

ORIGINAL ARTICLE

The autism-associated gene chromodomain helicase DNA-binding protein 8 (*CHD8*) regulates noncoding RNAs and autism-related genes

B Wilkinson¹, N Grepo², BL Thompson³, J Kim^{2,4}, K Wang^{2,4}, OV Evgrafov^{2,4}, W Lu^{5,6}, JA Knowles^{2,4} and DB Campbell^{2,4}

Chromodomain helicase DNA-binding protein 8 (*CHD8*) was identified as a leading autism spectrum disorder (ASD) candidate gene by whole-exome sequencing and subsequent targeted-sequencing studies. *De novo* loss-of-function mutations were identified in 12 individuals with ASD and zero controls, accounting for a highly significant association. Small interfering RNA-mediated knockdown of *CHD8* in human neural progenitor cells followed by RNA sequencing revealed that *CHD8* insufficiency results in altered expression of 1715 genes, including both protein-coding and noncoding RNAs. Among the 10 most changed transcripts, 4 (40%) were noncoding RNAs. The transcriptional changes among protein-coding genes involved a highly interconnected network of genes that are enriched in neuronal development and in previously identified ASD candidate genes. These results suggest that *CHD8* insufficiency may be a central hub in neuronal development and ASD risk.

Translational Psychiatry (2015) 5, e568; doi:10.1038/tp.2015.62; published online 19 May 2015

INTRODUCTION

Autism spectrum disorder (ASD) is a complex neurodevelopmental disorder characterized by impairments in social interaction, communication and behavioral flexibility.¹ Due to the vast clinical and genetic heterogeneity of ASD, the identification of causal genetic determinants has proven challenging.^{2–4} However, multiple independent studies have now provided substantial evidence for the contribution of *de novo* loss-of-function (LoF) mutations in chromodomain helicase DNA-binding protein 8 (*CHD8*) to ASD.^{5–9} Initially, exome sequencing on 209 ASD probands and their unaffected family members discovered two *de novo* LoF mutations in *CHD8*.⁶ Subsequent targeted sequencing of 44 candidate genes in 2446 ASD probands identified an additional seven *de novo* LoF mutations in *CHD8*, establishing a frequency of mutation that is the highest of all genes screened.⁷ Furthermore, a second targeted-sequencing study of an additional 3370 children with ASD or developmental delay identified three *de novo* LoF mutations in *CHD8*.⁸ Altogether, these highly significant findings have placed *CHD8* as a genuine ASD risk factor and account for 0.2% (12/6,176) of ASD cases. The *CHD8* LoF mutations have been found throughout the coding region of the gene, with truncating mutations as early as amino acid 62 of the 2581 amino acid *CHD8* protein. Truncating mutations were found in the chromodomain, the dex domain and the helicase domain. A detailed map of all the identified *CHD8* LoF mutations was published recently.⁹ In addition to *de novo* mutations in *CHD8*, six potentially disrupting mutations with unknown or inherited modes of transmission have also been identified^{8,9} along with a balanced chromosomal

abnormality leading to disruption of *CHD8*.¹⁰ *CHD8* LoF mutations have not been found in any of the 8792 controls included in these analyses, emphasizing the impact of *CHD8* LoF mutations on ASD risk.⁹ Phenotypic characterization of individuals with *CHD8* disrupting mutations indicate a subset of ASD that includes macrocephaly, distinct facial features and gastrointestinal difficulties.⁸

Although a critical role of *CHD8* in development is revealed by the embryonic lethality of *Chd8* knockout mice,¹¹ the function of *CHD8* in neural cell lineages has been largely unexplored. As *CHD8* actively associates with core transcriptional machinery,¹² transcription factors^{13,14} and histone-modifying complexes,¹⁵ transcriptional dysregulation conferred by *CHD8* insufficiency may provide evidence for the neurodevelopmental phenotypes observed in ASD. To emulate the potential effects of the identified LoF mutations, we performed small interfering RNA (siRNA)-mediated knockdown of *CHD8* followed by genome-wide transcriptional profiling through RNA sequencing (RNA-seq). Here, we show that knockdown of *CHD8* in SK-N-SH human neural progenitor cells results in altered expression of a highly interconnected network of genes, which are enriched in several processes essential for neuronal development. Remarkably, numerous previously identified ASD candidate genes are also differentially expressed in response to knockdown of *CHD8*, suggesting a convergence of multiple genes contributing to ASD onset. The data presented here point toward a role of *CHD8* in maintaining the active transcription of neural-specific genes and begins to elucidate the potential contributions of decreased functional *CHD8* to the pathogenesis of ASD.

¹Development, Stem Cells, and Regeneration Ph.D. Program, University of Southern California, Los Angeles, CA, USA; ²Zilkha Neurogenetic Institute, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA; ³Division of Occupational Science and Occupational Therapy, Herman Ostrow School of Dentistry, University of Southern California, Los Angeles, CA, USA; ⁴Department of Psychiatry and the Behavioral Sciences, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA; ⁵Broad Center for Regenerative Medicine and Stem Cell Research, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA and ⁶Department of Biochemistry and Molecular Biology, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA. Correspondence: Dr DB Campbell, Keck School of Medicine, University of Southern California, 213 Zilkha Neurogenetic Institute, 1501 San Pablo Street, Los Angeles, CA 90089, USA.

E-mail: dbcampbe@med.usc.edu

Received 29 October 2014; revised 12 February 2015; accepted 31 March 2015

MATERIALS AND METHODS

Cell culture

To measure gene expression in human neural progenitor cells, SK-N-SH cells (American Type Culture Collection; Manassas, VA, USA) were maintained in minimal essential medium supplemented with 10% heat-inactivated fetal bovine serum, 1% penicillin/streptomycin, non-essential amino acids, and 1.5 g l^{-1} sodium bicarbonate in 183-cm flasks at 37°C and 5% CO_2 .

siRNA transfection

To determine the impact of CHD8 knockdown on gene expression in human neural progenitor cells, SK-N-SH cells were seeded into six-well 10-cm plates and grown for 24 h (~70% confluency) before transfection. Transfections were carried out with either siRNA silencer select negative control No. 1 (catalog no. 4390843, Ambion/Life Technologies; Carlsbad, CA, USA) or siRNA targeting *CHD8* (catalog no. 33582, Ambion/Life Technologies)¹⁶ at a concentration of 20 nM using Lipofectamine RNAiMAX Reagent (Invitrogen/Life Technologies; Carlsbad, CA, USA) according to the manufacturer's protocol. Cells were then collected 72 h post siRNA transfection and processed for downstream applications. Experiments were performed in quadruplicate.

