



Published in final edited form as:

Sci Signal. ; 11(513): . doi:10.1126/scisignal.aan6500.

Learning-dependent chromatin remodeling highlights noncoding regulatory regions linked to autism

John N. Koberstein¹, Shane G. Poplawski², Mathieu E. Wimmer³, Giulia Porcari⁴, Charly Kao⁵, Bruce Gomes¹, Davide Risso⁵, Hakon Hakonarson⁶, Nancy R. Zhang⁷, Robert T. Schultz⁸, Ted Abel⁹, and Lucia Peixoto^{1,*}

¹Department of Biomedical Sciences, Elson S. Floyd College of Medicine. Washington State University, Spokane, WA 99202, USA.

²Ibis Biosciences Inc. of Abbot, Carlsbad, CA 92008, USA.

³Department of Psychology and Program in Neuroscience, College of Liberal Arts, Temple University, Philadelphia, PA 19122, USA.

⁴Vanderbilt University Medical School, Nashville, TN 37232, USA.

⁵Division of Biostatistics and Epidemiology, Department of Healthcare Policy and Research, Weill Cornell Medicine, New York, NY 10065, USA.

⁶Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA.

⁷Department of Statistics, Wharton School, University of Pennsylvania, Philadelphia, PA 19104, USA.

⁸Center for Autism Research, Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA.

⁹Iowa Neuroscience Institute, University of Iowa Carver College of Medicine, Iowa City, IA 52242, USA.

Abstract

Autism spectrum disorder (ASD) is a prevalent neurodevelopmental disorder that is associated with genetic risk factors. Most human disease-associated single-nucleotide polymorphisms (SNPs) are not located in genes but rather are in regulatory regions that control gene expression. The function of regulatory regions is determined through epigenetic mechanisms. Parallels between the cellular basis of development and the formation of long-term memory have long been recognized, particularly the role of epigenetic mechanisms in both processes. We analyzed how learning alters chromatin accessibility in the mouse hippocampus using a new high-throughput sequencing

exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works

*Corresponding author: lucia.peixoto@wsu.edu.

Author contributions: Study design by L.P., M.E.W., and T.A. Data collection by L.P., M.E.W., S.G.P., G.P., and C.K. Data analysis and interpretation by L.P., J.N.K., D.R., M.E.W., S.G.P., N.R.Z., B.G., R.T.S., and T.A. Manuscript preparation by L.P. and J.N.K., with editing by M.E.W., S.G.P., H.H., D.R., N.R.Z., R.T.S., and T.A.

Competing interests: The authors declare that they have no competing interests.

Data and materials availability: HTS data will be publicly available through Gene Expression Omnibus after manuscript acceptance: GSE98317 (reviewer's access link: www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=idmlisiztyfhhmj&acc=GSE98317). DEScan code is freely available, and instructions for download and installation can be found in the user manual (file S1).

bioinformatics strategy we call DEScan (differential enrichment scan). DEScan, which enabled the analysis of data from epigenomic experiments containing multiple replicates, revealed changes in chromatin accessibility at 2365 regulatory regions—most of which were promoters. Learning-regulated promoters were active during forebrain development in mice and were enriched in epigenetic modifications indicative of bivalent promoters. These promoters were disproportionately intronic, showed a complex relationship with gene expression and alternative splicing during memory consolidation and retrieval, and were enriched in the data set relative to known ASD risk genes. Genotyping in a clinical cohort within one of these promoters (*SHANK3* promoter 6) revealed that the SNP rs6010065 was associated with ASD. Our data support the idea that learning recapitulates development at the epigenetic level and demonstrate that behaviorally induced epigenetic changes in mice can highlight regulatory regions relevant to brain disorders in patients.

INTRODUCTION

Autism spectrum disorder (ASD) is a neurodevelopmental disorder of high prevalence that involves genetic risk factors. Genetic analysis has connected a high number of genetic variants to ASD (1–5). Studies using exome sequencing that focused on identifying de novo mutations have discovered hundreds of potential ASD-associated genes (6–8). So far, most approaches for understanding genetic risk factors in ASD have focused on identifying mutations in genes (or exons). However, one of the hallmark findings from the Encyclopedia of DNA Elements (ENCODE) project is that most of the single-nucleotide polymorphisms (SNPs) associated with human disease are outside of protein-coding genes (9). Genetic variants in regulatory regions, the regions of the genome that control gene expression, are also more common (10). It is likely that a large proportion of genetic variants associated with ASD will be common variants that lie within regulatory regions and, thus, remain largely unexplored.

Several examples of regulatory genetic variants were found to have functional links to ASD (11–14). Reanalysis of available genome-wide association study data shows that loci associated with autism are more likely to influence gene expression in the brain (expression quantitative trait loci) (15). Focusing on regions that can regulate gene expression is a promising avenue to discover new ASD-linked genetic variants. However, it remains challenging to pinpoint regulatory regions that are likely to be important for a specific disease. In any given tissue at any given time, not all regulatory regions are able to be used or “accessible.” DNA exists within the nucleus of a cell as tightly packed chromatin, which can be unpackaged and repackaged as needed. The accessibility of chromatin is regulated through epigenetic modifications. The regulatory regions accessible in the brain, in response to a relevant stimulus, may be functionally relevant to ASD.

Conceptual parallels between the cellular basis of development and the formation of long-term memory have long been recognized. In particular, the epigenetic mechanisms used for deciding fate during development are the same as those that form the basis of long-term memory storage (16). Here, we explored whether epigenomic studies of learning and memory could guide the discovery of regulatory regions of relevance to neurodevelopmental disorders, such as ASD. We used sonication of cross-linked chromatin followed by high-

throughput sequencing (HTS) [Sono-seq (17)] to determine which promoters in the mouse genome are accessible in the hippocampus after a learning experience. We developed a new bioinformatics strategy called DEScan (differential enrichment scan) to analyze the data and found that learning produces increases in chromatin accessibility at 2365 regions, most of which are transcriptional start sites (TSS), and furthermore, are alternative (intronic) promoters. Learning-regulated promoters disproportionately overlap with those active during forebrain development in mice and are enriched in epigenetic modifications that are hallmarks of bivalent promoters, typically considered to poise expression of developmental genes (18). We then combined the results of our chromatin accessibility study with learning-induced genome-wide gene expression data [RNA sequencing (RNA-seq)] that we have previously generated (19, 20). Our results suggest that the relationship between changes in chromatin state and changes in gene expression is complex, with a highly significant enrichment of learning-regulated promoters relative to alternative splicing. In addition, learning-regulated promoters were disproportionately associated with known ASD risk genes. Genotyping of common SNPs in a clinical cohort within one of these regions (*SHANK3* promoter 6) revealed that SNP rs6010065 was associated with ASD. Our data support the idea that learning recapitulates development at the epigenetic level and demonstrate that behaviorally induced epigenetic changes obtained in mice can highlight functionally relevant regulatory regions for genetic association studies of brain disorders.

RESULTS

Learning increases chromatin accessibility genome-wide

Learning and long-term memory require transcription, protein synthesis, and epigenetic processes that regulate gene expression (21–27). However, it is not yet clear which specific regulatory regions are epi-genetically controlled by experience-based learning. We used Sono-seq (17) to determine which promoters in the mouse genome are regulated upon contextual fear conditioning. Sono-seq has been shown to be a useful and simple method by which to map local alterations in chromatin structure that correlate well with RNA polymerase II (RNAPol II) binding. Fear conditioning is a form of learning in which an aversive stimulus (such as a shock) is associated with a neutral context. Reexposure to the context triggers retrieval of the memory for the context-shock association, which is quantified as freezing in mice. Fear conditioning is highly reproducible among individual mice, requiring a single exposure to learn. In addition, the timeline of sensitivity for transcriptional and epigenetic inhibition is established with this test, making it an ideal learning task for our genomic study. Fear conditioning requires activity in the hippocampus, a brain region essential for long-term memory formation. Hippocampal tissue was collected at the first established sensitive period for transcriptional and epigenetic inhibition during memory consolidation, which was 30 min after fear conditioning (28, 29). Some animals were not sacrificed to confirm learning of the task, which was typical with average freezing of 55% ($\pm 10\%$) after reexposure to the context 24 hours later. Animals that were handled but not trained were dissected at the same time of day as controls. After tissue dissection, DNA was immediately cross-linked and prepared for Sono-seq (Fig. 1A). The results indicated that learning increases chromatin accessibility genome-wide (Fig. 1B).

