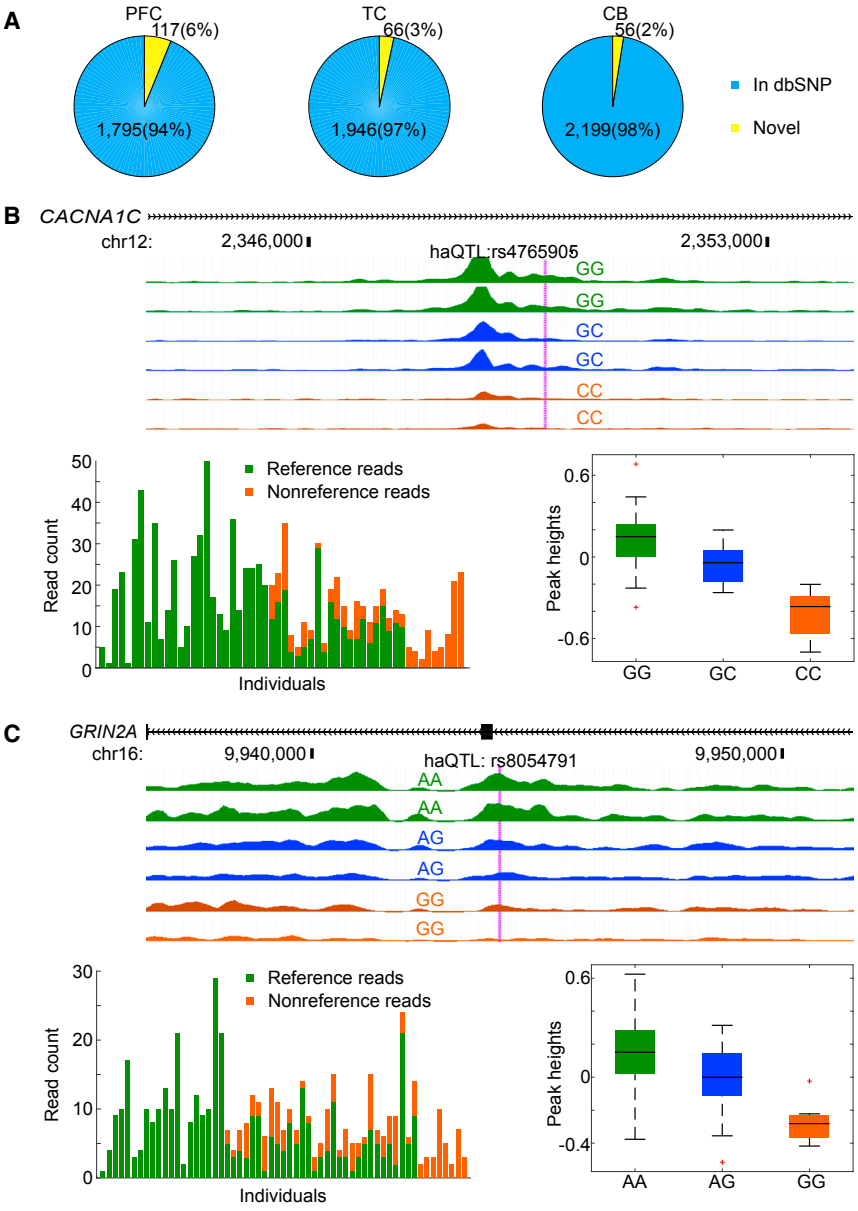



**Figure 5. Histone Acetylation Quantitative Trait Loci and Linkage with GWAS SNPs**

(A) Pie chart, number of histone acetylation quantitative trait loci (haQTLs) called in PFC, TC, and CB.

(B) A SNP within an intron of *CACNA1C* is an haQTL (Q = 1.5e-4) in CB and is in LD with four genome-wide significant SNPs from GWAS of five psychiatric disorders (Table S7). Histone acetylation tracks (chr12: 2,343,551-2,354,513), read depth analysis (bar graph) and peak height boxplots indicate that the reference "G" allele has higher histone acetylation than the non-reference "C" allele. All acetylation tracks are plotted on the same fold-enrichment scale (y axis: 0–120).

(C) A SNP within an intron of *GRIN2A* (chr16: 9,936,580-9,951,221) is an haQTL (Q = 2.2e-2) in TC and CB and is in LD with a genome-wide significant SNP (rs9922678; $R^2$ = 0.91) for schizophrenia. Histone acetylation tracks (chr12: 2,343,551-2,354,513), read depth analysis (bar graph), and peak height boxplots indicate that the reference "A" allele has higher histone acetylation than the non-reference "G" allele. All acetylation tracks are plotted on the same fold-enrichment scale (y axis: 0–50).

See also Table S7.

ASD (Parikshak et al., 2013). Loci with decreased acetylation in ASD also converged on shared functional categories, such as digestive tract morphogenesis, chemokine signaling, HDAC activity, and immune processes related to microglia. Note that it is possible for functional categories to appear systematically enriched in DA peaks merely because of the contribution of a single highly enriched "jackpot" gene. However, our results are likely to be robust to such artifacts, because we discarded functional terms that had fewer than five genes near DA peaks and then manually inspected the remaining top hits (shown in Figure 2) for jackpot effects. While the primary causes of ASD are highly heterogeneous, it appears that they nevertheless converge on shared downstream epigenomic changes associated with specific functions. It is possible that these shared chromatin alterations could in turn drive some of the shared symptoms of ASD.

The above functional enrichments have intriguing links to ASD epidemiology and results from model organisms. In addition to the well-studied roles of synaptic, ion-channel, and glutamate-pathway genes in ASD (Schmunk and Gargus, 2013; Parikshak et al., 2015), exposure to HDAC inhibitors in utero has been linked to ASD and ASD-like symptoms in humans and has also been demonstrated to cause social deficits in rodents (Chomiak et al., 2013; Christensen et al., 2013). HDAC suppression could thus be a common epigenomic feature of ASD. Chemokine pathway changes in ASD are also plausible. Suppression of the chemokine receptor gene *CX3CR1*, which flanks eight downregulated peaks in TC (p = 1.4e-8, Table 1), causes microglial activation (Wolf et al., 2013). Moreover, *CX3CR1* knockout mice have two phenotypes observed in autism: impaired social interaction and increased repetitive behavior (Zhan et al., 2014). Finally, the enrichment of downregulated DA peaks near digestive tract morphogenesis genes could point to the existence of pleiotropic loci potentially contributing to the comorbidity of gastrointestinal problems with ASD (McElhanon et al., 2014).

In addition to pathway-level chromatin aberrations, we found strong enrichment of DA peaks near individual genes. The chemokine pathway genes *CCL3L1/CCL3L3* (p = 3.1e-9) and *CX3CR1* (p = 1.4e-8) were both among the top five genome-wide for enrichment in downregulated TC peaks (Tables 1 and S6). The top-ranked gene in the same downregulated peak list

was *GRB10* (p = 4.3e-10), an imprinted gene expressed via the paternal allele in neurons and the maternal allele in most other adult mouse tissues (Plasschaert and Bartolomei, 2015). Deletion of the paternal allele specifically affects social behavior in mice (Garfield et al., 2011). Moreover, the GRB10-interacting GYF proteins GIGYF1 and GIGYF2 are known to harbor de novo loss-of-function mutations in ASD (Krumm et al., 2015). At a functional level, GRB10 mediates a negative feedback loop that damps mTORC1 signaling (Yu et al., 2011), a pathway with multiple links to ASD. mTORC1 hyperactivity alters the synaptic excitation/inhibition ratio and causes multiple autism-like symptoms in mice (Gkogkas et al., 2013). In addition, mTORC1 is negatively regulated by four syndromic autism genes (Wang and Doering, 2013). Thus, *GRB10* deacetylation could represent a common epigenetic mechanism of idiopathic ASD via a pathway that is also affected by rare genetic variants in syndromic ASD. *HDAC4* provides yet another example of mechanistic parallelism between rare genetic and common epigenetic mechanisms. The *HDAC4* gene is mutated in a syndromic form of ASD (Williams et al., 2010) and flanks 16 peaks deacetylated in ASD (Table S4). The syndromic ASD gene *CNTNAP2* provides yet another example—it ranks fourth in the genome for deacetylated TC peaks in ASD (Table 1). On a broader scale, the convergence of rare genetic mutations and common "epimutations" on similar pathways in ASD is supported by the genome-wide similarity of histone acetylation changes between dup15q syndrome and idiopathic ASD (Figure 1E).