Western blot analyses

To determine the degree to which siRNA knockdown of *CHD8* transcript results in decreased CHD8 protein, total protein was isolated using the Illustra triplePrep kit (GE Healthcare; Waukesha, WI, USA) and protein concentration was determined using the DC protein assay (Bio-Rad; Hercules, CA, USA). Total protein (10 μg) was then separated on a 4–20% gradient criterion TGX gel (Bio-Rad) and transferred to a nitrocellulose membrane by capillary transfer at 80 V for 3 h using a Bio-Rad Criterion blotter system. Blots were incubated overnight at 4°C with anti-CHD8 (catalog no. 7656, Cell Signaling Technology; Danvers, MA, USA) and anti-GAPDH (catalog no. 1228, Cell Signaling Technology) primary antibodies diluted 1:1000 in blocking buffer. Anti-rabbit immunoglobulin G, horseradish-peroxidase-linked secondary antibody (catalog no. 7074, Cell Signaling Technology) was diluted 1:2000 and incubated with the membrane for 1.5 h at room temperature. Signal was developed using SuperSignal West Dura Extended Duration Substrate (Thermo Fisher Scientific; Waltham, MA, USA) and captured on a BioSpectrum Imaging System (UVP; Upland, CA, USA). The total density ratio of CHD8 to GAPDH was determined using the VisionWorks LS version 6.8 software (UVP) and analyzed by Student's paired *t*-test for comparison ($n=4$).

RNA-seq

To determine genes that are differentially expressed in response to knockdown of *CHD8* on a genome-wide scale, cells were homogenized with QIAshredder spin columns (Qiagen; Valencia, CA, USA) followed by total cellular RNA extraction using the Qiagen RNeasy kit according to the manufacturer's protocol. RNeasy extraction typically excludes fragments < 200 bp. Therefore, the data presented here focuses on long noncoding RNAs (ncRNAs). Quality (RNA integrity number) was assessed using an Agilent Technologies 2200 TapeStation Instrument (Agilent Technologies; Santa Clara, CA, USA) and 260/280 absorbance ratios were obtained using the NanoDrop ND-1000 Spectrophotometer (Thermo Fisher Scientific). cDNA library preparation was carried out using the Illumina Truseq Stranded Total RNA Sample Preparation kit and sequencing was performed on the HiSeq2000 sequencer to generate 101 bp single-end reads (Illumina; San Diego, CA, USA). An approximate average of 45 million reads per replicate was obtained for samples treated with siRNA targeting *CHD8* and an average of 40 million reads per replicate was obtained for samples treated with negative control siRNA (Supplementary Table 1).

RNA-seq analysis

Quality control pre-processing of raw reads was carried out using Cutadapt.¹⁷ Reads were first mapped to the transcriptome of either UCSC Hg19 or ENSEMBL GrCH37 release 75 and the remaining reads were then mapped to the unannotated regions of the respective human genome build using Tophat.¹⁸ The percentage of reads that were aligned to the genome for each replicate along with their respective quality control metrics are outlined in Supplementary Table 1. Following alignment, differential gene expression analysis was carried out using Cuffdiff¹⁹ along

with a supplied GTF file of the corresponding transcriptome to be analyzed. Cuffdiff analysis was performed with default settings with the exception of the masking of all annotated ribosomal RNAs and the no-effective-length-correction option, which was used to reduce over-estimation of short transcript expression. Genes with a false discovery rate (Q-value) of < 0.05 were considered significant. All gene ontology (GO) and protein–protein interaction analyses were carried out using the list of significantly differentially expressed genes identified using the UCSC Hg19 human genome annotation. To explore changes in genes transcribing ncRNAs, we also carried out the RNA-seq analysis using the ENSEMBL GrCH37, release 75, gene annotation, which contains a much larger quantity of ncRNA transcripts compared with the UCSC Hg19 annotated transcriptome.

Quantitative PCR

To measure the degree to which siRNA knockdown of *CHD8* results in decreased transcript levels and verify the quality of RNA-seq data analysis, cell pellets were processed for RNA isolation using the Qiagen RNeasy kit (Qiagen). Superscript III (Life Technologies; Grand Island, NY, USA) was used for reverse transcription of cDNA libraries. Quantitative PCR (qPCR) was performed using Life Technologies Taqman Gene Expression qPCR kits on a Life Technologies OneStepPlus real-time PCR machine. Each sample was run in triplicate along with the housekeeping gene, *GAPDH*. Assays used for validation (Life Technologies) include *ABCA1* (Hs01059118_m1), *CHD8* (Hs00394229_m1), *BDNF* (Hs02718934_s1), *CUL1* (Hs01117001_m1), *GAPDH* (Hs9999905_m1), *GJA1* (Hs00748445_s1), *HMGGA2* (Hs00171569_m1), *HSPB3* (Hs00937412_s1), *IGF2* (Hs04188276_m1), *IGFBP3* (Hs00365742_g1), *IL1R1* (Hs00991002_m1), *IL8* (Hs00174103_m1), *LINC00473* (Hs00293257_m1), *LINC01021* (Hs01388536_m1), *MMP3* (Hs00968305_m1), *PITX3* (Hs01013935_g1), *SCN2A* (Hs01109877_m1), *SERPINE1* (Hs01126606_m1), *SEMA3D* (Hs00380877_m1), *SYTL2* (Hs00909223_m1), *TGM2* (Hs00190278_m1) and *TNS3* (Hs00994933_m1). Relative quantity of the assayed transcript was determined by the $2^{-\Delta\Delta\text{Ct}}$ method using *GAPDH* as a reference.²⁰ For RNA-seq validation, the observed qPCR changes were compared with those obtained in RNA-seq data analysis through the use of Spearman and Pearson correlation tests carried out in the R statistical environment using the (cor.test) package.