To define where in the genome these changes were occurring, we developed a new bioinformatics strategy, which we termed DEScan. The main motivation behind DEScan was to leverage biological variation to ensure reproducibility. A common goal in epigenomic sequencing studies is to identify differences between conditions; however, how replicates affect the results has not been thoroughly explored. DEScan is an R programming language-integrated peak and differential caller, specifically designed to accommodate epigenomic signal of variable width and leverage biological reproducibility. DEScan has two modules (fig. S1A): a novel peak-calling module and a differential enrichment module (see Materials and Methods for further details). The first module integrates peak calling on individual samples using a Poisson model distribution and a variable-width window (without the need of a separate background sample) with evaluation of reproducibility of peak location across replicates to produce a count matrix for statistical testing. The results show that most of the peaks found in a sample are not reproducible (fig. S1B), highlighting the need for three or more biological replicates in differential epigenomic studies. The differential enrichment module in DEScan implements normalization and statistical testing in a manner analogous to RNA-seq data analysis described in other popular epigenetic sequencing differential analysis packages. The main difference being that, in DEScan, data normalization is tailored to epigenomic data. Traditional methods of normalization based on total read counts per sample do not properly distinguish the controls from fear-conditioned samples (fig. S2A) and remove the differential signal from the analysis (fig. S2, B and C). Using a method of normalization that can properly correct for unwanted variability in sequencing data, called removal of unwanted variation (RUV) (27), we were able to capture the experimental variable. Our results caution against using normalization factors derived from total read counts, pointing to a fundamental difference between transcriptomic and epigenomic data. Overall, DEScan detected 2365 learning-regulated regions at a false discovery rate (FDR) < 0.05 (table S1). Of those, 25 are decreases and 2340 are increases, confirming that the overall effect of learning is an increase in chromatin accessibility. Learning-regulated regions were closely associated with genes, with 2019 of them either overlapping the start of a known transcript or located inside a known gene. They were disproportionately biased toward alternative (not first exon) promoters and CpG islands (Fig. 2A).

Learning-regulated regions are active during development, associated with differential expression, and disproportionately linked to known ASD risk genes

To further understand the functional characteristics of learning-regulated regions, we used all available data obtained from brain tissue samples from the ENCODE3 project to investigate whether these regions were enriched for specific epigenetic modifications. We found that learning-regulated regions showed significant enrichment for histone marks associated with promoting transcription [“active”: di- or trimethylated histone H3-Lys⁴ (H3K4me_{2/3}), acetylated histone H3-Lys⁹, and -Lys²⁷ (H3K9ac, H3K27ac)] as well as repressing transcription (“repressive”: H3K27me₃) (Fig. 2A). These are characteristics of bivalent promoters; thus, herein, we refer to learning-regulated regions as learning-regulated promoters. Enrichment for active marks H3K4me_{2/3} and H3K9ac is significantly higher in the mouse brain at embryonic day 12.5 (E12.5) than postnatally (Fig. 2B). These results suggest that promoters induced by learning are preferentially active during development. Comparing chromatin-accessible regions in the developing mouse brain at E12.5 [obtained

by assay for transposase-accessible chromatin using sequencing (ATAC-seq) through ENCODE] with chromatin-accessible regions in the mouse hippocampus as determined by our study indicated that learning-regulated promoters are also, for the most part, accessible during development (Fig. 2A).

To investigate the relationship between an increase in chromatin accessibility and differential expression after learning, we examined the overlap between our learning-regulated promoters and genome-wide gene expression data after fear conditioning we have previously defined using RNA-seq (19). The genome-wide gene expression data include gene expression measured 30 min after fear conditioning (the same time the Sono-seq samples were taken) and 30 min after a retrieval test 24 hours after. Our results showed a trend toward enrichment of learning-regulated promoters in genes differentially expressed 30 min after fear conditioning ($P = 0.12$). This enrichment is borderline significant for genes differentially expressed 30 min after retrieval ($P = 0.05$; Fig. 2C). Because there is a bias toward alternative promoters, we hypothesized that they may be driving alternative exon usage or splicing. To test this, we looked at the relationship between learning-regulated promoters and genes that show differential exon usage independently from global differential expression we previously defined (20). In this case, there was a highly significant enrichment for alternative spliced genes 30 min after fear conditioning ($P < 0.001$). This enrichment was also significant 30 min after retrieval, although to a lesser magnitude ($P = 0.01$; Fig. 2C). Overall, our data suggest that there is a relationship between chromatin organization and differential expression. The relationship is complex, in that it involves a highly significant role of learning-regulated promoters in alternative splicing as well as a moderate role in global gene expression during both memory consolidation and retrieval.

To determine whether learning-regulated promoters were associated with specific sets of genes, we conducted classical gene set enrichment analysis using the Database for Annotation Visualization and Integrated Discovery [DAVID (30)]. Only genes that directly overlapped with learning-regulated promoters were considered. We first focused on established pathways [Kyoto Encyclopedia of Genes and Genomes (KEGG)] and found that only “pathways in cancer” enriched at an FDR < 0.01 . We then looked at Gene Ontology (GO; Biological Function) and found only eight GO terms enriched at FDR < 0.01 : transcription, regulation of transcription, and positive regulation of transcription (all DNA-templated); positive and negative regulation of transcription from RNAPol II promoter; and nervous system development, in utero embryonic development, and positive regulation of neuron differentiation. Overall, learning-regulated promoters are associated with genes involved in transcription, brain development, and cancer. Given our results and the recent awareness that ASD and cancer share risk genes and pathways (31), we asked whether learning-regulated regions were preferentially associated with known risk genes for either disorder. We observed a greater than threefold enrichment of ASD risk genes in learning-regulated regions (Table 1 and table S2). This enrichment held true and was greater when the strength of evidence required to define a “risk gene” was increased. Genes known to cause syndromic forms of ASD were enriched 4.8-fold relative to learning-regulated promoters. Some ASD risk genes, such as *Rai1*, *Rere*, *Otx1*, *Dlx*, *Pcca*, and *Cacna1*, were associated with more than one learning-regulated promoter. All members of the SHANK family of

scaffolding proteins (encoded by *Shank1*, *Shank2*, and *Shank3*) were associated with learning-regulated promoters and therefore likely epi-genetically regulated by learning.

Shank3* promoter 6 is a learning-regulated promoter associated with ASD and differential isoform expression of *Shank3

To further investigate the functional relevance of learning-regulated promoters, we focused on one such region associated with a well-established ASD risk gene: *Shank3* promoter 6. SHANK3 is a synaptic scaffolding protein for which mutations have been found to be associated with a substantially increased risk for ASD (32). In humans, complete deletion of *SHANK3* leads to Phelan-McDermid syndrome (33), a rare disease strongly associated with ASD. Our Sono-seq data showed that, although the canonical first exon promoter of *Shank3* is accessible in basal conditions, accessibility of promoter 6 increased specifically in response to learning (Fig. 3A). Given that this is an internal promoter and *Shank3* has a wide array of isoforms (34), we then asked whether learning affected isoform-specific expression of *Shank3*. We repeated our contextual fear conditioning experiment, but this time, we isolated RNA and evaluated gene expression using quantitative polymerase chain reaction (qPCR). We found that learning down-regulated expression of the major isoform of *Shank3* (*Shank3a*), which mirrors the activity-dependent reduction of *Shank3a* that has been reported in vitro (Fig. 3B) (34). The mechanism that links activation of *Shank3* promoter 6 and repression of *Shank3a* is not clear at this point. To investigate whether *SHANK3* promoter 6 has any impact on ASD diagnosis, we genotyped a noncoding SNP located within this region (rs6010065) in a clinical cohort. Our results show that rs6010065 is associated with ASD (Fig. 3C). Given the uneven distribution of minor allele frequencies observed at this site based on the 1000 genomes project (www.internationalgenome.org), we split our cohort based on genetic background. We found that the association of the heterozygous allele with ASD was only significant in Caucasians (fig. S3A). In non-Caucasians, the same genotype was associated with learning ability (fig. S3B), although the sample is too small to draw definite conclusions.