Among the TFs highlighted by motif analysis of DA peaks upregulated in ASD, the neurodevelopmental factor MEF2C (Li et al., 2008) has substantial evidence for genetic association with ASD (Novara et al., 2010; Neale et al., 2012). Encouragingly, it has also been identified through motif analysis of co-regulated gene networks containing ASD risk genes (Parikshak et al., 2013). The MEF2 complex is known to interact with HDAC4 (Gruffat et al., 2002), which raises the hypothesis that downregulation of HDACs in ASD cerebral cortex could relieve the repression of MEF2C target sites, thus increasing their histone acetylation level. AP-1, another TF enriched in DA peaks, has also been shown to interact with HDAC4 (Yamaguchi et al., 2005). As the most highly enriched motif in PFC (p = 4e-19) and TC (p = 8e-31) DA peaks, RFX is particularly noteworthy. Though there is no known genetic association of RFX TFs with ASD, members of this family play key roles in neurodevelopment (Benadiba et al., 2012; Bae et al., 2014), and our results raise the hypothesis that they could serve as mediators of diverse upstream causal factors. The enrichment of MR binding sites at upregulated peaks in TC is also noteworthy—the *NR3C2* gene encoding MR was recently shown to be significantly associated with autism in a recent exome-sequencing study (De Rubeis et al., 2014).

Although ASD is the focus of the current study, the haQTLs detected in PFC, TC, and CB (Table S7) are not specific to ASD. Thus, they can serve as a resource for prioritizing causal SNPs for a broad range of brain-related disorders. For example, the haQTL set included candidate causal SNPs at four GWAS loci for schizophrenia and other psychiatric disorders, including *GRIN2A* and *CACNA1C* (Figure 5; Table S7). At these loci, we hypothesize that perturbed histone acetylation could constitute a mechanistic intermediary between genotype and phenotype. As the number of genetic association studies increases and cohort sizes grow ever larger, the haQTLs identified here will serve as a valuable resource for mapping causal regulatory mutations within brain disease-associated loci. This is, to the best of our knowledge, the first cohort-scale HAWAS study, and as such it lays a foundation for future studies of histone modification changes in disease. Moreover, this initial analysis of human brain haQTLs paves the way for multiple future studies of chromatin-altering variants in primary samples based on ChIP-seq, DNase-seq (Degner et al., 2012), assay for transposase-accessible chromatin using sequencing (ATAC-seq), and other assays.

## STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Human Subjects
- METHOD DETAILS
  - ChIP-Seq on Brain Tissue
  - Read Alignment and Peak Calling
  - Peak Height Normalization
  - Quality Control of 229 ChIP-Seq Datasets
  - Removal of Confounding Factors
  - Analysis of Differentially Acetylated (DA) Peaks
  - Functional Enrichment of DA Peaks
  - Enrichment of DA Loci for Expression at Specific Developmental Stages
  - Motif Analysis
  - SNP-Calling Pipeline
  - haQTL Calling
  - LD between Pyschiatric Disorder GWAS SNPs and haQTLs
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Statistical Method of Computation
  - Inclusion and Exclusion Criteria of Any Data
- DATA AND SOFTWARE AVAILABILITY
  - Data Resources

### SUPPLEMENTAL INFORMATION

Supplemental Information includes five figures and seven tables and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2016.10.031.

### AUTHOR CONTRIBUTIONS

J.P. prepared ChIP-seq libraries with help from H.S.H., R.R., and B.E. W.S. carried out data analysis. R.C.-H.d.R. helped to call haQTLs and intersected them with GWAS data. J.P. performed GREAT and motif analysis. V.K. performed the comparison of histone acetylation and gene expression. N.N.P., T.G.B., and C.C.Y.W. helped with data sample management, data interpretation and analysis. W.S. and S.P. wrote the manuscript with help from J.P. and comments from D.H.G. and J.M. D.H.G. and S.P. provided guidance on all experiments and analyses.

## ACKNOWLEDGMENTS

This work was funded by PsychENCODE Grant 1R01MH094714 and the Agency for Science, Technology and Research (A*STAR), Singapore. We are grateful to the patients and families who participate in the tissue programs from which our samples are obtained. Human tissue was obtained from the Autism BrainNet (sponsored by the Simons Foundation and Autism Speaks, formerly the Autism Tissue Program) and the University of Maryland Brain and Tissue Bank, which is a component of the NIH NeuroBioBank and the Oxford Brain Bank.

Received: January 23, 2016
Revised: July 14, 2016
Accepted: October 18, 2016
Published: November 17, 2016

## REFERENCES

Abrahams, B.S., and Geschwind, D.H. (2010). Connecting genes to brain in the autism spectrum disorders. Arch. Neurol. *67*, 395–399.

Abrahams, B.S., Arking, D.E., Campbell, D.B., Mefford, H.C., Morrow, E.M., Weiss, L.A., Menashe, I., Wadkins, T., Banerjee-Basu, S., and Packer, A. (2013). SFARI Gene 2.0: a community-driven knowledgebase for the autism spectrum disorders (ASDs). Mol. Autism *4*, 36.

Akbarian, S., Liu, C., Knowles, J.A., Vaccarino, F.M., Farnham, P.J., Crawford, G.E., Jaffe, A.E., Pinto, D., Dracheva, S., Geschwind, D.H., et al.; PsychENCODE Consortium (2015). The PsychENCODE project. Nat. Neurosci. *18*, 1707–1712.

Bae, B.I., Tietjen, I., Atabay, K.D., Evrony, G.D., Johnson, M.B., Asare, E., Wang, P.P., Murayama, A.Y., Im, K., Lisgo, S.N., et al. (2014). Evolutionarily dynamic alternative splicing of GPR56 regulates regional cerebral cortical patterning. Science *343*, 764–768.

Benadiba, C., Magnani, D., Niquille, M., Morlé, L., Valloton, D., Nawabi, H., Ait-Lounis, A., Otsmane, B., Reith, W., Theil, T., et al. (2012). The ciliogenic transcription factor RFX3 regulates early midline distribution of guidepost neurons required for corpus callosum development. PLoS Genet. *8*, e1002606.

Bernstein, B.E., Stamatoyannopoulos, J.A., Costello, J.F., Ren, B., Milosavljevic, A., Meissner, A., Kellis, M., Marra, M.A., Beaudet, A.L., Ecker, J.R., et al. (2010). The NIH Roadmap Epigenomics Mapping Consortium. Nat. Biotechnol. *28*, 1045–1048.

Bolstad, B.M., Irizarry, R.A., Astrand, M., and Speed, T.P. (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics *19*, 185–193.

Bourgeron, T. (2015). From the genetic architecture to synaptic plasticity in autism spectrum disorder. Nat. Rev. Neurosci. *16*, 551–563.

Boyle, A.P., Hong, E.L., Hariharan, M., Cheng, Y., Schaub, M.A., Kasowski, M., Karczewski, K.J., Park, J., Hitz, B.C., Weng, S., et al. (2012). Annotation of functional variation in personal genomes using RegulomeDB. Genome Res. *22*, 1790–1797.

BrainSpan (2015). BrainSpan: Atlas of the Developing Human Brain [Internet]. Available from: http://brainspan.org.

Chomiak, T., Turner, N., and Hu, B. (2013). What We Have Learned about Autism Spectrum Disorder from Valproic Acid. Pathol. Res. Int. *2013*, 712758.

Christensen, J., Grønborg, T.K., Sørensen, M.J., Schendel, D., Parner, E.T., Pedersen, L.H., and Vestergaard, M. (2013). Prenatal valproate exposure and risk of autism spectrum disorders and childhood autism. JAMA *309*, 1696–1703.

Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc. Natl. Acad. Sci. USA *107*, 21931–21936.

Cross-Disorder Group of the Psychiatric Genomics Consortium (2013). Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. Lancet *381*, 1371–1379.

de la Torre-Ubieta, L., Won, H., Stein, J.L., and Geschwind, D.H. (2016). Advancing the understanding of autism disease mechanisms through genetics. Nat. Med. *22*, 345–361.

De Rubeis, S., He, X., Goldberg, A.P., Poultney, C.S., Samocha, K., Cicek, A.E., Kou, Y., Liu, L., Fromer, M., Walker, S., et al.; DDD Study; Homozygosity Mapping Collaborative for Autism; UK10K Consortium (2014). Synaptic, transcriptional and chromatin genes disrupted in autism. Nature *515*, 209–215.

Degner, J.F., Pai, A.A., Pique-Regi, R., Veyrieras, J.-B.B., Gaffney, D.J., Pickrell, J.K., De Leon, S., Michelini, K., Lewellen, N., Crawford, G.E., et al. (2012). DNase I sensitivity QTLs are a major determinant of human expression variation. Nature *482*, 390–394.

del Rosario, R.C., Poschmann, J., Rouam, S.L., Png, E., Khor, C.C., Hibberd, M.L., and Prabhakar, S. (2015). Sensitive detection of chromatin-altering polymorphisms reveals autoimmune disease mechanisms. Nat. Methods *12*, 458–464.

DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. *43*, 491–498.

Devlin, B., and Scherer, S.W. (2012). Genetic architecture in autism spectrum disorder. Curr. Opin. Genet. Dev. *22*, 229–237.