Functional annotation and network construction

To begin to functionally characterize and cluster the differentially expressed transcripts, the Database for Annotation, Visualization and Integrated Discovery (DAVID)²¹ was used to analyze for the enrichment of GO terms and UP_TISSUE (Uniprot tissue annotation database) categories. All analyses were carried out with the default options. Terms that were redundant or had a corrected *P*-value > 0.05 were eliminated. Disease Association Protein–Protein Link Evaluator²² was used to determine direct interactions and connectivity among the protein products of the differentially expressed genes using the default settings. Network construction of the resulting interactions was carried out using Cytoscape version 3.1.1 software.²³

ASD candidate genes statistical analysis

To analyze whether or not a lack of functional *CHD8* results in the differential expression of other ASD candidate genes, a list of ASD candidate genes was obtained from the Simons Foundation Autism Research Initiative (SFARI) AutDB database.²⁴ A restricted list of ASD candidate genes was then curated by using only those genes with SFARI gene rankings 1–4 (strong evidence–minimal evidence) and S (syndromic) as previously reported.²⁵ Genes above the threshold of reliable detection in Cuffdiff (a total of 13 020 genes) served as the background for enrichment tests, which were measured by the two-tailed Fisher's exact test using the (fisher.test) package in R.

RESULTS

Quantitative modeling of *CHD8 de novo* loss-of-function mutations The ASD-associated *CHD8 de novo* LoF mutations likely result in an ~50% reduction in *CHD8* transcript and CHD8 protein in the developing brains of individuals with ASD. To quantitatively model these dosage-dependent consequences of *CHD8* LoF mutations, siRNA-mediated knockdown of *CHD8* was performed in the

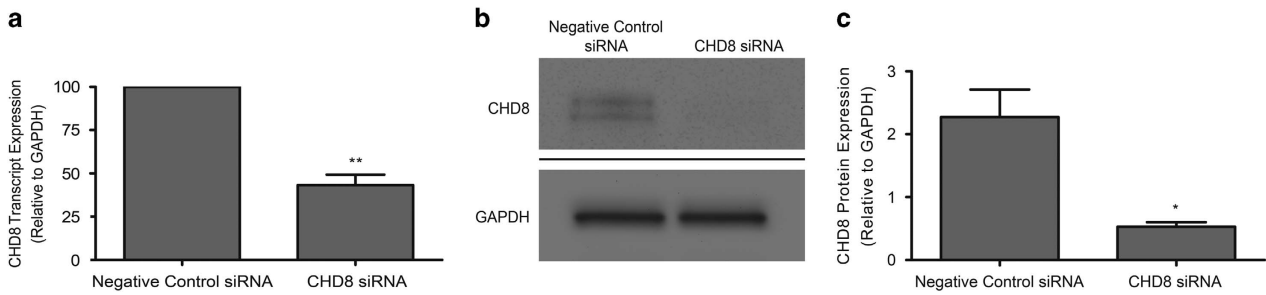


Figure 1. Quantitative modeling of disruptive mutations in CHD8. **(a)** Quantitative PCR showed that a 20 nM concentration of *CHD8* siRNA resulted in an ~50% reduction of *CHD8* transcript. **(b)** Representative western blot indicated decreased *CHD8* protein following *CHD8* siRNA-mediated knockdown. **(c)** Quantitative analysis of western blot data shows a significant reduction in *CHD8* protein following *CHD8* siRNA treatment. Error bars are s.e.m., ** $P < 0.01$, * $P < 0.05$. *CHD8*, chromodomain helicase DNA-binding protein 8; siRNA, small interfering RNA.

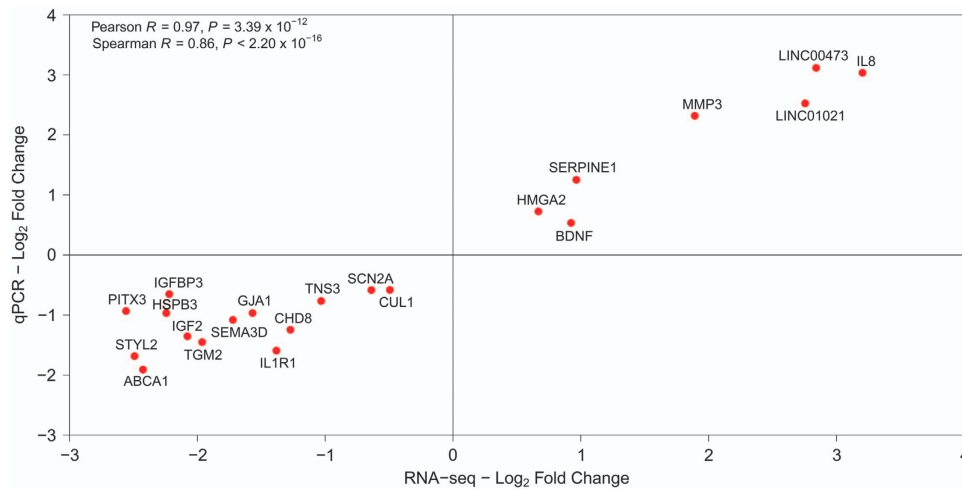


Figure 2. qPCR validation of differentially expressed genes identified by RNA-Seq. Fifteen genes excluding *CHD8* were assayed via qPCR to establish the validity of the differential expression analysis based on the obtained RNA-seq data. qPCR and RNA-seq values showed a statistically significant correlation with both Pearson and Spearman methods. qPCR, quantitative PCR.

human neural progenitor cell line, SK-N-SH. Following this, the knockdown efficiency of both *CHD8* transcript and *CHD8* protein levels were confirmed via qPCR and western blotting, respectively (Figure 1).