DISCUSSION

Here, we identified the promoters in the mouse hippocampus altered by learning after a classical conditioning task through the use of HTS technology and the development of a new bioinformatics strategy. We found that learning increases chromatin accessibility genome-wide, supporting previous findings that neuronal activation increases chromatin accessibility (35). Our results, however, provide a genome-wide map of chromatin remodeling in response to behavior. Our data reveal a complex relationship between chromatin organization and differential expression during memory consolidation and retrieval, with a likely role in regulating alternative splicing and moderate role in regulating global levels of gene expression. We also present DEScan, a new bioinformatics strategy that allows us to define the identity of the regions recruited by learning in a reproducible way. It has long been recognized that the epigenetic mechanisms involved in development are co-opted during learning. Our results further support the conceptual parallels between learning and development. We show that the promoters recruited and the epigenetic modifications within them also show parallels between learning and brain development. The overlap between

learning and brain development is particularly significant for mouse E12.5, although the reason behind this is unclear. Studies have shown that specification of hippocampal fate in mice begins as early as E10.5 to E12.5 (36), raising the possibility that the high concordance observed between chromatin accessibility in forebrain at E12.5 and hippocampus in adulthood is rooted in similarities between epigenetic regulation due to learning and hippocampal fate determination. Overall, our data suggest that learning may recapitulate brain development at the epigenetic level by activating an overlapping subset of promoters associated with genes involved in transcription, brain development, and cancer. Learning-regulated regions are also disproportionately associated with ASD risk genes, further strengthening the link between brain development and long-term memory formation. This is not surprising, considering that intellectual disability, which is characterized by learning impairments, is the most common ASD comorbidity (37).

As whole-genome sequencing (WGS) becomes widespread, looking beyond the 2% of the genome that codes for genes and understanding the contribution of genetic risk factors in regulatory regions are essential for understanding the genetic contribution to any disorder. Work from the ENCODE consortium has already shown that most of the SNPs associated with human disease are located in regulatory regions (9). However, it remains challenging to pinpoint regulatory regions that are likely to be important for a specific disease. Here, we strengthen the conceptual parallels between learning and brain development and show that those parallels can be used to identify regulatory regions relevant to neurodevelopmental disorders, such as ASD. We go one step further and illustrate how that information can be exploited to identify a noncoding variant linked to ASD (rs6010065 located in *SHANK3* promoter 6). Still, an open question is how the observed changes in chromatin accessibility ultimately translate into differences in gene expression that can affect phenotype. The challenges in that regard are twofold. First, changes in gene expression are expected to follow epigenetic changes, but at which point in time is uncertain and that timeline could be different for each regulatory region and gene. Second, because the changes in chromatin accessibility disproportionately affect alternative promoters, they are likely to affect isoform-specific expression (as illustrated in our *Shank3* example), which is notoriously harder to quantify reliably genome-wide. With the continued WGS of ASD individuals, future studies will evaluate whether learning-regulated regions can identify novel regulatory genetic variants as well as explore the potential role of those variants in learning ability within ASD.

MATERIALS AND METHODS

Mice and behavior

C57BL/6J mice were maintained under standard conditions, with food and water available ad libitum. Adult male mice 3 months of age were kept on a 12-hour light/12-hour dark cycle with lights on at 7 a.m. All behavioral and biochemical experiments were performed during the light cycle, with training starting at about 10 a.m. (Zeitgeber time 3). All procedures were approved by the University of Pennsylvania Institutional Animal Care and Use Committee. Contextual fear conditioning was performed as previously described (38). Briefly, mice were handled for 3 days before training, placed in a novel context within soundproof chambers, and allowed to explore for 148 s. A 2-s 1.5-mA footshock was

delivered, and mice were kept in the novel context for an additional 30 s. Mice were removed to their home cage after training. One mouse was trained and dissected per day to allow for dissection to occur at the same circadian time every day.

DNA isolation and sequencing

Thirty minutes after fear conditioning, mice were sacrificed and the hippocampus was rapidly dissected. The hippocampus was finely chopped and placed into 500- μ l 2% paraformaldehyde for 10 min to cross-link tissue. Cross-linking was stopped by the addition of 100- μ l 1 M glycine, and tissue was washed 3 \times in ice-cold phosphate-buffered saline containing protease inhibitor cocktail (Sigma-Aldrich) and frozen at -80°C . Chromatin was prepared by dounce-homogenizing the tissue in 1 ml of cell lysis buffer [10 mM tris-HCl (pH 8.1), 10 mM NaCl, 3 mM MgCl_2 , and 0.5% NP-40), centrifuging at 5500g for 5 min at 4°C , removing the supernatant, and resuspending the pellet in 300 μ l of chromatin immunoprecipitation (ChIP) nuclear lysis buffer [50 mM tris (pH 8.1), 5 mM EDTA, and 1% SDS]. Chromatin was sonicated using a Bioruptor sonicator with the following conditions: 30 s on and 1 min off for 15 min, add ice and cool the sonicator, and 30 s on and 1 min off for another 15 min. Chromatin was centrifuged for 5 min at 14000g at 4°C , and the supernatant was removed. Chromatin size was tested by removing an aliquot, reversing cross-links, cleaning DNA using the Qiagen MinElute PCR purification kit (Qiagen), quantifying by NanoDrop (Thermo Fisher Scientific), and running the DNA on a 1% agarose gel. Sono-seq DNA was prepared by removing 1 μ g of sonicated chromatin. The next morning, 4- μ l 0.5 M EDTA, 8- μ l 1 M tris-HCl (pH 7.5), and 1- μ l Proteinase K (Roche Diagnostics) were added, and samples were incubated for an additional hour at 65°C . Samples were purified into 30 μ l using the Qiagen MinElute PCR purification kit. Library preparation for Sono-seq samples was performed using a combination of the ChIP sequencing (ChIP-seq) DNA sample prep kit and the Multiplexing Sample Preparation Oligonucleotide Kit (Illumina). End repair and adenylation were performed as described by the ChIP protocol, whereas adapter ligation was performed as described by the multiplexing protocol. A gel slice 300 ± 25 base pair (bp) was cut from the gel for size selection, the Qiagen gel extraction kit was used to purify DNA, and 23 PCR cycles were used after the PCR reactions described by the multiplexing kit. The final product was run on a gel and purified into 15 μ l by the Qiagen gel extraction kit again. Sequencing libraries were quantified using the KAPA Library Quantification Kit, normalized to 10 nM, and submitted to the Penn Genome Frontiers Institute sequencing core at the University of Pennsylvania for 100-bp single-end sequencing on an Illumina HiSeq2000. The average sequence depth was 148 million reads, four independent biological replicates per condition.