Garfield, A.S., Cowley, M., Smith, F.M., Moorwood, K., Stewart-Cox, J.E., Gilroy, K., Baker, S., Xia, J., Dalley, J.W., Hurst, L.D., et al. (2011). Distinct physiological and behavioural functions for parental alleles of imprinted Grb10. Nature *469*, 534–538.

Geschwind, D.H., and State, M.W. (2015). Gene hunting in autism spectrum disorder: on the path to precision medicine. Lancet Neurol. *14*, 1109–1120.

Gkogkas, C.G., Khoutorsky, A., Ran, I., Rampakakis, E., Nevarko, T., Weatherill, D.B., Vasuta, C., Yee, S., Truitt, M., Dallaire, P., et al. (2013). Autism-related deficits via dysregulated eIF4E-dependent translational control. Nature *493*, 371–377.

Grabrucker, A.M. (2013). Environmental factors in autism. Front. Psychiatry *3*, 118.

Grubert, F., Zaugg, J.B., Kasowski, M., Ursu, O., Spacek, D.V., Martin, A.R., Greenside, P., Srivas, R., Phanstiel, D.H., Pekowska, A., et al. (2015). Genetic control of chromatin states in humans involves local and distal chromosomal interactions. Cell *162*, 1051–1065.

Gruffat, H., Manet, E., and Sergeant, A. (2002). MEF2-mediated recruitment of class II HDAC at the EBV immediate early gene BZLF1 links latency and chromatin remodeling. EMBO Rep. *3*, 141–146.

Guintivano, J., Aryee, M.J., and Kaminsky, Z.A. (2013). A cell epigenotype specific model for the correction of brain cellular heterogeneity bias and its application to age, brain region and major depression. Epigenetics *8*, 290–302.

Harrow, J., Frankish, A., Gonzalez, J.M., Tapanari, E., Diekhans, M., Kokocinski, F., Aken, B.L., Barrell, D., Zadissa, A., Searle, S., et al. (2012). GENCODE: the reference human genome annotation for The ENCODE Project. Genome Res. *22*, 1760–1774.

Heintzman, N.D., Hon, G.C., Hawkins, R.D., Kheradpour, P., Stark, A., Harp, L.F., Ye, Z., Lee, L.K., Stuart, R.K., Ching, C.W., et al. (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature *459*, 108–112.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell *38*, 576–589.

Jin, F., Li, Y., Dixon, J.R., Selvaraj, S., Ye, Z., Lee, A.Y., Yen, C.-A.A., Schmitt, A.D., Espinoza, C.A., and Ren, B. (2013). A high-resolution map of the three-dimensional chromatin interactome in human cells. Nature *503*, 290–294.

Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, D. (2002). The human genome browser at UCSC. Genome Res. *12*, 996–1006.

Krumm, N., Turner, T.N., Baker, C., Vives, L., Mohajeri, K., Witherspoon, K., Raja, A., Coe, B.P., Stessman, H.A., He, Z.-X.X., et al. (2015). Excess of rare, inherited truncating mutations in autism. Nat. Genet. *47*, 582–588.

Kubota, T., Miyake, K., and Hirasawa, T. (2012). Epigenetic understanding of gene-environment interactions in psychiatric disorders: a new concept of clinical genetics. Clin. Epigenetics *4*, 1.

Kumar, V., Muratani, M., Rayan, N.A., Kraus, P., Lufkin, T., Ng, H.H., and Prabhakar, S. (2013). Uniform, optimal signal processing of mapped deep-sequencing data. Nat. Biotechnol. *31*, 615–622.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.

Li, H., Radford, J.C., Ragusa, M.J., Shea, K.L., McKercher, S.R., Zaremba, J.D., Soussou, W., Nie, Z., Kang, Y.J., Nakanishi, N., et al. (2008). Transcription factor MEF2C influences neural stem/progenitor cell differentiation and maturation *in vivo*. Proc. Natl. Acad. Sci. USA *105*, 9397–9402.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

Loke, Y.J., Hannan, A.J., and Craig, J.M. (2015). The role of epigenetic change in autism spectrum disorders. Front. Neurol. *6*, 107.

Louveau, A., Smirnov, I., Keyes, T.J., Eccles, J.D., Rouhani, S.J., Peske, J.D., Derecki, N.C., Castle, D., Mandell, J.W., Lee, K.S., et al. (2015). Structural and functional features of central nervous system lymphatic vessels. Nature *523*, 337–341.

Lunnon, K., Smith, R., Hannon, E., De Jager, P.L., Srivastava, G., Volta, M., Troakes, C., Al-Sarraj, S., Burrage, J., Macdonald, R., et al. (2014). Methylomic profiling implicates cortical deregulation of ANK1 in Alzheimer's disease. Nat. Neurosci. *17*, 1164–1170.

Matelski, L., and Van de Water, J. (2016). Risk factors in autism: thinking outside the brain. J. Autoimmun. *67*, 1–7.

Mathelier, A., Fornes, O., Arenillas, D.J., Chen, C.-Y.Y., Denay, G., Lee, J., Shi, W., Shyr, C., Tan, G., Worsley-Hunt, R., et al. (2016). JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. Nucleic Acids Res. *44*(D1), D110–D115.

Matys, V., Kel-Margoulis, O.V., Fricke, E., Liebich, I., Land, S., Barre-Dirrie, A., Reuter, I., Chekmenev, D., Krull, M., Hornischer, K., et al. (2006). TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. Nucleic Acids Res. *34*, D108–D110.

Maurano, M.T., Humbert, R., Rynes, E., Thurman, R.E., Haugen, E., Wang, H., Reynolds, A.P., Sandstrom, R., Qu, H., Brody, J., et al. (2012). Systematic localization of common disease-associated variation in regulatory DNA. Science *337*, 1190–1195.

McElhanon, B.O., McCracken, C., Karpen, S., and Sharp, W.G. (2014). Gastrointestinal symptoms in autism spectrum disorder: a meta-analysis. Pediatrics *133*, 872–883.

McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., and Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. Nat. Biotechnol. *28*, 495–501.

Mill, J., and Heijmans, B.T. (2013). From promises to practical strategies in epigenetic epidemiology. Nat. Rev. Genet. *14*, 585–594.

Montano, C., Taub, M.A., Jaffe, A., Briem, E., Feinberg, J.I., Trygvadottir, R., Idrizi, A., Runarsson, A., Berndsen, B., Gur, R.C., et al. (2016). Association of DNA methylation differences with schizophrenia in an epigenome-wide association study. JAMA Psychiatry *73*, 506–514.

Neale, B.M., Kou, Y., Liu, L., Ma'ayan, A., Samocha, K.E., Sabo, A., Lin, C.F., Stevens, C., Wang, L.S., Makarov, V., et al. (2012). Patterns and rates of exonic de novo mutations in autism spectrum disorders. Nature *485*, 242–245.

Novara, F., Beri, S., Giorda, R., Ortibus, E., Nageshappa, S., Darra, F., Dalla Bernardina, B., Zuffardi, O., and Van Esch, H. (2010). Refining the phenotype associated with MEF2C haploinsufficiency. Clin. Genet. *78*, 471–477.

O'Roak, B.J., Deriziotis, P., Lee, C., Vives, L., Schwartz, J.J., Girirajan, S., Karakoc, E., Mackenzie, A.P., Ng, S.B., Baker, C., et al. (2011). Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. Nat. Genet. *43*, 585–589.

Parikshak, N.N., Luo, R., Zhang, A., Won, H., Lowe, J.K., Chandran, V., Horvath, S., and Geschwind, D.H. (2013). Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. Cell *155*, 1008–1021.

Parikshak, N.N., Gandal, M.J., and Geschwind, D.H. (2015). Systems biology and gene networks in neurodevelopmental and neurodegenerative disorders. Nat. Rev. Genet. *16*, 441–458.

Parikshak, N.N., Swarup, V., Belgard, T.G., Irimia, M., Ramaswami, G., Gandal, M.J., Hartl, C., Leppa, V., de la Torre Ubieta, L., Huang, J., et al. (2016). Genome-wide changes in lncRNA, alternative splicing, and cortical patterning in autism. Nature. http://dx.doi.org/10.1038/nature20612.

Pazin, M.J., and Kadonaga, J.T. (1997). What's up and down with histone deacetylation and transcription? Cell *89*, 325–328.

Plasschaert, R.N., and Bartolomei, M.S. (2015). Tissue-specific regulation and function of Grb10 during growth and neuronal commitment. Proc. Natl. Acad. Sci. USA *112*, 6841–6847.

Quail, M.A., Kozarewa, I., Smith, F., Scally, A., Stephens, P.J., Durbin, R., Swerdlow, H., and Turner, D.J. (2008). A large genome center's improvements to the Illumina sequencing system. Nat. Methods *5*, 1005–1010.

Ripke, S., Neale, B., Corvin, A., Walters, J., Farh, K.-H., Holmans, P., Lee, P., Bulik-Sullivan, B., Collier, D., Huang, H., et al.; Schizophrenia Working Group of the Psychiatric Genomics Consortium (2014). Biological insights from 108 schizophrenia-associated genetic loci. Nature *511*, 421–427.