CHD8 insufficiency results in genome-wide transcriptional changes

As transcriptional changes due to *CHD8* insufficiency may be able to inform the mechanisms by which a lack of *CHD8* influences ASD pathogenesis, we performed genome-wide transcriptome profiling of *CHD8*-depleted human neuronal progenitor cells, SK-N-SH, via RNA-seq. Differential gene expression analysis using the UCSC Hg19 genome annotation revealed 1715 differentially expressed genes (Supplementary Tables 2 and 3), with a majority of protein-coding genes downregulated (~63%). Subsequent qPCR validation was performed for genes categorized as either having large fold changes in expression or possessing functions related to neurodevelopmental processes identified by GO analysis. Comparison of gene expression levels measured by RNA-Seq and qPCR indicated a highly significant Pearson correlation coefficient (Figure 2), validating the magnitude of gene expression changes identified by RNA-seq. Both protein-coding and ncRNAs with altered expression are distributed across the genome in a pattern consistent with length of the chromosome; there does not appear to be any clustering of the genes with altered expression (Supplementary Figures 1 and 2).

CHD8 influences the expression of ncRNA

ncRNAs represent a relatively unexplored source of functional transcripts that may contribute to neurodevelopmental phenotypes.²⁶ Using the ENSEMBL GrCh37, release 75, gene annotation, we identified 657 differentially expressed genes and found that ~12.5% of these were ncRNAs (Table 1, Supplementary Tables 4 and 5). The identified ncRNA account for 40% of the top 10 differentially expressed genes (Table 1). In contrast to protein-coding genes, the majority of ncRNAs are upregulated (~60%) instead of downregulated following *CHD8* knockdown.

Loss of functional CHD8 affects numerous ASD susceptibility genes

Comparing the list of differentially expressed genes to those genes contained in the SFARI AutDB gene database,²⁴ there is an overlap of 65 genes excluding *CHD8* (Supplementary Table 6). Of the 631 genes in the AutDB database, 409 were expressed in SK-N-SH cells. Among the 1715 genes differentially expressed following *CHD8* knockdown in SK-N-SH cells, 65 were overlapping (Supplementary Table 7). Although the entire list of differentially expressed genes does not reach statistical significance for enrichment ($P = 0.059$, OR (odds ratio) = 1.307, Fisher's exact test), considering only those genes that are downregulated significantly strengthens this enrichment ($P = 4.793 \times 10^{-3}$, OR = 1.59). To focus on genes with substantial evidence for a role in ASD, a restricted list of 250 genes (159 of which are expressed in SK-N-SH cells) was curated from the SFARI database taking into account

Table 1. Top 10 differentially expressed genes following knockdown of *CHD8*

Gene ^a	Fold change	Locus ^a	Protein- or noncoding	Q-value
IL8	7.60	4:74606222–74609433	Protein-coding	2.76E–3
LINC00473	6.53	6:165740775–166401536	Noncoding	2.03E–2
LINC00877	5.59	3:72084450–72291716	Noncoding	2.93E–2
ACPP	5.03	3:132036210–132087142	Protein-coding	1.53E–2
LINC10121	4.66	5:27472398–27496508	Noncoding	2.76E–3
HSPB3	–4.64	5:53751444–53752207	Protein-coding	1.68E–2
ABCA1	–4.89	9:107543282–107691173	Protein-coding	2.76E–3
SYTL2	–5.51	11:85405266–85522184	Protein-coding	2.76E–3
PTX3	–5.89	3:156893011–157251408	Protein-coding	2.76E–3
RP11-777B9.5	–6.81	4:78929665–78929842	Noncoding	2.76E–3

Abbreviation: CHD8, chromodomain helicase DNA-binding protein 8. ^aRefers to ENSEMBL human genome GrCH37 build 37.

Table 2. ASD-associated genes differentially expressed following knockdown of *CHD8*

ASD-associated gene ^a	Fold change	Locus ^b	Q-value
HTR1B	4.22	6:78171947–78173120	3.54E–2
CNTNAP4	1.88	16:76311175–76593135	1.27E–2
NBEA	1.65	13:35516423–36705514	1.05E–3
CNTN4	1.63	3:2140549–3099645	1.05E–3
ITGB3	1.55	17:45331207–45390077	9.79E–3
SYNE1	1.47	6:152442818–152958534	1.05E–3
STK39	1.28	2:168810529–169104105	2.55E–2
GTF2I	–1.27	7:74072029–74175022	2.20E–2
UBE2H	–1.31	7:129470572–129592800	7.70E–3
RIMS3	–1.34	1:41086351–41131324	4.08E–3
ICA1	–1.34	7:8152814–8302242	7.65E–3
RPS6KA2	–1.37	6:166822853–167275771	1.92E–3
NRXN2	–1.38	11:64373645–64490660	1.71E–2
ADK	–1.39	10:75910942–76469061	1.71E–2
CEP41	–1.40	7:130033611–130081051	7.65E–3
HDAC4	–1.40	2:239969863–240322643	1.12E–2
NF1	–1.41	17:29421944–29704695	3.40E–3
TSPAN7	–1.45	X:38420730–38548172	2.47E–2
SCN2A	–1.56	2:166095911–166248820	1.05E–3
REEP3	–1.59	10:65281122–65384883	1.05E–3
CHRNA7	–1.63	15:32322685–32462384	1.92E–3
OXTR	–1.66	3:8792094–8811300	1.40E–2
NRP2	–1.68	2:206547223–206662857	1.05E–3
DHCR7	–1.70	11:71145456–71159477	1.05E–3
SEMA5A	–1.81	5:9035137–9546233	1.05E–3
KCNMA1	–1.83	10:78629358–79397577	1.05E–3
NPAS2	–2.45	2:101436612–101613287	1.05E–3

Abbreviations: ASD, autism spectrum disorder; CHD8, chromodomain helicase DNA-binding protein 8. ^aAccording to genes within SFARI database possessing gene scores of '1–4' and 'S'. ^bRefers to UCSC Hg19.