HTS data analysis

Reads were mapped to the mouse genome (mm9) using Bowtie2 with default parameters (39). Mapped reads were filtered for quality using custom scripts. Reads were required to have a Bowtie2 mapping quality alignment score greater than 2 to limit multimapping reads. To avoid artifacts from clonal PCR amplification, identical reads (three or more reads of the same sequence) were removed if the surrounding reads did not also show high replication. This strategy was adopted to avoid removing identical reads that may result from repeated sampling of highly accessible chromatin regions, a possible outcome of the high level of

sequencing depth used in this study. Details of sequencing, mapping, and filtering statistics are available in table S1. TSS plots were generated using NGSplot (40) of aggregate read counts over all replicates per condition. To analyze differential enrichment of Sono-seq signal between fear conditioned and home cage controls, we developed a novel bioinformatics strategy using the R statistical language we named “DESCan” (fig. S1). DESCan is optimized to perform best on epigenetic HTS data that are not defined by genomic sequence and produce broad peaks. DESCan performs peak calling in individual replicates based on a Poisson distribution using an adaptive window scan and the surrounding 5 or 10 kb of sequence as a background. Peaks in individual replicates are then aligned, and only genomic regions with a defined number of replicates containing a peak (“carriers”) are kept for further analysis. The union of the coordinates of the replicated peaks is used to set the final peak coordinates. The peak-calling module of DESCan cannot be directly compared to currently available peak-caller software for three reasons. First, DESCan was designed to accommodate epigenomic data that produce both narrow (transcription factor) and broad (chromatin accessibility and histone acetylation) signal. Hence, it implements a scan of adaptive width, not a fixed window scan as all other peak finders. Second, other peak-calling algorithms usually rely on “input” as background sample to determine significance for peak calling. DESCan does not require an input sample and calculates peak significance based on the surrounding genomic sequence (Sono-seq and input are also essentially the same data set). Although this approach could lead to increased false positives, peaks on individual samples are never considered in isolation of other biological replicates in DESCan. The last and most important difference is that DESCan requires biological replicates. It considers the reproducibility of peaks in both location and significance simultaneously. Because no current peak caller explicitly accounts for reproducibility, the results are not directly comparable. Replicated genomic regions are used to derive a count matrix using featureCounts (41). The differential module of DESCan implements a strategy similar to that used for RNA-seq differential expression analysis. In that sense, it is analogous to other available approaches used to estimate differential epigenomic regulation, with one important difference: normalization. Normalization is performed using RUV (27). We have previously shown in the context of RNA-seq that RUV normalization is essential for the accuracy and reproducibility of differential analysis for the types of data sets described in this article (19). After normalization, testing for differential enrichment between fear-conditioned and control animals was performed using edgeR (42). The results of differential Sono-seq analysis reported in this article were obtained using the following DESCan parameters: adaptive window size, 50 to 500 bp; number of carriers, 4; and FDR > 0.01. The choice of a proper k for RUV normalization was determined as previously described (19), resulting in $k = 4$. The DESCan manual is available in file S1.

Analysis of overlap with genomic regions was performed using mm10 ChIP-seq and ATAC-seq data derived from forebrain tissue available from ENCODE3 and CpG island and alternative promoter annotations obtained from the University of California, Santa Cruz (UCSC) genome browser. Coordinate differences between mm9 and mm10 mouse genome version were accounted for by using UCSC liftOver. Significance of the number of overlapping regions with his-tone marks, ATAC-seq, CpG islands, and alternative promoters was assessed through permutation testing (learning-regulated regions compared to 10,000

rounds of random resampling from replicated Sono-seq peaks) using the R/Bioconductor package *regionR*. Comparisons between developmental stages were performed using ENCODE E12.5 and postnatal day 0 ChIP-seq data derived from forebrain, midbrain, and hindbrain. Significance of overlap with differentially expressed or spliced genes was assessed through permutation testing (learning- regulated regions compared to 10,000 rounds of random resampling from all genes expressed in the hippocampus). Comparisons to genes implicated in disease were performed using data sets obtained from the Simons Foundation Autism Research Initiative Human Gene list (881 total genes, 7 October 2017, updated version) and the Catalogue of Somatic Mutations in Cancer cancer gene census (561 total genes, version 82). Human HUGO Gene Nomenclature Committee symbols were then translated to Ensembl mouse gene IDs using the *getLDS* function from the R/Bioconductor *biomaRt* package, resulting in 880 and 573 mouse genes, respectively. Learning-regulated regions were annotated with the Ensembl gene ID of the nearest TSS using the R/Bioconductor package *ChIPpeakAnno*. Functional annotation was based on Ensembl gene IDs and performed using the database for annotation, visualization, and integrated discovery version 6.7 (DAVID, <https://david.ncifcrf.gov>). The following functional categories were used: GO biological process and KEGG pathways. Enrichment cutoff relative to background (mouse genome) = FDR < 0.01.

Quantitative reverse transcription PCR

qPCR was performed as previously described (43). RNA extraction was performed using a Qiagen RNeasy lipid tissue kit with the following modifications: RNA precipitation was performed using 100% ethanol, and washes were performed using Qiagen's RWT buffer. Concentration and purity were quantified by NanoDrop spectrophotometry (Thermo Fisher Scientific). For quantitative real-time reverse transcription PCR (RT-PCR), reactions were prepared in 384-well optical reaction plates (ABI) with optical adhesive covers (ABI). Two technical replicates were used. Reactions were carried out in ViiA7 real-time PCR system (Invitrogen) according to the manufacturer's instructions. For mRNA qPCR, generation of complementary DNA was carried out by the RETROscript Kit (Ambion) with 1 μ g of RNA as template. Primer sequences for *Shank3* isoform genes were obtained from (44). Data were normalized to *Hprt* before calculation of differences. Fold change was calculated using the C_t method (45). The data presented are the calculated mean for the biological replicates with $n = 8$ (that is, the number of mice examined). We used t tests to compare fold change values for each gene in each comparison of interest. Two-tailed P values are reported. Fos induction was used as a positive control on all qPCR runs.

Participants

Our data set consisted of 768 individuals, including both male (678) and female (90) participants, among which 554 children (age, 3.3 to 15.2 years) were affected by ASD and 214 children (age, 3.2 to 16.3 years) were typically developing controls (TDCs). Participants were recruited and diagnosed by expert clinicians at the Center for Autism Research at the Children's Hospital of Philadelphia (CHOP). Children with known genetic conditions associated with ASD were excluded from the study. DNA samples were collected through cheek swabs and/or blood draw. All procedures were approved by the Institutional Review Board at CHOP.

Patient genotyping

DNA extraction was carried out using a detergent-based method using AMPure beads automated on a Biomek FX liquid handler instrument from Beckman Coulter. Genotyping of rs6010065 was performed using TaqMan assay from Life Technologies (catalog no. C__29723448_10). Reactions were carried out in ViiA7 real-time PCR system (Invitrogen) according to the manufacturer's instructions. On the basis of data from the 1000 genome project (www.internationalgenome.org), we expected the minor allele for rs6010065 to be different in Caucasian populations relative to non-Caucasians. Accordingly, Hardy-Weinberg equilibrium was not achieved whether all individuals were analyzed together, but it was achieved whether non-Caucasians and Caucasians were analyzed separately. Self-reported ethnicity was used to stratify patients based on genetic background. Patients that did not report ethnicity were excluded from further studies. Final sample sizes were as follows: Caucasian, $n = 302$ (ASD) and $n = 125$ (TDCs); non-Caucasian: $n = 67$ (ASD) and $n = 46$ (TDCs). Tests of association were performed as χ^2 test on 2×2 contingency tables per genotype relative to the sum of all other genotypes. P values reported are one-tailed.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments:

We thank H. Schoch for helpful discussions. **Funding:** This work was supported by National Research Service Award training grants T32NS007413 [to L.P.; M. Robinson, principal investigator (PI)] and T32HL007953 (to M.E.W.; A.I. Pack, PI). T.A. acknowledges funding from the Brush Family Professorship, Defense Advanced Research Projects Agency 58077 LSDRP (S. Bhatnagar, PI), and R01 MH087463. H.H. and R.T.S. acknowledge funding from the Lurie Family Foundation. R.T.S. acknowledges funding from the Pennsylvania Department of Health (Student Assistance Program no. 4100047863), U54 HD86984, and RC1 MH08879.