Rodriguez, J.I., and Kern, J.K. (2011). Evidence of microglial activation in autism and its possible role in brain underconnectivity. Neuron Glia Biol. *7*, 205–213.

Sanders, S.J., Murtha, M.T., Gupta, A.R., Murdoch, J.D., Raubeson, M.J., Willsey, A.J., Ercan-Sencicek, A.G., DiLullo, N.M., Parikshak, N.N., Stein, J.L., et al. (2012). De novo mutations revealed by whole-exome sequencing are strongly associated with autism. Nature *485*, 237–241.

Schmunk, G., and Gargus, J.J. (2013). Channelopathy pathogenesis in autism spectrum disorders. Front. Genet. *4*, 222.

Shulha, H.P., Cheung, I., Whittle, C., Wang, J., Virgil, D., Lin, C.L., Guo, Y., Lessard, A., Akbarian, S., and Weng, Z. (2012). Epigenetic signatures of autism: trimethylated H3K4 landscapes in prefrontal neurons. Arch. Gen. Psychiatry *69*, 314–324.

Splawski, I., Timothy, K.W., Sharpe, L.M., Decher, N., Kumar, P., Bloise, R., Napolitano, C., Schwartz, P.J., Joseph, R.M., Condouris, K., et al. (2004). Ca(V)1.2 calcium channel dysfunction causes a multisystem disorder including arrhythmia and autism. Cell *119*, 19–31.

Stein, J.L., Parikshak, N.N., and Geschwind, D.H. (2013). Rare inherited variation in autism: beginning to see the forest and a few trees. Neuron *77*, 209–211.

The 1000 Genomes Project Consortium, Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T., and McVean, G.A. (2012). An integrated map of genetic variation from 1,092 human genomes. Nature *491*, 56–65.

RefSeq. (2002). The Reference Sequence Project. The NCBI Handbook. Bethesda (MD): National Library of Medicine, National Center for Biotechnology Information. http://www.ncbi.nlm.nih.gov/books/NBK21091/.

van de Geijn, B., McVicker, G., Gilad, Y., and Pritchard, J.K. (2015). WASP: allele-specific software for robust molecular quantitative trait locus discovery. Nat. Methods *12*, 1061–1063.

Voineagu, I., Wang, X., Johnston, P., Lowe, J.K., Tian, Y., Horvath, S., Mill, J., Cantor, R.M., Blencowe, B.J., and Geschwind, D.H. (2011). Transcriptomic analysis of autistic brain reveals convergent molecular pathology. Nature *474*, 380–384.

Wang, H., and Doering, L.C. (2013). Reversing autism by targeting downstream mTOR signaling. Front. Cell. Neurosci. *7*, 28.

Wang, Z., Zang, C., Rosenfeld, J.A., Schones, D.E., Barski, A., Cuddapah, S., Cui, K., Roh, T.-Y.Y., Peng, W., Zhang, M.Q., and Zhao, K. (2008). Combinatorial patterns of histone acetylations and methylations in the human genome. Nat. Genet. *40*, 897–903.

Williams, S.R., Aldred, M.A., Der Kaloustian, V.M., Halal, F., Gowans, G., McLeod, D.R., Zondag, S., Toriello, H.V., Magenis, R.E., and Elsea, S.H. (2010). Haploinsufficiency of HDAC4 causes brachydactyly mental retardation syndrome, with brachydactyly type E, developmental delays, and behavioral problems. Am. J. Hum. Genet. *87*, 219–228.

Willsey, A.J., Sanders, S.J., Li, M., Dong, S., Tebbenkamp, A.T., Muhle, R.A., Reilly, S.K., Lin, L., Fertuzinhos, S., Miller, J.A., et al. (2013). Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. Cell *155*, 997–1007.

Wolf, Y., Yona, S., Kim, K.-W., and Jung, S. (2013). Microglia, seen from the CX3CR1 angle. Front. Cell. Neurosci. *7*, 26.

Yamaguchi, K., Lantowski, A., Dannenberg, A.J., and Subbaramaiah, K. (2005). Histone deacetylase inhibitors suppress the induction of c-Jun and its target genes including COX-2. J. Biol. Chem. *280*, 32569–32577.

Yasuda, H., Yoshida, K., Yasuda, Y., and Tsutsui, T. (2011). Infantile zinc deficiency: association with autism spectrum disorders. Sci. Rep. *1*, 129.

Yen, A., and Kellis, M. (2015). Systematic chromatin state comparison of epigenomes associated with diverse properties including sex and tissue type. Nat. Commun. *6*, 7973.

Yu, Y., Yoon, S.-O.O., Poulogiannis, G., Yang, Q., Ma, X.M., Villén, J., Kubica, N., Hoffman, G.R., Cantley, L.C., Gygi, S.P., and Blenis, J. (2011). Phosphoproteomic analysis identifies Grb10 as an mTORC1 substrate that negatively regulates insulin signaling. Science *332*, 1322–1326.

Zhan, Y., Paolicelli, R.C., Sforazzini, F., Weinhard, L., Bolasco, G., Pagani, F., Vyssotski, A.L., Bifone, A., Gozzi, A., Ragozzino, D., and Gross, C.T. (2014). Deficient neuron-microglia signaling results in impaired functional brain connectivity and social behavior. Nat. Neurosci. *17*, 400–406.

## STAR★METHODS

### KEY RESOURCES TABLE

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| Antibodies | | |
| H3K27ac | Active Motif | Cat#39133; RRID: AB_2561016 |
| Chemicals, Peptides, and Recombinant Proteins | | |
| Protein G Dynal magnetic beads | Invitrogen | Cat#1003D |
| Ligase | Enzymatics | Cat#L6030-HC-L |
| Polynucleotide kinase | Enzymatics | Cat#Y9040L |
| T4 DNA polymerase | Enzymatics | Cat#P7080L |
| Phusion Polymerase | NEB | Cat#M0530L |
| Klenow exo- | Enzymatics | Cat#P7010 |
| PCR_Primer_Index_4: CAAGCAGAAGACGGCATACGAGATTGG TCAGTGACTGGAGTTCAGACGTGTGCTC T**T**C**C**GATCT | Illumina | N/A |
| PCR_Primer_Index_5: CAAGCAGAAGACGGCATACGAGATCAC TGTGTGACTGGAGTTCAGACGTGTGCT CT**T**C**C**GATCT | Illumina | N/A |
| PCR_Primer_Index_6: CAAGCAGAAGACGGCATACGAGATATT GGCGTGACTGGAGTTCAGACGTGTGCT CT**T**C**C**GATCT | Illumina | N/A |
| PCR_Primer_Index_7: CAAGCAGAAGACGGCATACGAGATGAT CTGGTGACTGGAGTTCAGACGTGTGCT CT**T**C**C**GATCT | Illumina | N/A |
| Deposited Data | | |
| Raw and analyzed data | This paper | https://www.synapse.org/#!Synapse:syn4587616 |
| human reference genome NCBI build 37, GRCh37 | Genome Reference Consortium | http://www.ncbi.nlm.nih.gov/projects/genome/assembly/grc/human/ |
| Raw and analyzed data for RNA-seq | Parikshak et al., 2016 | http://biorxiv.org/content/early/2016/09/23/077057 |
| Raw data | Bernstein et al., 2010; NHGRI Epigenome Atlas | http://www.genboree.org/epigenomeatlas/index.rhtml |
| Please see Table S3 for the GREAT results | This paper | N/A |
| Please see Table S5 for the curated ASD Gene list | SFARI gene | https://gene.sfari.org/autdb/GS_Home.do |
| Refseq gene set | Refseq | https://www.ncbi.nlm.nih.gov/refseq/ |
| Gencode gene set | Gencode | https://www.gencodegenes.org/ |
| Analyzed data for human brain development RNA-seq | BrainSpan | http://brainspan.org |
| Transcription factor motif database | TRANSFAC | http://www.gene-regulation.com/pub/databases.html |
| Transcription factor motif database | JASPAR | http://jaspar.genereg.net/ |
| "Self Chain" regions of the genome | UCSC Genome Browser | http://genome.ucsc.edu/index.html |
| EUR SNPs and indels database | 1000 Genome | http://www.internationalgenome.org/ |
| GWAS SNPs on schizophrenia | Ripke et al., 2014 | N/A |

(*Continued on next page*)