SFARI gene rankings as previously described.²⁵ Following this narrowing, the list of differentially expressed genes due to knockdown of *CHD8* contained a total of 27 overlapping genes excluding *CHD8* (Table 2), which did not reach statistical significance for enrichment ($P=0.153$, $OR=1.378$). Again, considering only those genes that are downregulated significantly strengthened the enrichment ($P=0.038$, $OR=1.673$). In comparing all genes from the SFARI database and the narrow list of ASD genes, upregulated genes fail to show significant enrichment in both the cases. These results suggest no advantage of narrowing the list of genes by evidence-based scoring. These results further suggest that *CHD8* insufficiency may regulate a network of genes that contribute to ASD risk. Although previously identified ASD candidate genes are enriched among the 1715 differentially expressed genes following *CHD8* knockdown, the magnitude of gene expression changes is similar between ASD candidate genes and genes not previously implicated in ASD. These results suggest

that, in addition to altering expression of genes implicated in ASD, *CHD8* knockdown also impacts expression of genes that are either unrelated to ASD or not yet identified as contributors to ASD risk.

Transcriptional consequences of *CHD8* insufficiency are enriched in processes essential for neuronal development

GO analysis using the DAVID web server for all differentially expressed genes shows an enrichment of genes within GO categories describing several functions that have previously been associated with *CHD8* including regulation of proliferation, apoptosis and transcription (Figure 3a, Supplementary Table 8).^{11,12,27} We also performed GO analysis grouping the differentially expressed genes by their direction of differential expression. Although upregulated genes show GO enrichment in mitochondrial membrane-related terms (Supplementary Table 9), GO analysis of downregulated genes shows an enrichment in several neuron-specific processes in addition to those categories

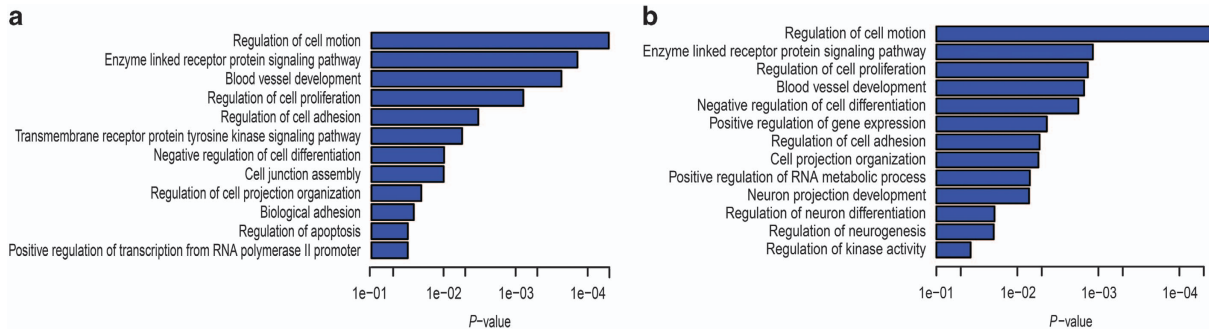


Figure 3. Representative enrichment of gene ontology among differentially expressed genes. **(a)** Gene ontology (GO) enrichment analysis on all differentially expressed genes due to knockdown of *CHD8*. **(b)** GO enrichment analysis on genes downregulated in response to knockdown of *CHD8*. Upregulated genes subjected to GO analysis indicated enrichment of mitochondrial membrane genes only (not shown). *P*-value represents Benjamini–Hotchberg correction for multiple tests. *CHD8*, chromodomain helicase DNA-binding protein 8.