REFERENCES AND NOTES

- Pinto D, Pagnamenta AT, Klei L, Anney R, Merico D, Regan R, Conroy J, Magalhaes TR, Correia C, Abrahams BS, Almeida J, Bacchelli E, Bader GD, Bailey AJ, Baird G, Battaglia A, Berney T, Bolshakova N, Bölte S, Bolton PF, Bourgeron T, Brennan S, Brian J, Bryson SE, Carson AR, Casallo G, Casey J, Chung BHY, Cochrane L, Corsello C, Crawford EL, Crossett A, Cytrynbaum C, Dawson G, de Jonge M, Delorme R, Drmic I, Duketis E, Duque F, Estes A, Farrar P, Fernandez BA, Folstein SE, Fombonne E, Freitag CM, Gilbert J, Gillberg C, Glessner JT, Goldberg J, Green A, Green J, Guter SJ, Hakonarson H, Heron EA, Hill M, Holt R, Howe JL, Hughes G, Hus V, Iglizzi R, Kim C, Klauck SM, Kolevzon A, Korvatska O, Kustanovich V, Lajonchere CM, Lamb JA, Laskawiec M, Leboyer M, Le Couteur A, Leventhal BL, Lionel AC, Liu X-Q, Lord C, Lotspeich L, Lund SC, Maestrini E, Mahoney W, Mantoulan C, Marshall CR, McConachie H, McDougale CJ, McGrath J, McMahan WM, Merikangas A, Migita O, Minshew NJ, Mirza GK, Munson J, Nelson SF, Noakes C, Noor A, Nygren G, Oliveira G, Papanikolaou K, Parr JR, Parrini B, Paton T, Pickles A, Pilorge M, Piven J, Ponting CP, Posey DJ, Poustka A, Poustka F, Prasad A, Ragoussis J, Renshaw K, Rickaby J, Roberts W, Roeder K, Roge B, Rutter ML, Bierut LJ, Rice JP, Salt J, Sansom K, Sato D, Segurado R, Sequeira AF, Senman L, Shah N, Sheffield VC, Soorya L, Sousa I, Stein O, Sykes N, Stoppioni V, Strawbridge C, Tancredi R, Tansey K, Thiruvahindrapuram B, Thompson AP, Thomson S, Tryfon A, Tsiantis J, Van Engeland H, Vincent JB, Volkmar F, Wallace S, Wang K, Wang Z, Wassink TH, Webber C, Weksberg R, Wing K, Wittemeyer K, Wood S, Wu J, Yaspan BL, Zurawiecki D, Zwaigenbaum L, Buxbaum JD, Cantor RM, Cook EH, Coon H, Cuccaro ML, Devlin B, Ennis S, Gallagher L, Geschwind DH, Gill M, Haines JL, Hallmayer J, Miller J, Monaco AP, Nurnberger JI, Jr., Paterson AD, Pericak-Vance MA, Schellenberg GD, Szatmari P,

- Vicente AM, Vieland VJ, Wijsman EM, Scherer SW, Sutcliffe JS, Betancur C, Functional impact of global rare copy number variation in autism spectrum disorders. *Nature* 466, 368–372 (2010). [PubMed: 20531469]
2. Glessner JT, Wang K, Cai G, Korvatska O, Kim CE, Wood S, Zhang H, Estes A, Brune CW, Bradfield JP, Imielinski M, Frackelton EC, Reichert J, Crawford EL, Munson J, Sleiman PMA, Chiavacci R, Annaiah K, Thomas K, Hou C, Glaberson W, Flory J, Otieno F, Garris M, Soorya L, Klei L, Piven J, Meyer KJ, Anagnostou E, Sakurai T, Game RM, Rudd DS, Zurawiecki D, McDougle CJ, Davis LK, Miller J, Posey DJ, Michaels S, Kolevzon A, Silverman JM, Bernier R, Levy SE, Schultz RT, Dawson G, Owley T, McMahon WM, Wassink TH, Sweeney JA, Nurnberger JI, Coon H, Sutcliffe JS, Minshew NJ, Grant SFA, Bucan M, Cook EH, Buxbaum JD, Devlin B, Schellenberg GD, Hakonarson H, Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. *Nature* 459, 569–573 (2009). [PubMed: 19404257]
 3. Wang K, Zhang H, Ma D, Bucan M, Glessner JT, Abrahams BS, Salyakina D, Imielinski M, Bradfield JP, Sleiman PMA, Kim CE, Hou C, Frackelton E, Chiavacci R, Takahashi N, Sakurai T, Rappaport E, Lajonchere CM, Munson J, Estes A, Korvatska O, Piven J, Sonnenblick LI, Alvarez Retuerto AI, Herman EI, Dong H, Hutman T, Sigman M, Ozonoff S, Klin A, Owley T, Sweeney JA, Brune CW, Cantor RM, Bernier R, Gilbert JR, Cuccaro ML, McMahon WM, Miller J, State MW, Wassink TH, Coon H, Levy SE, Schultz RT, Nurnberger JI, Haines JL, Sutcliffe JS, Cook EH, Minshew NJ, Buxbaum JD, Dawson G, Grant SFA, Geschwind DH, Pericak-Vance MA, Schellenberg GD, Hakonarson H, Common genetic variants on 5p14.1 associate with autism spectrum disorders. *Nature* 459, 528–533 (2009). [PubMed: 19404256]
 4. Allen-Brady K, Robison R, Cannon D, Varvil T, Villalobos M, Pingree C, Leppert MF, Miller J, McMahon WM, Coon H, Genome-wide linkage in Utah autism pedigrees. *Mol. Psychiatry* 15, 1006–1015 (2010). [PubMed: 19455147]
 5. Marshall CR, Noor A, Vincent JB, Lionel AC, Feuk L, Skaug J, Shago M, Moessner R, Pinto D, Ren Y, Thiruvahindrapuram B, Fiebig A, Schreiber S, Friedman J, Ketelaars CEJ, Vos YJ, Ficicioglu C, Kirkpatrick S, Nicolson R, Sloman L, Summers A, Gibbons CA, Teebi A, Chitayat D, Weksberg R, Thompson A, Vardy C, Crosbie V, Luscombe S, Baatjes R, Zwaigenbaum L, Roberts W, Fernandez B, Szatmari P, Scherer SW, Structural variation of chromosomes in autism spectrum disorder. *Am. J. Hum. Genet* 82, 477–488 (2008). [PubMed: 18252227]
 6. O’Roak BJ, Vives L, Girirajan S, Karakoc E, Krumm N, Coe BP, Levy R, Ko A, Lee C, Smith JD, Turner EH, Stanaway IB, Vernet B, Malig M, Baker C, Reilly B, Akey JM, Borenstein E, Rieder MJ, Nickerson DA, Bernier R, Shendure J, Eichler EE, Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature* 485, 246–250 (2012). [PubMed: 22495309]
 7. Sanders SJ, Murtha MT, Gupta AR, Murdoch JD, Raubeson MJ, Willsey AJ, Ercan-Sencicek AG, DiLullo NM, Parikshak NN, Stein JL, Walker MF, Ober GT, Teran NA, Song Y, El-Fishawy P, Murtha RC, Choi M, Overton JD, Bjornson RD, Carriero NJ, Meyer KA, Bilguvar K, Mane SM, Šestan N, Lifton RP, Günel M, Roeder K, Geschwind DH, Devlin B, State MW, De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* 485, 237–241 (2012). [PubMed: 22495306]
 8. Neale BM, Kou Y, Liu L, Ma’ayan A, Samocha KE, Sabo A, Lin C-F, Stevens C, Wang L-S, Makarov V, Polak P, Yoon S, Maguire J, Crawford EL, Campbell NG, Geller ET, Valladares O, Schafer C, Liu H, Zhao T, Cai G, Lihm J, Dannenfelser R, Jabado O, Peralta Z, Nagaswamy U, Muzny D, Reid JG, Newsham I, Wu Y, Lewis L, Han Y, Voight BF, Lim E, Rossin E, Kirby A, Flannick J, Fromer M, Shakir K, Fennell T, Garimella K, Banks E, Poplin R, Gabriel S, DePristo M, Wimbish JR, Boone BE, Levy SE, Betancur C, Sunyaev S, Boerwinkle E, Buxbaum JD, Cook EH, Jr., Devlin B, Gibbs RA, Roeder K, Schellenberg GD, Sutcliffe JS, Daly MJ, Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature* 485, 242–245 (2012). [PubMed: 22495311]
 9. ENCODE Project Consortium, An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57–74 (2012). [PubMed: 22955616]
 10. Rockman MV, Wray GA, Abundant raw material for *cis*-regulatory evolution in humans. *Mol. Biol. Evol* 19, 1991–2004 (2002). [PubMed: 12411608]
 11. Lepagnol-Bestel A-M, Maussion G, Boda B, Cardona A, Iwayama Y, Delezoide A-L, Moalic J-M, Muller D, Dean B, Yoshikawa T, Gorwood P, Buxbaum JD, Ramoz N, Simonneau M, SLC25A12