*Continued*

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| GWAS SNPs on 5 psychiatric disorders | Cross-Disorder Group of the Psychiatric Genomics Consortium, 2013 | N/A |
| Other | | |
| Autism Tissue Program (ATP) | Harvard Brain Tissue Resource Center | http://www.autismtissueprogram.org/site/c.nlKUL7MQIsG/b.5183271/k.BD86/Home.htm |
| UMB BTB | University of Maryland Brain and Tissue Bank | http://medschool.umaryland.edu/btbank/ |
| the Oxford Brain Bank | University of Oxford | https://www.ndcn.ox.ac.uk/ |
| Software and Algorithms | | |
| BWA | Li and Durbin, 2009 | http://bio-bwa.sourceforge.net/ |
| SAMtools | Li et al., 2009 | http://www.htslib.org/ |
| DFilter and EFilter | Kumar et al., 2013 | http://collaborations.gis.a-star.edu.sg/~cmb6/kumarv1/dfilter/ |
| CETS | Guintivano et al., 2013 | https://r-forge.r-project.org/projects/cets/ |
| GREAT | McLean et al., 2010 | http://bejerano.stanford.edu/great/public/html/ |
| HOMER: findMotifsGenome.pl script | Heinz et al., 2010 | http://homer.salk.edu/homer/motif/ |
| GATK (v3.2-2) | DePristo et al., 2011 | https://software.broadinstitute.org/gatk/ |
| G-SCI test | del Rosario et al., 2015 | http://collaborations.gis.a-star.edu.sg/~cmb6/G-SCI_test/ |
| R | The R Project for Statistical Computing | https://www.r-project.org/ |
| MATLAB | MathWorks | http://www.mathworks.com/products/matlab/?s_tid=hp_ff_p_matlab |

## CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for reagents may be directed to, and will be fulfilled by the Lead Contact Shyam Prabhakar (prabhakars@gis.a-star.edu.sg).

## EXPERIMENTAL MODEL AND SUBJECT DETAILS

### Human Subjects

Brain samples from 45 ASD and 49 control individuals were acquired from the Autism Tissue Program (ATP) at the Harvard Brain Tissue Resource Center, the University of Maryland Brain and Tissue Bank and the Oxford Brain Bank. Sample acquisition protocols were followed for each brain bank, and samples were de-identified prior to acquisition. Sample swaps were verified with independent genotyping. Brain sample and individual level metadata are provided in Table S1.

## METHOD DETAILS

### ChIP-Seq on Brain Tissue

For each ChIP-seq experiment approximately 100mg of frozen brain tissue per sample was aliquoted and thawed on ice in 1ml PBS buffer. Tissue was then homogenized using a manual glass douncer with 7-15 slow strokes on ice. The cell suspension was filtered with a 40uM cell strainer (Falcon) by spinning at 2000rpm for 1 min at 4C in a swing bucket centrifuge (Eppendorf Centrifuge 5810R). Pellets were then washed twice with cold PBS, crosslinked with 1% formaldehyde for 15 min at room temperature and excess formaldehyde quenched by addition of glycine (0.625M). Cells were lysed with FA and nuclei were collected and re-suspended in 300 μL SDS lysis buffer (1% SDS, 1% Triton X-100, 2 mM EDTA, 50 mM HEPES-KOH [pH 7.5], 0.1% Sodium deoxy cholate, Roche 1X Complete protease inhibitor). Nuclei were lysed for 15 min, after which sonication was used to fragment chromatin to an average size of 200–500 bp (Bioruptor Next gen, Diagenode). Protein-DNA complexes were immuno-precipitated using 3 μg of H3K27acetyl antibody of the same lot for all 257 ChIP experiments (catalog number 39133; Active Motif) coupled to 50 μL protein G Dynal beads (Invitrogen) overnight. The beads were washed and protein-DNA complexes were eluted with 150 μL of elution buffer (1% SDS, 10 mM EDTA, 50 mM Tris.HCl [pH 8]), followed by protease treatment and de-crosslinking at 65°C overnight. After phenol/chloroform extraction, DNA was purified by ethanol precipitation. 5% of sheared chromatin was aliquoted and treated with Pronase and RNase

treated following de-crosslinking in the same manner as the ChIP DNA. To prepare pooled input libraries for each brain region, DNA was then quantified and equal amounts from 10 ASD and 10 control samples (randomly chosen) were pooled. 100ng of pooled DNA were used for library preparation. Input libraries were multiplexed and sequenced in one HiSeq lane. Library preparation was performed as in Quail et al. (2008). After 15 cycles of PCR using indexing primers, libraries were size selected for 300-500 bp on low melting agarose gel and 4 libraries were pooled and sequenced in one lane of 2 × 100bp using the same Illumina HiSeq 2000 with V3 reagents.

### Read Alignment and Peak Calling

Reads from 257 ChIP-seq libraries were mapped to the human genome (hg19) using BWA (Li and Durbin, 2009). Duplicate reads were filtered out using SAMtools (Li et al., 2009). 17 libraries were then discarded due to low quality (< 200,000 reads or mapping rate < 6%). Peaks were called in the remaining 240 libraries using DFilter (Kumar et al., 2013) which used the input DNA library for the corresponding brain region as a control. On average 20,447 ChIP-seq peaks were called in PFC, 20,583 in TC and 21,685 in CB. 11 samples were then discarded because they contained fewer than 10,000 peaks, leaving 229 samples for further processing. In CB, singletons were defined as peaks detected in only a single individual (zero overlap with peaks in other libraries). Non-singleton peaks were then merged across individuals (overlap > 0 bp) to define the consensus set of 38,069 CB peaks. Because peaks in PFC and TC were highly overlapping, we combined these two brain regions to define the consensus set of 56,503 neocortical peaks. All subsequent analyses of PFC and TC were performed on this neocortical peak set.

### Peak Height Normalization

For each brain region, reads were counted in 100-bp bins for each library and scaled to normalize for sequencing depth (total read count). Binned counts were then adjusted by normalizing their GC-content against the average GC-content of all libraries in each brain region. In each peak region, the sum of bin-wise normalized counts was defined as the peak height. Finally, to reduce technical variation, the heights of peaks in the union peak set were quantile-normalized (Bolstad et al., 2003).

### Quality Control of 229 ChIP-Seq Datasets

The set of 229 ChIP-seq datasets contained 13 pairs of biological replicates. The average peak height correlation (Pearson) between replicate datasets was 0.92, which is similar to replicate correlations in H3K27ac ChIP-seq data from the NHGRI Epigenome Atlas (Inferior Temporal Lobe: 0.90; Mid-Frontal Lobe: 0.91) (Bernstein et al., 2010). For each of the 229 datasets, we quantified the mean Pearson correlation (mPC) with other datasets from the same brain region. Then, from each of the 13 pairs of replicates, we discarded the dataset with lower mPC. From the remaining 216 datasets, we discarded 7 that had low mPC values relative to the norm for the corresponding brain region (< first quartile – 2.5xinter-quartile range). The remaining 209 ChIP-seq datasets from 94 individuals were used for downstream analysis.

### Removal of Confounding Factors

First, the normalized peak heights were transformed into the log2 domain. Then principal-component analysis (PCA) was performed to detect potential confounding factors by correlating the top 5 principal components (PCs) with the biological covariates (diagnosis, age, sex, neuronal cell fraction, ethnicity, and agonal state) and technical covariates (sequencing batches, brain bank, fragment median insert size from paired-end sequencing, percentage of duplicated reads, sequencing depth of each library and number of peaks for each library). The neuronal cell fraction for each sample was estimated using CETS (Guintivano et al., 2013) from DNA methylation data generated in a parallel study on the same cohort (C.C.Y.W., R. Smith, E. Hannon, L. Schalwyk, A. Kepa, J.P., W.S., N.N.P., S.P., D.H.G., and J.M., unpublished data). Neuronal cell fractions for samples that were not included in the parallel DNA methylation study were assigned with the median neuronal cell fraction across all samples. Covariates that significantly correlated with top 5 PCs were regressed out from the peak height matrix. For PFC, regressed out covariates included age, sex, neuronal cell fraction, sequencing batches, brain bank, fragment median insert size, percentage of duplicated reads, sequencing depth and number of peaks. For TC, regressed out covariates included age, sex, neuronal cell fraction, sequencing batches, brain bank, fragment median insert size, sequencing depth and number of peaks. For CB, regressed out covariates included age, sex, neuronal cell fraction, sequencing batches, brain bank, fragment median insert size, percentage of duplicated reads, sequencing depth and number of peaks. PCA was performed again after regression to confirm that no confounding factors correlated strongly with the top 5 PCs (Figure S2). Downstream analyses were based on the peak height matrix after covariate regression.