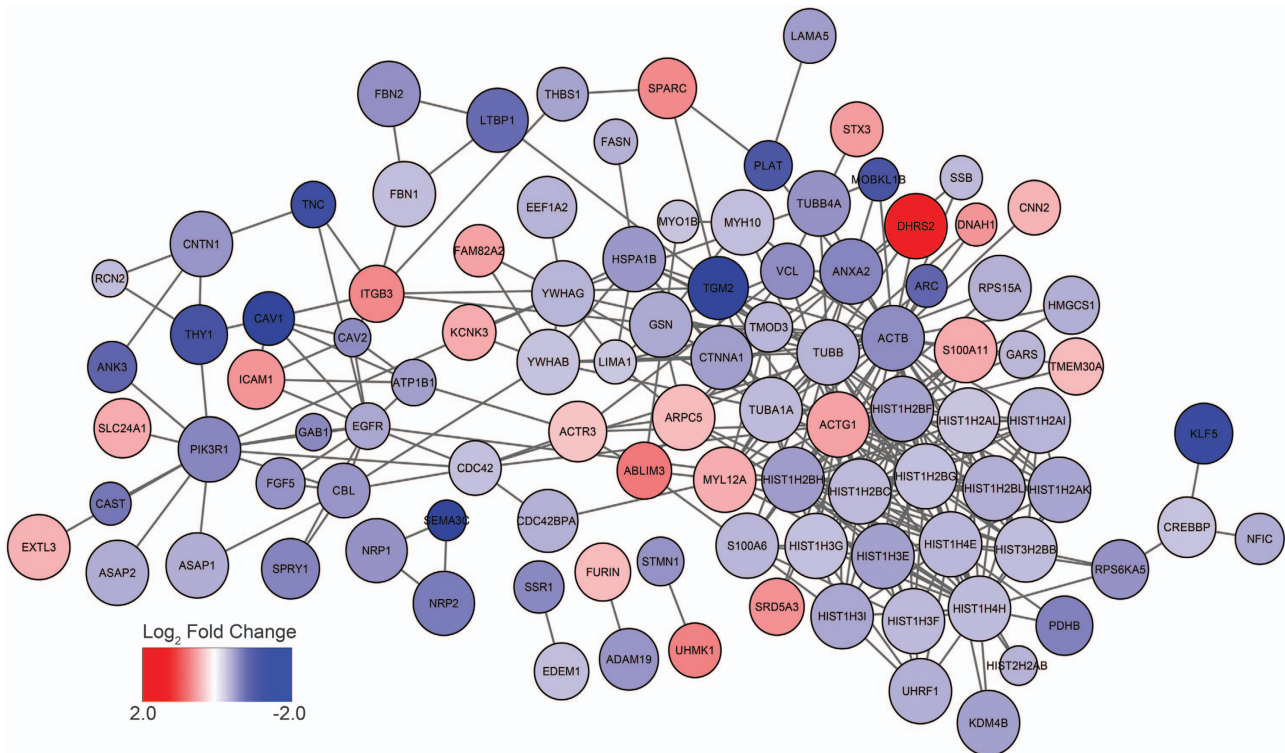


Figure 4. Protein–protein interaction analysis of differentially expressed genes—Disease Association Protein–Protein Link Evaluator (DAPPLE) was used to determine direct relationships among the protein products of the genes that were differentially expressed due to knockdown of *CHD8*. A subnetwork of genes with associated corrected *P*-values < 0.05, indicating seed proteins that are more likely to be connected to other seed proteins within the network than by chance, was generated and visualized using Cytoscape software. Increasing significance corresponds to increasing node size. Differential expression of the associated gene product is shown with increasing densities of blue and red indicating the degree of down- and upregulation conferred by knockdown of *CHD8*, respectively. *CHD8*, chromodomain helicase DNA-binding protein 8.

observed for all differentially expressed genes. These include GO categories such as regulation of neuron differentiation, neurogenesis and neuron projection development (Figure 3b, Supplementary Table 10), all of which may be processes disrupted during ASD pathogenesis. In addition, utilizing the Uniprot tissue (UP_TISSUE) annotation database, we found the downregulated genes to be highly enriched in brain tissue ($P = 2.01 \times 10^{-15}$, Benjamini–Hotchberg correction), whereas upregulated genes were not enriched in any specific tissue (Supplementary Table 11). Overall, these results suggest that *CHD8* retains conserved functions and also has an essential role in the maintenance of tissue-specific genes involved in neuron development.

CHD8 regulates a highly interconnected network of genes To assess the connectivity among the differentially expressed genes following knockdown of *CHD8*, we utilized the Disease Association Protein–Protein Link Evaluator.²² Analysis of all differentially expressed genes shows significant enrichment in both direct and indirect connectivity ($P = 3.00 \times 10^{-3}$ for both), identifying 2022 direct interactions (Figure 4, Supplementary Table 12). Analysis of the downregulated genes indicated a significant increase in indirect connectivity ($P = 3.00 \times 10^{-3}$ for direct connectivity and $P = 9.99 \times 10^{-4}$ for indirect connectivity). Although upregulated genes showed no enrichment in ASD candidate genes and weak enrichment in GO categories, Disease

Association Protein–Protein Link Evaluator showed an unexpected enrichment in direct and indirect connectivity among upregulated genes ($P=2.00 \times 10^{-3}$ and $P=3.20 \times 10^{-2}$, respectively). These results suggest a network of connected genes upregulated by *CHD8*.

DISCUSSION

From the initial discovery through exome sequencing to confirmation by targeted sequencing of additional ASD probands, *de novo* LoF mutations in *CHD8* are highly implicated in contributing to ASD.^{5–9} Due to the heterogeneity of ASD, it is likely to be influenced by the combinatorial effects of hundreds of genes under abnormal regulation. Although genetic influences may differ among ASD cases, strong evidence suggests that the functional contribution of these various factors may converge on similar pathways.^{4,28} Our data suggest that *CHD8* may be a central node in one of these pathways as *CHD8* insufficiency in human neural progenitor cells results in altered expression of previously identified ASD candidate genes. Among recently identified genes disrupted in autism,²⁹ *CACNA2D3* and *ETFB* were among the 1715 genes with altered expression following *CHD8* knockdown, suggesting potential biological pathways that contribute to ASD.

Considering the enrichment of ASD candidate genes among the differentially expressed protein-coding genes, the ncRNA genes identified in our analysis may also serve as contributing factors to ASD phenotypes. The ncRNA represents a relatively unexplored category of genes that participate in several regulatory functions and may potentially contribute to the neurodevelopmental phenotypes observed in ASD.^{26,30,31} For example, the ncRNAs colorectal neoplasia differentially expressed (CRNDE) and telomerase RNA component, have both been shown to undergo differential expression during neuronal differentiation of pluripotent stem cells suggesting their participation in neurogenesis.^{32,33} In particular, CRNDE is intimately linked with the insulin/IGF signaling pathway, as treatment with IGF1 or IGF2 downregulates nuclear-retained CRNDE transcripts.³⁴ This is of importance to ASD as administration of IGF1 has been shown to rescue phenotypes associated with Rett Syndrome^{35,36} and is now in clinical trials.³⁷ Our findings of upregulation of CRNDE following decreased expression of *CHD8* suggest that this ncRNA may represent a novel therapeutic target. In addition, our data shows downregulation of the ncRNA TUG1, which has been shown to directly associate with the chromatin modifying complex, PRC2, and to be regulated by p53.³⁸ This may represent a novel ncRNA target of *CHD8*, as *CHD8* is known to interact with p53 and colocalize to promoters of p53-regulated genes.¹¹

Consistent with previous studies showing the participation of *CHD8* in large protein complexes involving members of the mixed lineage leukemia (MLL) family,³⁹ depletion of endogenous *CHD8* in neural cells results in decreased expression of the MLL genes, *MLL1* and *MLL3*, along with *MN1*, which associates with MLL to regulate transcription.⁴⁰ In addition, our data show the conservation of specific changes in gene expression including genes involved in Wnt signaling (*DKK1* and *NKD2*) and the p53 target, *CDKN1A*, observed in *CHD8*-depleted HCT116 colorectal cancer cells and U2OS osteosarcoma cells, respectively.^{11,41} Approximately 13% of differentially expressed genes in our analysis of human neural progenitor cells were shown to have *CHD8* bound to their promoters in previous work characterizing the binding sites of *CHD8* in human C33A cervical carcinoma cells (Supplementary Table 13).¹⁴ Although these data point toward the conservation of specific regulatory mechanisms involving *CHD8*, our analyses also suggest a tissue-specific function of *CHD8* in maintaining the active transcription of genes related to neuronal development. This is illustrated by the significant enrichment of neuron-development GO categories, brain tissue specificity and the significant connectivity of genes downregulated by *CHD8* insufficiency. In contrast, upregulated genes

lack tissue enrichment, suggesting that they are not specific to any one particular tissue type.