- expression is associated with neurite outgrowth and is upregulated in the prefrontal cortex of autistic subjects. *Mol. Psychiatry* 13, 385–397 (2008). [PubMed: 18180767]
12. Campbell DB, Sutcliffe JS, Ebert PJ, Militeri R, Bravaccio C, Trillo S, Elia M, Schneider C, Melmed R, Sacco R, Persico AM, Levitt P, A genetic variant that disrupts *MET* transcription is associated with autism. *Proc. Natl. Acad. Sci. U.S.A.* 103, 16834–16839 (2006). [PubMed: 17053076]
 13. Kantojärvi K, Oikkonen J, Kotala I, Kallela J, Vanhala R, Onkamo P, Järvelä I, Association and promoter analysis of AVPR1A in Finnish autism families. *Autism Res* 8, 634–639 (2015). [PubMed: 25707602]
 14. Melke J, Goubran Botros H, Chaste P, Betancur C, Nygren G, Anckarsäter H, Rastam M, Ståhlberg O, Gillberg IC, Delorme R, Chabane N, Mouren-Simeoni M-C, Fauchereau F, Durand CM, Chevalier F, Drouot X, Collet C, Launay J-M, Leboyer M, Gillberg C, Bourgeron T, Abnormal melatonin synthesis in autism spectrum disorders. *Mol. Psychiatry* 13, 90–98 (2008). [PubMed: 17505466]
 15. Davis LK, Gamazon ER, Kistner-Griffin E, Badner JA, Liu C, Cook EH, Sutcliffe JS, Cox NJ, Loci nominally associated with autism from genome-wide analysis show enrichment of brain expression quantitative trait loci but not lymphoblastoid cell line expression quantitative trait loci. *Mol. Autism* 3, 3 (2012). [PubMed: 22591576]
 16. Day JJ, Sweatt JD, Epigenetic mechanisms in cognition. *Neuron* 70, 813–829 (2011). [PubMed: 21658577]
 17. Auerbach RK, Euskirchen G, Rozowsky J, Lamarre-Vincent N, Moqtaderi Z, Lefrançois P, Struhl K, Gerstein M, Snyder M, Mapping accessible chromatin regions using Sono-Seq. *Proc. Natl. Acad. Sci. U.S.A.* 106, 14926–14931 (2009). [PubMed: 19706456]
 18. Voigt P, Tee W-W, Reinberg D, A double take on bivalent promoters. *Genes Dev.* 27, 1318–1338 (2013). [PubMed: 23788621]
 19. Peixoto L, Risso D, Poplawski SG, Wimmer ME, Speed TP, Wood MA, Abel T, How data analysis affects power, reproducibility and biological insight of RNA-seq studies in complex datasets. *Nucleic Acids Res.* 43, 7664–7674 (2015). [PubMed: 26202970]
 20. Poplawski SG, Peixoto L, Porcari GS, Wimmer ME, McNally AG, Mizuno K, Giese KP, Chatterjee S, Koberstein JN, Risso D, Speed TP, Abel T, Contextual fear conditioning induces differential alternative splicing. *Neurobiol. Learn. Mem* 134 Pt B, 221–235 (2016). [PubMed: 27451143]
 21. Agranoff BW, Davis RE, Casola L, Lim R, Actinomycin D blocks formation of memory of shock-avoidance in goldfish. *Science* 158, 1600–1601 (1967). [PubMed: 6060370]
 22. Flood JF, Rosenzweig MR, Bennett EL, Orme AE, The influence of duration of protein synthesis inhibition on memory. *Physiol. Behav* 10, 555–562 (1973). [PubMed: 4736141]
 23. Zovkic IB, Guzman-Karlsson MC, Sweatt JD, Epigenetic regulation of memory formation and maintenance. *Learn. Mem* 20, 61–74 (2013). [PubMed: 23322554]
 24. Peixoto L, Abel T, The role of histone acetylation in memory formation and cognitive impairments. *Neuropsychopharmacology* 38, 62–76 (2013). [PubMed: 22669172]
 25. Nelson ED, Monteggia LM, Epigenetics in the mature mammalian brain: Effects on behavior and synaptic transmission. *Neurobiol. Learn. Mem* 96, 53–60 (2011). [PubMed: 21396474]
 26. Alberini CM, Transcription factors in long-term memory and synaptic plasticity. *Physiol. Rev* 89, 121–145 (2009). [PubMed: 19126756]
 27. Bourchouladze R, Abel T, Berman N, Gordon R, Lapidus K, Kandel ER, Different training procedures recruit either one or two critical periods for contextual memory consolidation, each of which requires protein synthesis and PKA. *Learn. Mem* 5, 365–374 (1998). [PubMed: 10454361]
 28. Igaz LM, Vianna MRM, Medina JH, Izquierdo I, Two time periods of hippocampal mRNA synthesis are required for memory consolidation of fear-motivated learning. *J. Neurosci* 22, 6781–6789 (2002). [PubMed: 12151558]
 29. Risso D, Ngai J, Speed TP, Dudoit S, Normalization of RNA-seq data using factor analysis of control genes or samples. *Nat. Biotechnol* 32, 896–902 (2014). [PubMed: 25150836]

30. Dennis G, Jr., Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA, DAVID: Database for annotation, visualization, and integrated discovery. *Genome Biol.* 4, P3 (2003). [PubMed: 12734009]
31. Crawley JN, Heyer W-D, LaSalle JM, Autism and cancer share risk genes, pathways, and drug targets. *Trends Genet.* 32, 139–146 (2016). [PubMed: 26830258]
32. Moessner R, Marshall CR, Sutcliffe JS, Skaug J, Pinto D, Vincent J, Zwaigenbaum L, Fernandez B, Roberts W, Szatmari P, Scherer SW, Contribution of *SHANK3* mutations to autism spectrum disorder. *Am. J. Hum. Genet* 81, 1289–1297 (2007). [PubMed: 17999366]
33. Phelan K, McDermid HE, The 22q13.3 deletion syndrome (Phelan-McDermid syndrome). *Mol. Syndromol* 2, 186–201 (2012). [PubMed: 22670140]
34. Wang X, Xu Q, Bey AL, Lee Y, Jiang Y.-h., Transcriptional and functional complexity of Shank3 provides a molecular framework to understand the phenotypic heterogeneity of SHANK3 causing autism and Shank3 mutant mice. *Mol. Autism* 5, 30 (2014). [PubMed: 25071925]
35. Su Y, Shin J, Zhong C, Wang S, Roychowdhury P, Lim J, Kim D, G.-l. Ming, H. Song, Neuronal activity modifies the chromatin accessibility landscape in the adult brain. *Nat. Neurosci* 20, 476–483 (2017). [PubMed: 28166220]
36. Grove EA, Tole S, Patterning events and specification signals in the developing hippocampus. *Cereb. Cortex* 9, 551–561 (1999). [PubMed: 10498273]
37. Matson JL, Shoemaker M, Intellectual disability and its relationship to autism spectrum disorders. *Res. Dev. Disabil* 30, 1107–1114 (2009). [PubMed: 19604668]
38. Peixoto LL, Wimmer ME, Poplawski SG, Tudor JC, Kenworthy CA, Liu S, Mizuno K, Garcia BA, Zhang NR, Giese KP, Abel T, Memory acquisition and retrieval impact different epigenetic processes that regulate gene expression. *BMC Genomics* 16 (Suppl. 5), S5 (2015).
39. Langmead B, Salzberg SL, Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359 (2012). [PubMed: 22388286]
40. Shen L, Shao N, Liu X, Nestler E, ngs.plot: Quick mining and visualization of next-generation sequencing data by integrating genomic databases. *BMC Genomics* 15, 284 (2014). [PubMed: 24735413]
41. Liao Y, Smyth GK, Shi W, featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30, 923–930 (2014). [PubMed: 24227677]
42. Robinson MD, McCarthy DJ, Smyth GK, edgeR: A bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140 (2010). [PubMed: 19910308]
43. Vecsey CG, Peixoto L, Choi JHK, Wimmer M, Jaganath D, Hernandez PJ, Blackwell J, Meda K, Park AJ, Hannenhalli S, Abel T, Genomic analysis of sleep deprivation reveals translational regulation in the hippocampus. *Physiol. Genomics* 44, 981–991 (2012). [PubMed: 22930738]
44. Wang X, McCoy PA, Rodriguiz RM, Pan Y, Je HS, Roberts AC, Kim CJ, Berrios J, Colvin JS, Bousquet-Moore D, Lorenzo I, Wu G, Weinberg RJ, Ehlers MD, Philpot BD, Beaudet AL, Wetsel WC, Jiang Y.-h., Synaptic dysfunction and abnormal behaviors in mice lacking major isoforms of *Shank3*. *Hum. Mol. Genet* 20, 3093–3108 (2011). [PubMed: 21558424]
45. Schmittgen TD, Livak KJ, Analyzing real-time PCR data by the comparative CT method. *Nat. Protoc* 3, 1101–1108 (2008). [PubMed: 18546601]
46. Jiang Y.-h., Ehlers MD, Modeling autism by SHANK gene mutations in mice. *Neuron* 78, 8–27 (2013). [PubMed: 23583105]