### Analysis of Differentially Acetylated (DA) Peaks

In each brain region, an initial set of differentially acetylated (DA) peaks between ASD and control was constructed based on the above-described peak height matrix (fold-change $\geq$ 1.3; $Q \leq 0.05$; Wilcoxon rank sum test; Benjamini-Hochberg correction). Using this initial set of DA peaks, we calculated the pairwise Pearson correlation coefficient matrix R of peak heights, and raised each element of the matrix to the ninth power ($R_{ij}^9$). The resulting row vectors were used to define the coordinates of ASD and control samples in correlation space, for the purpose of calculating Euclidean distances. For each sample, two distances were calculated: the median Euclidean distance to all the ASD samples (Distance_A) and the median Euclidean distance to all the control samples (Distance_C). Any ASD sample with Distance_A > 1.05xDistance_C was discarded (Figure S3). Similarly, any control sample with

Distance_C > 1.05xDistance_A was discarded. A final set of DA peaks was constructed between ASD and control using the remaining samples, with the same *Q*-value and fold change cutoffs as above (Table S4). To test whether these DA peaks were genuine, we generated 1,000 randomized datasets by permuting sample labels (ASD, control). For each permuted dataset, we called DA peaks using the above-described two-step approach. For each brain region, the *P*-value of the number of DA peaks in the actual data was calculated as the fraction of permuted datasets with an equal or greater number of DA peaks (Figure 1B).

### Functional Enrichment of DA Peaks

First, we masked the dup15q locus in the complete peak height matrix. We then used the GREAT tool (McLean et al., 2010) to determine the enrichment of gene categories in DA peaks. Genes were associated with regulatory regions using the basal+extension association rule defined by GREAT. The hypergeometric test was performed to determine if a gene category was enriched for genes associated to DA peaks (foreground set) compared to genes associated to all peaks (background set). Gene categories with fold-change $\geq$ 1.5 and $Q \leq 0.01$ were retained. Additionally, we discarded the enriched gene category if less than 5 genes were associated with DA peaks in that category. To display the non-redundant significantly enriched gene categories in Figure 2, we further selected the top 3 non-redundant gene categories in biological process and molecular function gene ontologies. Top 1 gene category from cellular component, PANTHER pathway, mouse phenotype and disease ontology are shown in the figure as well. The complete GREAT results can be found in Table S3.

### Enrichment of DA Loci for Expression at Specific Developmental Stages

A similar analysis was performed to determine the enrichment of DA peaks in SFARI genes (Tables S5 and S6), DA peaks near individual genes (Tables 1 and S6) and DA peaks near developmental stage-specific genes (Figure 4). Again, the dup15q locus was excluded. The Refseq gene set (RefSeq, 2002) was used in the first two analyses. Gencode v10 gene set (Harrow et al., 2012) was used in enrichment analysis in brain development. Human brain RNA-seq profiles were downloaded from BrainSpan (BrainSpan, 2015). We defined expressed genes as RPKM > 5 in at least 2 dataset. Then quantile-quantile normalization was performed on the RPKM values across each developmental time point (8 post-conception weeks to 40 years old) in brain regions that develop into PFC and TC. At each time point, the median RPKM values were used if there are replicate samples. We computed the coefficient of variation (CV) of each gene and clustered the samples across time points based on top 5,000 most variable expressed genes (high CVs). 11 and 12 stages were finally defined in PFC (Figure 4A) and TC (Figure 4B) by grouping similar time points based on the dendrogram, respectively. Samples at 2-3 years old were discarded due to low RNA quality (low RIN values). At each developmental stage, genes were ranked based on their gene expression fold change relative to the other stages. The top 2,276 (PFC) and 2,549 (TC) upregulated genes (fold change $\geq$ 1.5) at each developmental stage were tested for enrichment of DA peaks.

### Motif Analysis

For motif enrichment analysis, we used the HOMER ChIP-seq pipeline's findMotifsGenome.pl script with the "-mknown" option (Heinz et al., 2010). Motif models were drawn from the TRANSFAC vertebrate database (Matys et al., 2006) and the analysis was performed separately on Up and Down DA peaks from each of the 3 brain regions (6 DA peak sets in total), with all peaks from the same brain region as background. Motifs were classified as enriched based on fold enrichment ($\geq$1.3), FDR ($\leq$0.01) and number of foreground peaks that had a motif match ($\geq$20). The list of enriched motifs was almost identical when we used the JASPAR database (Mathelier et al., 2016) instead of TRANSFAC (data not shown).

### SNP-Calling Pipeline

ChIP-seq reads were aggregated across all three brain regions for each individual and then passed to the multi-sample SNP-calling pipeline. Reads used for SNP calling were de-duplicated and retained only if they were mapped to the genome in the correct orientation. We performed indel realignment, base-quality-score recalibration and SNP calling using GATK version 3.2-2 (DePristo et al., 2011). 1,297,168 SNPs within peaks in all three brain regions were called using GATK's Haplotype Caller at a SNP quality threshold of 50. Subsequently, SNP calls were filtered out with the following criteria: MQ0Fraction > 0.001, QD < 4.3, within 6 bp of an indel, more than seven SNPs within a 100-bp region, Mapping Quality < 45, Homopolymer Run > 10, MQ0 > 9.5, Dels > 0.255. Moreover, only SNP calls covered by at least 5 non-reference reads across all libraries and 3 or more non-reference reads in at least one library were retained. SNPs that violated Hardy-Weinberg equilibrium with a binomial test *P*-value $1 \times 10^{-3}$ were discarded. To eliminate mapping artifacts, SNPs in highly paralogous regions of the genome implicated by the "Self Chain" track on the UCSC Genome Browser (Kent et al., 2002) (normalized score $\geq$ 90) were filtered out. Finally, a high-confidence set of 821,606 SNPs within PFC and TC peaks and 560,972 SNPs called within CB peaks were obtained. Note that we did not perform genotype calling, since the G-SCI test does not require prior knowledge of genotypes. Rather, it integrates over the likelihoods of all three genotypes for each individual, given the data (del Rosario et al., 2015).

### haQTL Calling

haQTLs were called in the 84 Caucasian samples using G-SCI test (del Rosario et al., 2015). The diagnosis status and top PCs which account for more than 5% variance were regressed out from peak heights before haQTLs calling. We then performed the G-SCI test on each of the 821,606 SNPs within peaks for PFC and TC regions and the 560,972 SNPs within peaks for CB. For each SNP, an

adjusted *P*-value was computed using a permutation test from 10,000 to 1 million permutations until a nonzero *P*-value was obtained. After 1 million permutations, if the adjusted *P*-value was still 0, it was set to $5 \times 10^{-7}$. We then used the Benjamini and Hochberg multiple testing correction to calculate the FDR. At FDR threshold of 10%, 9094, 7468 and 9860 candidate haQTLs were identified in PFC, TC and CB. To detect possible artificial haQTLs due to different mapping rates to the reference genome between alleles, we simulated all possible 100 bp paired-end reads covering the haQTL and flanking SNPs and indels. The union of our SNP and indel (quality > 50 by GATK) calls and the 1000 Genome EUR SNPs and indels (The 1000 Genomes Project Consortium, 2012) were used. The fragment length of the simulated paired-end reads was set to be equal to 180 which is the median fragment size of all libraries. The simulated reads were then mapped to the reference genome using BWA. 1510, 1192 and 693 haQTLs were discarded because their inferred allelic imbalances from the ChIP-seq data were smaller than five times the mapping bias estimated from the simulation. The remaining haQTLs were further filtered by an effect-size filter which calculated the Pearson correlation between peak height and the fraction of Q30 nonreference bases and haQTLs with $R^2 < 0.1$ were discarded. The final set of 1912, 2012 and 2255 haQTLs in PFC, TC and CB were from the remaining haQTLs after effect-size filter and only the most significant SNP in each ChIP-seq peak was retained.

### LD between Pyschiatric Disorder GWAS SNPs and haQTLs

We downloaded two sets of GWAS SNPs, one on schizophrenia (Ripke et al., 2014) and another on 5 psychiatric disorders (Cross-Disorder Group of the Psychiatric Genomics Consortium, 2013). For the schizophrenia study we used all 128 SNPs while for the 5 psychiatric disorder study, we used a *P*-value threshold of 5e-8 (99 SNPs). The LD was calculated on the EUR population, hence for this analysis we only used SNPs that are polymorphic in the 1000 Genomes EUR population (The 1000 Genomes Project Consortium, 2012), yielding 1,863 BA41 haQTLs, 1,714 BA9 haQTLs and 2,141 Vermis haQTLs. An haQTL was considered to be in LD with a GWAS SNP if $R^2$ was at least 0.8.

## QUANTIFICATION AND STATISTICAL ANALYSIS

### Statistical Method of Computation

Statistical methods and software used in this study are cited in the STAR Methods and the Figure legends. The statistical analyses were performed in MATLAB and R. The initial and final sets of DA peaks were constructed using the Wilcoxon rank sum test and Benjamini-Hochberg multiple testing correction with fold-change $\geq 1.3$ and $Q \leq 0.05$ (e.g., Table S4). The *P*-value of the permutation test was calculated as the fraction of permuted datasets with an equal or greater number of DA peaks (e.g., Figure 1B). In the Venn diagram, the *P*-values were calculated using the hypergeometric test with the set of all peaks as background (e.g., Figure 1D). The *P*-values in the dotplots (e.g., Figures 1D, 1E, and S4) and the violin plot (Figure S5) were calculated assuming a *t*-distributed Pearson correlation coefficient. The *P*-value in the dotplot (e.g., Figure 1F) was calculated using the hypergeometric test. The gene category enrichment (e.g., Figure 2; Table S3), the enrichment of DA peaks in SFARI genes (e.g., Table S5), DA peaks near individual genes (e.g., Tables 1 and S6) and DA peaks near developmental stage-specific genes (e.g., Figure 4) were evaluated using the hypergeometric test. When calling haQTLs using G-SCI test, an adjusted *P*-value was computed for each SNP using a permutation test from 10,000 to 1 million permutations until a nonzero *P*-value was obtained. After 1 million permutations, if the adjusted *P*-value was still 0, it was set to $5 \times 10^{-7}$. We then used the Benjamini and Hochberg multiple testing correction to calculate the FDR with a threshold of 10% (e.g., Figure 5A; Table S7).