Sugathan *et al.*⁴² recently described the binding sites of *CHD8* and changes in transcription resulting from knockdown of *CHD8* in iPSC-derived neural progenitor cells where they found *CHD8* to be localized to 5568 promoters and 1756 differentially expressed genes, respectively. Comparing the Sugathan *et al.*⁴² list of differentially expressed genes with the list of 1715 genes observed in the current study, there is an overlap of 159 genes (9.37%, Supplementary Table 14). There were 633 (37.30%) differentially expressed genes in our data set that Sugathan *et al.*⁴² reported to have *CHD8* bound in close proximity to the transcription start site (Supplementary Table 15), indicating that there is a substantial proportion of genes directly regulated by *CHD8*. In addition, the differentially expressed genes that are shared between our data and that of Sugathan *et al.*⁴² include *H1FO*, *TP53INP1* and *MLL3*, suggesting conserved functions of *CHD8* to participate in protein complexes containing members of the MLL family³⁹ and regulate p53 signaling through interactions with Histone H1.¹¹ As we also find *CHD8* knockdown results in an enrichment of differentially expressed ASD-associated genes, this illustrates the involvement of *CHD8* in regulating neurodevelopmental genes, which may differ between cell types and emphasizes the need for these transcriptional networks to be analyzed in several cellular models.

Our data indicate that a decrease of functional *CHD8* in human neural progenitor cells induces numerous transcriptional changes that potentially contribute to the pathogenesis of ASD. As the goal of this study was to uncover the consequences of *CHD8* insufficiency as a whole rather than each of the particular *CHD8* LoF mutations identified in individual patients, the modeling of the exact *CHD8* LoF mutations through genome-editing techniques or induced pluripotent stem cell generation should be considered a priority. With the function of *CHD8* in neurons just beginning to be explored, this work lays a foundation for further functional studies and identifies numerous molecular targets, including targetable ncRNAs, that could potentially be modulated for the purposes of therapeutic intervention.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

This work was supported by the US National Institutes of Health grants R01MH100172 to DBC and R21MH099504 to DBC.

REFERENCES

- American Psychiatric Association and American Psychiatric Association DSM-5 Task Force. *Diagnostic and Statistical Manual of Mental Disorders: DSM-5*. American Psychiatric Association: Arlington, VA, USA, 2013.
- Krumm N, O'Roak BJ, Shendure J, Eichler EE. A *de novo* convergence of autism genetics and molecular neuroscience. *Trends Neurosci* 2014; **37**: 95–105.
- Jeste SS, Geschwind DH. Disentangling the heterogeneity of autism spectrum disorder through genetic findings. *Nat Rev Neurol* 2014; **10**: 74–81.
- Levitt P, Campbell DB. The genetic and neurobiologic compass points toward common signaling dysfunctions in autism spectrum disorders. *J Clin Invest* 2009; **119**: 747–754.
- Sanders SJ, Murtha MT, Gupta AR, Murdoch JD, Raubeson MJ, Willsey AJ *et al.* *De novo* mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* 2012; **485**: 237–241.
- O'Roak BJ, Vives L, Girirajan S, Karakoc E, Krumm N, Coe BP *et al.* Sporadic autism exomes reveal a highly interconnected protein network of *de novo* mutations. *Nature* 2012; **485**: 246–250.
- Neale BM, Kou Y, Liu L, Ma'ayan A, Samocha KE, Sabo A *et al.* Patterns and rates of exonic *de novo* mutations in autism spectrum disorders. *Nature* 2012; **485**: 242–245.