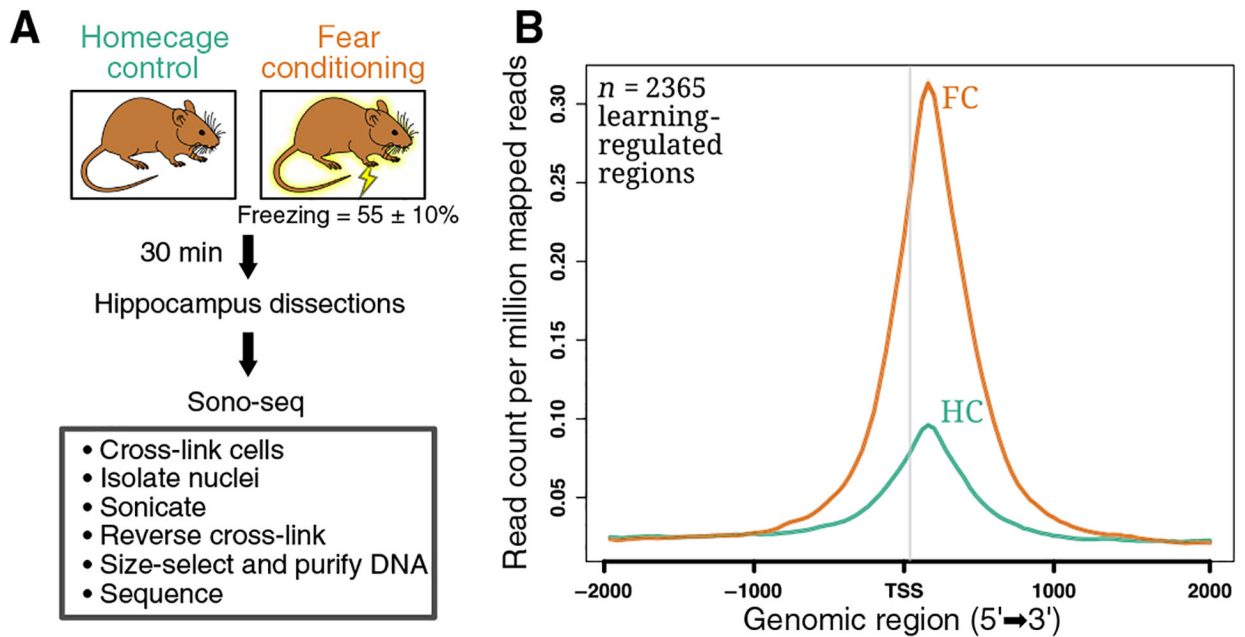


Fig. 1. Learning increases chromatin accessibility in the mouse hippocampus.

(A) Outline of experimental design to measure changes in chromatin accessibility after contextual fear conditioning (FC) versus time of day matched controls (HC) using sonication of cross-linked chromatin followed by sequencing (Sono-seq) (B) The effect of learning in chromatin accessibility at transcriptional start sites (TSS) genome-wide. The plot shows normalized read counts relative to genomic position around TSS for all promoters in the mouse genome (mm9). Average per condition is shown ($n = 4$ samples per group). Average sequencing depth per sample: 100 million 100-bp single-end reads. The number of learning-regulated regions detected at false discovery rate (FDR) < 0.05 requiring four replicates to reproduce the peak using DESeq is shown (see figs. S1 and S2 and table S1).

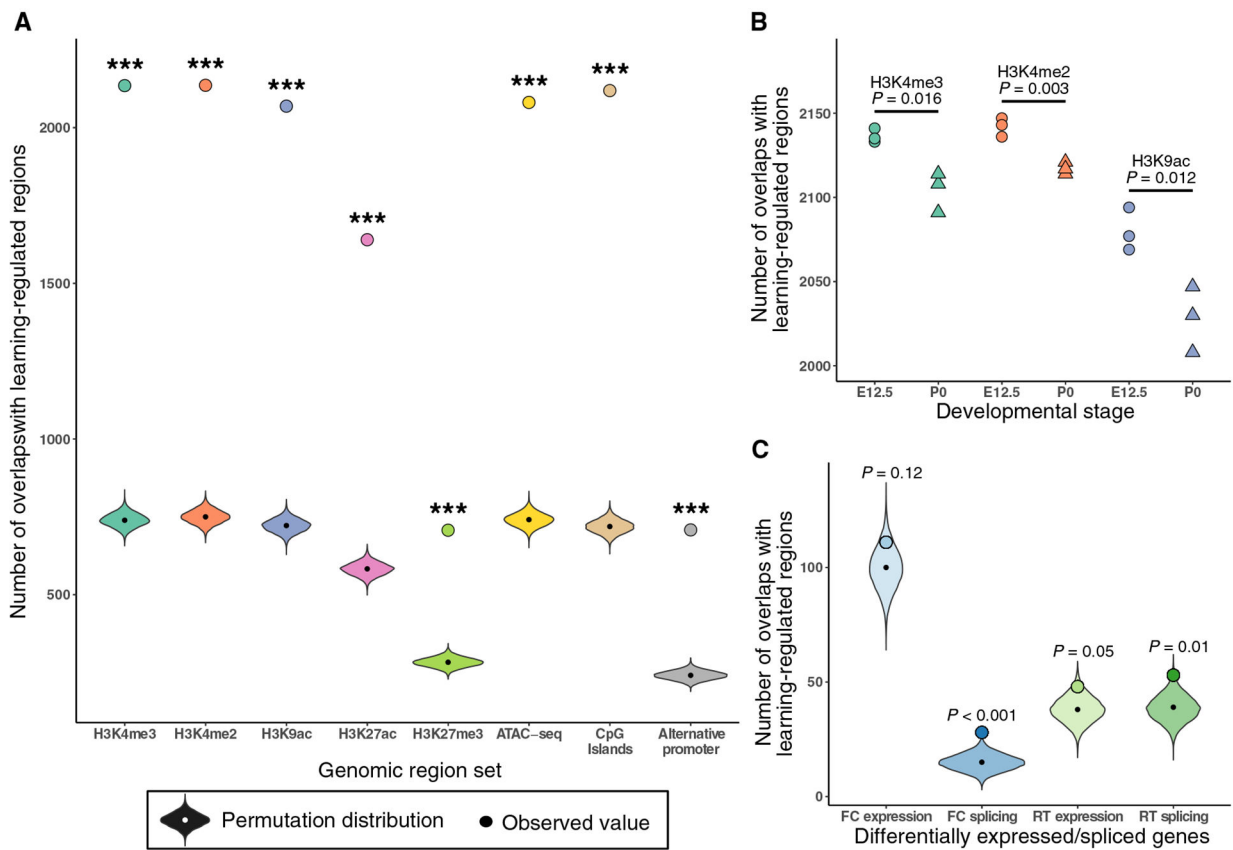


Fig. 2. Learning recapitulates development at the epigenetic level and is linked to changes in gene expression.