### Inclusion and Exclusion Criteria of Any Data

After mapping to the reference genome, 17 samples were discarded due to low quality (< 200,000 reads or mapping rate < 6%). 11 samples were discarded after peak calling because they contained fewer than 10,000 peaks.

## DATA AND SOFTWARE AVAILABILITY

### Data Resources

The accession number for the ChIP-seq data reported in this paper is Synapse: syn4587616.

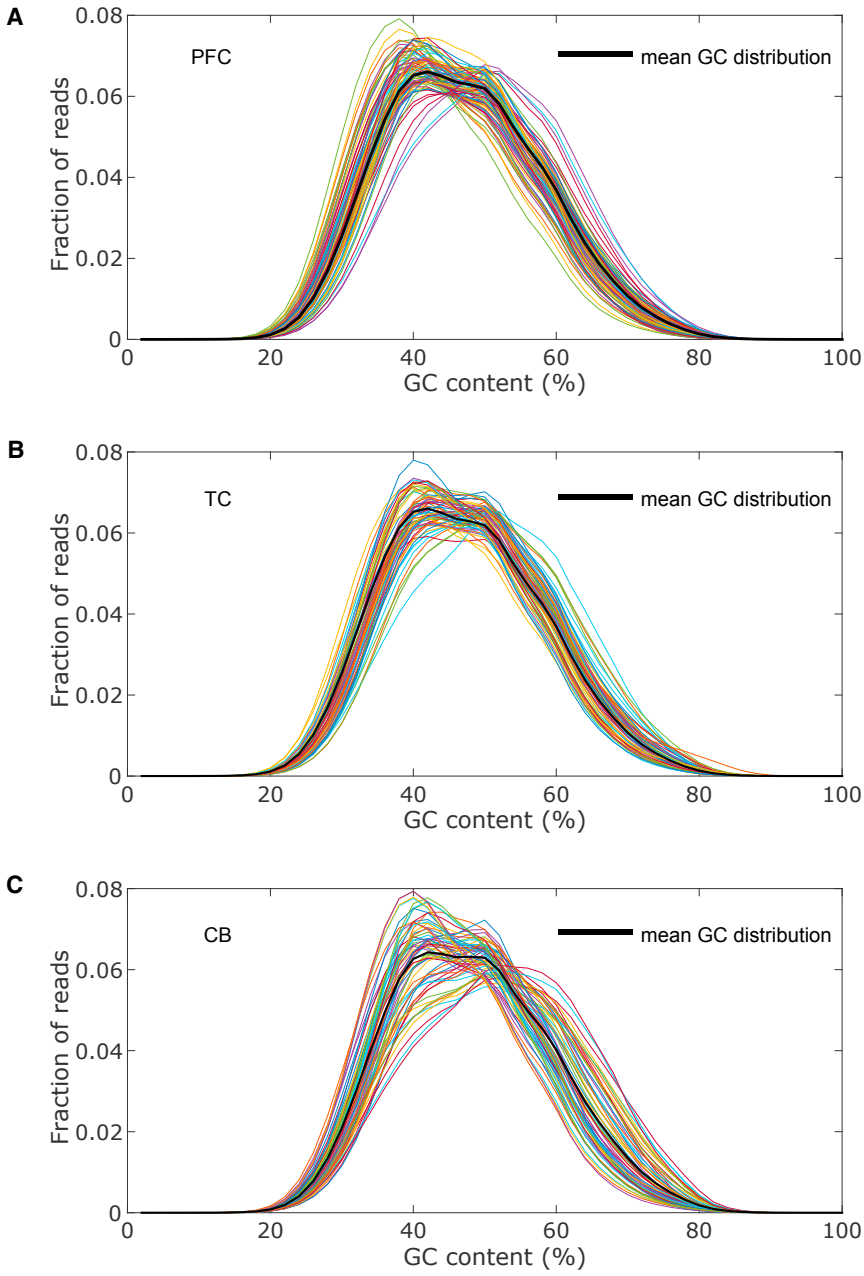



**Figure S1. GC Content Distribution of Samples in Three Brain Regions, Related to Figure 1**

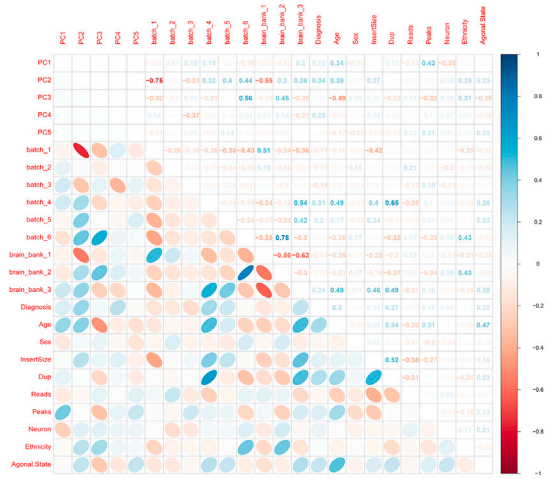
(A) GC content distributions of 81 samples in PFC were normalized to the mean GC distribution in PFC.

(B) GC content distributions of 66 samples in TC were normalized to the mean GC distribution in TC.

(C) GC content distributions of 62 samples in CB were normalized to the mean GC distribution in CB.

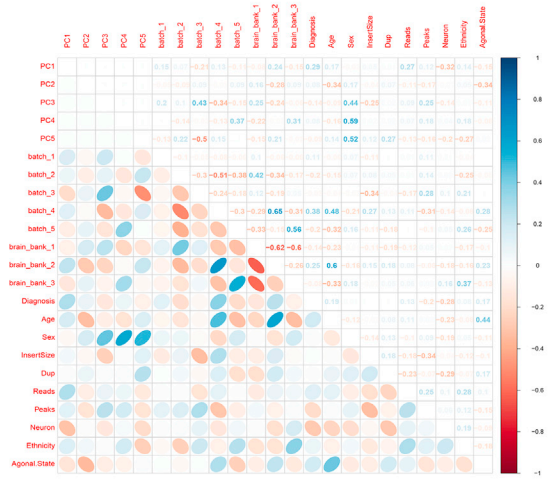
A PFC

Before regression

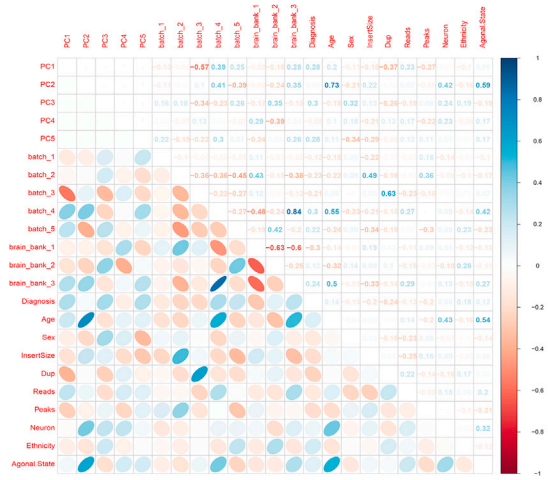


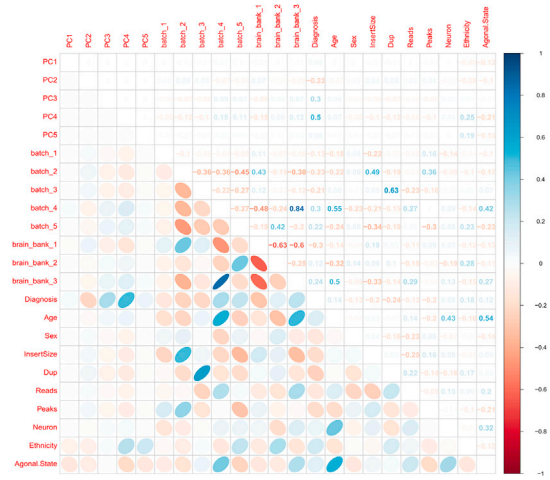
After regression

B TC



C CB





*(legend on next page)*

**Figure S2. Correlation between Top 5 Principal Components and Covariates in Three Brain Regions before and after Regression, Related to Figure 1**

(A) PFC.

(B) TC.