- 8 O'Roak BJ, Vives L, Fu W, Egertson JD, Stanaway IB, Phelps IG *et al*. Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. *Science* 2012; **338**: 1619–1622.
- 9 Bernier R, Golzio C, Xiong B, Stessman HA, Coe BP, Penn O *et al*. Disruptive CHD8 mutations define a subtype of autism early in development. *Cell* 2014; **158**: 263–276.
- 10 Talkowski ME, Rosenfeld JA, Blumenthal I, Pillalamarri V, Chiang C, Heilbut A *et al*. Sequencing chromosomal abnormalities reveals neurodevelopmental loci that confer risk across diagnostic boundaries. *Cell* 2012; **149**: 525–537.
- 11 Nishiyama M, Oshikawa K, Tsukada Y, Nakagawa T, Iemura S, Natsume T *et al*. CHD8 suppresses p53-mediated apoptosis through histone H1 recruitment during early embryogenesis. *Nat Cell Biol* 2009; **11**: 172–182.
- 12 Rodríguez-Paredes M, Ceballos-Chávez M, Esteller M, García-Domínguez M, Reyes JC. The chromatin remodeling factor CHD8 interacts with elongating RNA polymerase II and controls expression of the cyclin E2 gene. *Nucleic Acids Res* 2009; **37**: 2449–2460.
- 13 Ishihara K, Oshimura M, Nakao M. CTCF-dependent chromatin insulator is linked to epigenetic remodeling. *Mol Cell* 2006; **23**: 733–742.
- 14 Subtil-Rodríguez A, Vázquez-Chávez E, Ceballos-Chávez M, Rodríguez-Paredes M, Martín-Subero JI, Esteller M *et al*. The chromatin remodeller CHD8 is required for E2F-dependent transcription activation of S-phase genes. *Nucleic Acids Res* 2014; **42**: 2185–2196.
- 15 Yates JA, Menon T, Thompson BA, Bochar DA. Regulation of HOXA2 gene expression by the ATP-dependent chromatin remodeling enzyme CHD8. *FEBS Lett* 2010; **584**: 689–693.
- 16 Caldon CE, Sergio CM, Schütte J, Boersma MN, Sutherland RL, Carroll JS *et al*. Estrogen regulation of cyclin E2 requires cyclin D1 but not c-Myc. *Mol Cell Biol* 2009; **29**: 4623–4639.
- 17 Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J* 2011; **17**, pp 10–12.
- 18 Kim D, Pertea G, Trapnell C, Pimental H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* 2013; **14**: R36.
- 19 Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR *et al*. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and cufflinks. *Nat Protoc* 2012; **7**: 562–578.
- 20 Schmittgen TD, Livak KJ. Analyzing real-time PCR data by the comparative C (T) method. *Nat Protoc* 2008; **3**: 1101–1108.
- 21 Dennis G, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC *et al*. DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol* 2003; **4**: P3.
- 22 Rossin EJ, Lage K, Raychaudhuri S, Xavier RJ, Tatar D, Benita Y *et al*. Proteins encoded in genomic regions associated with immune-mediated disease physically interact and suggest underlying biology. *PLoS Genet* 2011; **7**: e1001273.
- 23 Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D *et al*. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003; **13**: 2498–2504.
- 24 Abrahams BS, Arking DE, Campbell DB, Mefford HC, Morrow EM, Weiss LA *et al*. SFARI Gene 2.0: a community-driven knowledgebase for the autism spectrum disorders (ASDs). *Mol Autism* 2013; **4**: 36.
- 25 Parikshak NN, Luo R, Zhang A, Won H, Lowe JK, Chandran V *et al*. Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. *Cell* 2013; **155**: 1008–1021.
- 26 Wilkinson B, Campbell DB. Contribution of long noncoding RNAs to autism spectrum disorder risk. *Int Rev Neurobiol* 2013; **113**: 35–59.
- 27 Menon T, Yates JA, Bochar DA. Regulation of androgen-responsive transcription by the chromatin remodeling factor CHD8. *Mol Endocrinol* 2010; **24**: 1165–1174.
- 28 Berg JM, Geschwind DH. Autism genetics: searching for specificity and convergence. *Genome Biol* 2012; **13**: 247.
- 29 De Rubeis S, He X, Goldberg AP, Poultney CS, Samocha K, Cicek AE *et al*. Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* 2014; **515**: 209–215.
- 30 Kerin T, Ramanathan A, Rivas K, Grepo N, Coetzee GA, Campbell DB. A noncoding RNA antisense to moesin at 5p14.1 in autism. *Sci Transl Med* 2012; **4**: 128ra40.
- 31 Ziats MN, Rennett OM. Aberrant expression of long noncoding RNAs in autistic brain. *J Mol Neurosci* 2013; **49**: 589–593.
- 32 Lin M, Pedrosa E, Shah A, Hrabovsky A, Maqbool S, Zheng D *et al*. RNA-Seq of human neurons derived from iPSCs reveals candidate long non-coding RNAs involved in neurogenesis and neuropsychiatric disorders. *PLoS One* 2011; **6**: e23356.
- 33 Ng SY, Johnson R, Stanton LW. Human long non-coding RNAs promote pluripotency and neuronal differentiation by association with chromatin modifiers and transcription factors. *EMBO J* 2012; **31**: 522–533.
- 34 Ellis BC, Graham LD, Molloy PL. CRNDE, a long non-coding RNA responsive to insulin/IGF signaling, regulates genes involved in central metabolism. *Biochim Biophys Acta* 2014; **1843**: 372–386.
- 35 Marchetto MC, Carromeu C, Acab A, Yu D, Yeo GW, Mu Y *et al*. A model for neural development and treatment of Rett syndrome using human induced pluripotent stem cells. *Cell* 2010; **143**: 527–539.
- 36 Tropea D, Giacometti E, Wilson NR, Beard C, McCurry C, Fu DD *et al*. Partial reversal of Rett Syndrome-like symptoms in Mecp2 mutant mice. *Proc Natl Acad Sci USA* 2009; **106**: 2029–2034.
- 37 Khwaja OS, Ho E, Barnes KV, O'Leary HM, Pereira LM, Finkelstein Y *et al*. Safety, pharmacokinetics, and preliminary assessment of efficacy of mecasermin (recombinant human IGF-1) for the treatment of Rett syndrome. *Proc Natl Acad Sci USA* 2014; **111**: 4596–4601.
- 38 Khalil AM, Guttman M, Huarte M, Garber M, Raj A, Rivea Morales D *et al*. Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc Natl Acad Sci USA* 2009; **106**: 11667–11672.
- 39 Dou Y, Milne TA, Tackett AJ, Smith ER, Fukuda A, Wysocka J *et al*. Physical association and coordinate function of the H3 K4 methyltransferase MLL1 and the H4 K16 acetyltransferase MOF. *Cell* 2005; **121**: 873–885.
- 40 Milne TA, Hughes CM, Lloyd R, Yang Z, Rozenblatt-Rosen O, Dou Y *et al*. Menin and MLL cooperatively regulate expression of cyclin-dependent kinase inhibitors. *Proc Natl Acad Sci USA* 2005; **102**: 749–754.
- 41 Thompson BA, Tremblay V, Lin G, Bochar DA. CHD8 is an ATP-dependent chromatin remodeling factor that regulates beta-catenin target genes. *Mol Cell Biol* 2008; **28**: 3894–3904.
- 42 Sugathan A, Biagioli M, Golzio C, Erdin S, Blumenthal I, Manavalan P *et al*. CHD8 regulates neurodevelopmental pathways associated with autism spectrum disorder in neural progenitors. *Proc Natl Acad Sci USA* 2014; **111**: E4468–E4477.



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

Supplementary Information accompanies the paper on the Translational Psychiatry website (<http://www.nature.com/tp>)