(A) Overlap of learning-regulated regions relative to regions of accessible chromatin during development (ATAC-seq), CpG islands, alternative promoters, and histone marks associated with active (H3K4me2/3, H3K9ac, and H3K27ac) or inactive, repressed (H3K27me3) transcription. Permutation testing was performed to evaluate statistical significance. The number of overlaps observed for each set with the learning-regulated regions was compared to overlaps with 10,000 random samplings of the same number of regions from a background set of all Sono-seq peaks present in two or more samples. The P value corresponds to the proportion of permuted samples with more overlaps than the observed value; *** $P < 0.001$, where no permutation showed greater overlap than the observed value. Violin plots display the full distribution of permuted set overlaps with each genomic region data set, whereas circles indicate the observed value of overlaps with learning-regulated promoters. All ChIP-seq and ATAC-seq data obtained from ENCODE are derived from mouse forebrain at E12.5. CpG island and alternative promoter (altEvents) annotations obtained from the UCSC genome browser. (B) Overlap of learning-regulated regions with active histone marks in the forebrain, midbrain, and hindbrain at E12.5 relative to the overlap in the same tissues at postnatal day 0 (P0). (C) Enrichment of learning-regulated promoters relative to gene expression measured using RNA-seq after FC and retrieval of memory 24 hours after (RT), assessed as the number of overlaps between the learning-regulated regions and each set of differentially expressed or spliced set of genes (FDR < 0.05) compared to overlaps with 10,000 random samplings of genes expressed in the

hippocampus. *P* value determined as described in (A). Violin plots display the full distribution of permuted set overlaps with each genomic region data set, whereas circles indicate the observed value of overlaps.

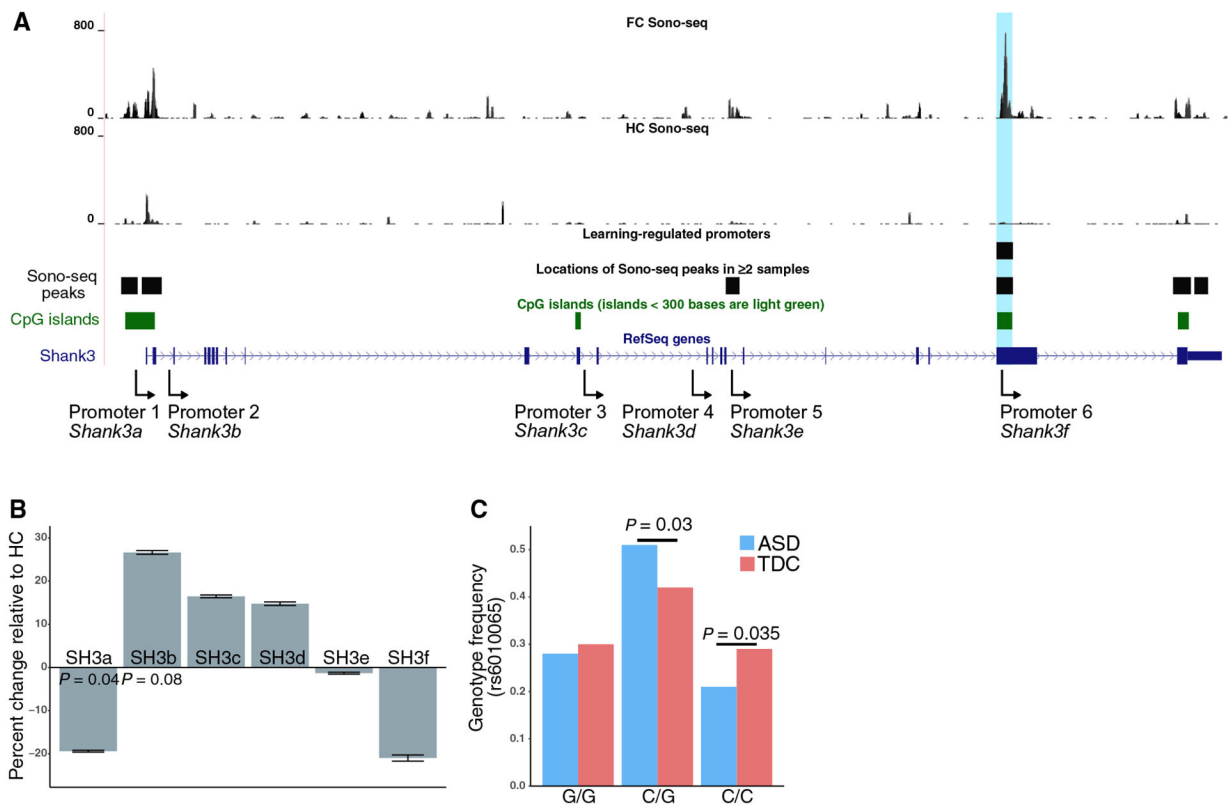


Fig. 3. Learning increases accessibility to *Shank3* promoter 6, highlighting a regulatory region linked to ASD.

(A) Chromatin accessibility at the *Shank3* locus. Sono-seq signals (mapped reads) from representative samples are shown as mapped to the UCSC genome browser (mm9) for both fear-conditioned (FC) and control (HC) animals. Dark blue track shows UCSC gene model for *Shank3*, with filled rectangles indicating the 22 exons. The light blue region highlights a region of significant enrichment compared to controls as detected by DEScan ($n = 4$ per group, FDR < 0.05). Green boxes represent CpG islands. The promoters of *Shank3* are shown for reference, adapted from (46). (B) Differential expression (by qPCR) of major *Shank3* isoforms after fear conditioning in an independent cohort of animals, normalized to *Hprt* expression and quantified using the C_t method. Data are mean \pm SE from $n = 7$ animals per condition. Only $P < 0.1$ is shown; by two-sided t tests. Primer pairs used to quantify *Shank3* isoform expression were designed to span *Shank3a* exons 2 to 10, *Shank3b* exons 3 to 6, *Shank3c* exons 12 to 14, *Shank3d* exons 13 to 15, *Shank3e* exons 17 to 19, and *Shank3f* exon 22. (C) Association of genetic variation in promoter 6 of *SHANK3* (SNP rs6010065) with ASD (blue, $n = 422$ patients) versus TDCs (red, $n = 182$ individuals) as assessed by the χ^2 test per genotype relative to the sum of all other genotypes. SNP rs6010065 does not overlap *SHANK3* exon 21. Further details on genotyping of patients are in fig. S3.

Learning-regulated regions are disproportionately associated with known ASD and cancer candidate genes. ASD candidate genes were obtained from the Simons Foundation Autism Research Initiative (SFARI) gene database. Cancer candidate genes were obtained from <http://cancer.sanger.ac.uk/> **census.** Statistical significance was calculated using χ^2 tests comparing observed versus expected frequencies. ASD genes associated with learning-regulated promoters can be found in table S2.

Table 1.

Gene sets	Disease genes/genes associated with learning-regulated regions (observed)	Disease genes/all genes (expected)	Fold enrichment	P value
ASD genes (all SFARI)	127 of 2279	880 of 53,110	3.4	<0.001
ASD genes (SFARI score 4)	81 of 2279	519 of 53,110	3.6	<0.001
ASD genes (syndromic)	20 of 2279	98 of 53,110	4.8	<0.001
Cancer genes	105 of 2279	573 of 53,110	4.3	<0.001