(C) CB.

Pearson correlation coefficient is shown at each grid point. After regressing out correlated confounding factors, the top 5 PCs correlated with none of the covariates except diagnosis. InsertSize: fragment median insert size; Dup: percentage of duplicated reads; Reads: sequencing depth; Peaks: number of peaks; Neuron: neuronal cell fraction.

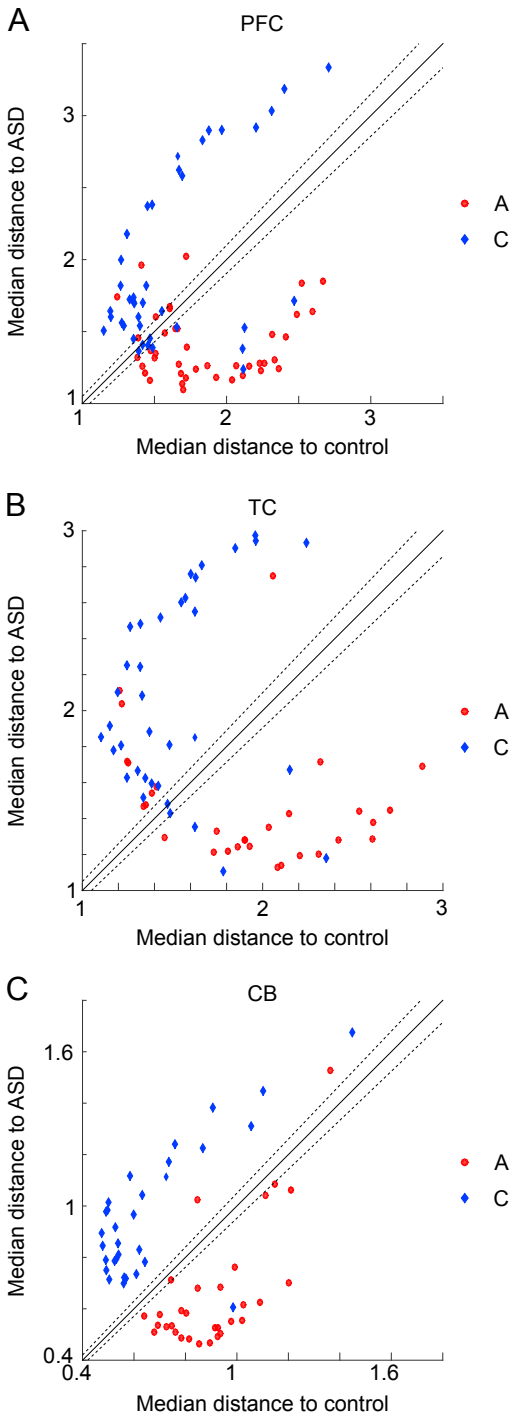



**Figure S3. Identification of Atypical Samples, Related to Figure 1**

Scatterplot of median divergence between acetylomes in PFC (A), TC (B) and CB (C). In this analysis, the acetylome is defined as the vector of peak heights at DA peaks. x axis: median Euclidean distance to other control acetylomes; y axis: median Euclidean distance to other ASD acetylomes (STAR Methods). Red dots: ASD samples; blue diamonds: control samples. Solid line: Y = X; Dotted lines: Y = 1.05X and X = 1.05Y. ASD samples above the Y = 1.05X line and control samples below the X = 1.05Y line were defined as atypical samples.

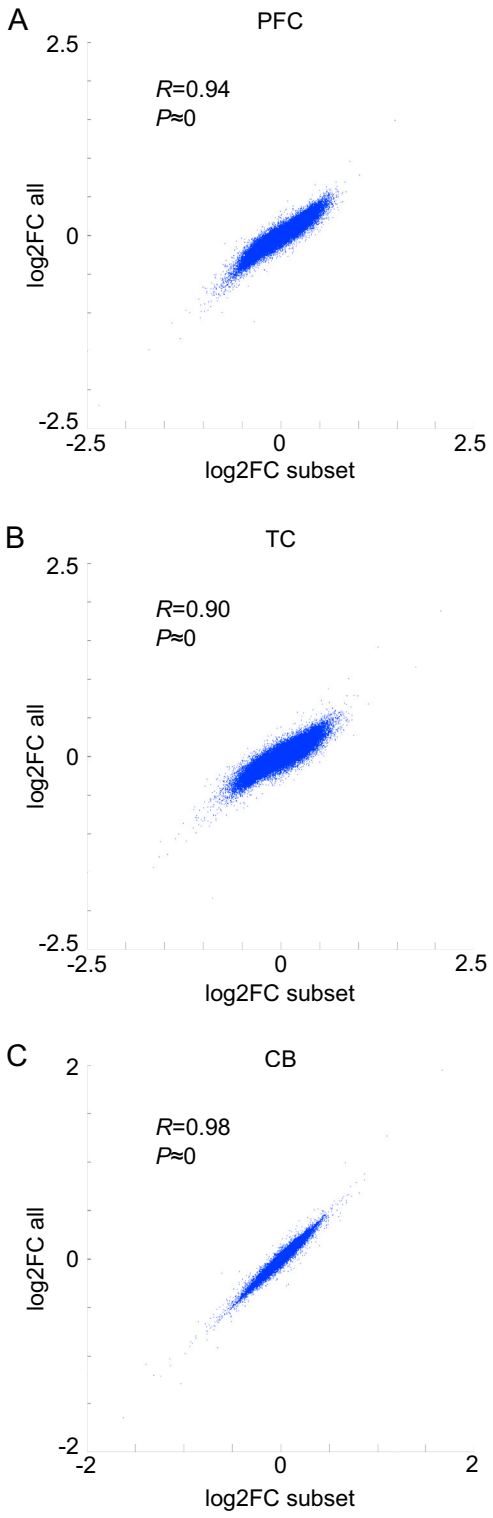



**Figure S4. Acetylation Fold Change between ASD and Control, Calculated Using All Samples Displayed on the Y Axis or Using Only Typical Samples Displayed on the X Axis, Related to Figure 1**

(A) PFC. The *P*-value of the fold-change correlation was calculated assuming a *t*-distributed Pearson correlation coefficient.

(B) Similar plot, TC.

(C) Similar plot, CB.

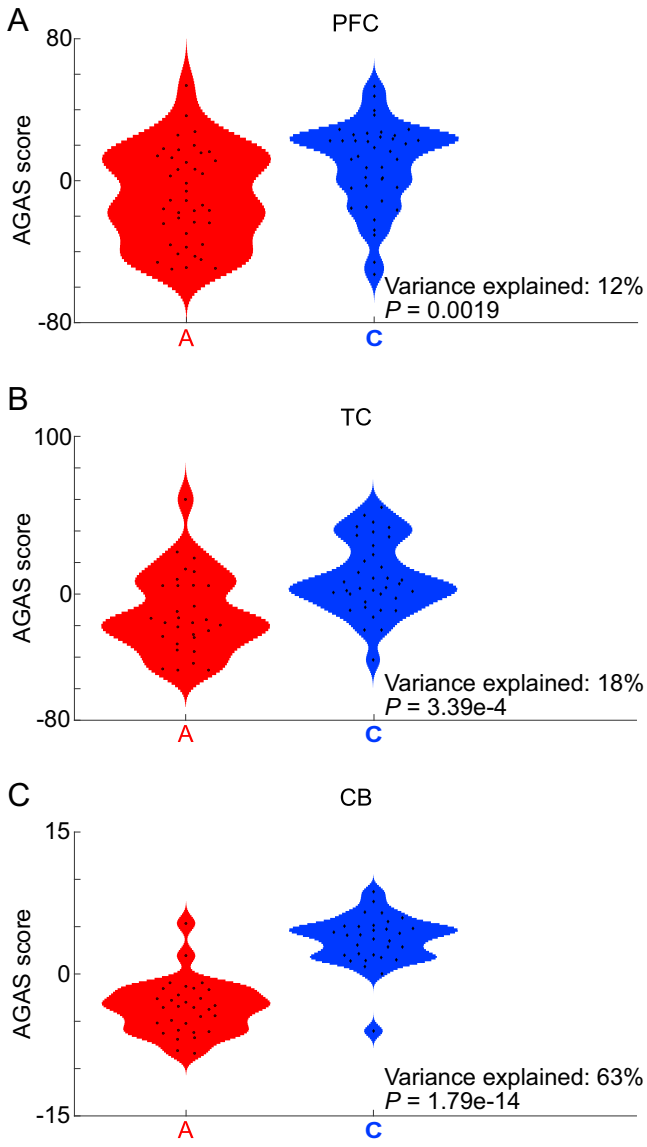



**Figure S5. ASD-Specific Global Acetylome Signature Scores, Related to Figure 1**

Violin plot of AGAS scores in PFC (A), TC (B) and CB (C). A: ASD samples; C: control samples. The *P*-value was calculated assuming a *t*-distributed Pearson correlation coefficient.