Genome-wide, integrative analysis implicates microRNA dysregulation in autism spectrum disorder

Ye E Wu1,2, Neelroop N Parikshak1,3,4, T Grant Belgard1,2,6 & Daniel H Geschwind1,2,4,5

Genetic variants conferring risk for autism spectrum disorder (ASD) have been identified, but the role of post-transcriptional mechanisms in ASD is not well understood. We performed genome-wide microRNA (miRNA) expression profiling in post-mortem brains from individuals with ASD and controls and identified miRNAs and co-regulated modules that were perturbed in ASD. Putative targets of these ASD-affected miRNAs were enriched for genes that have been implicated in ASD risk. We confirmed regulatory relationships between several miRNAs and their putative target mRNAs in primary human neural progenitors. These include hsa-miR-21-3p, a miRNA of unknown CNS function that is upregulated in ASD and that targets neuronal genes downregulated in ASD, and hsa_can_1002-m, a previously unknown, primate-specific miRNA that is downregulated in ASD and that regulates the epidermal growth factor receptor and fibroblast growth factor receptor signaling pathways involved in neural development and immune function. Our findings support a role for miRNA dysregulation in ASD pathophysiology and provide a rich data set and framework for future analyses of miRNAs in neuropsychiatric diseases.

ASD is a group of clinically heterogeneous neurodevelopmental disorders characterized by deficits in social functioning and the presence of repetitive and restricted behaviors or interests. ASD also manifests substantial genetic heterogeneity; hundreds of genomic loci have been implicated. Dozens of rare Mendelian disorders, including fragile X syndrome, neurofibromatosis, Rett syndrome, tuberous sclerosis complex and structural chromosomal variants, confer a high risk for ASD. Recent studies have also revealed the contribution of rare, de novo single-nucleotide mutations, none of which account for more than a small fraction of ASD cases. Thousands of common variants are also estimated to contribute to ASD, although the effect size of individual loci is small. Despite this remarkable heterogeneity, ASD-associated mutations have been suggested to target a few convergent biological processes, including synaptic function and neuronal activity, postsynaptic density protein metabolism, neuronal cell adhesion, WNT signaling, and chromatin remodeling during neurogenesis.

In contrast, much less is known about the contribution of post-transcriptional regulatory mechanisms to ASD. miRNAs, which are small non-coding regulatory RNAs that mediate mRNA destabilization and/or translational repression, represent a sparsely studied class of putative contributors to ASD pathophysiology. Each miRNA can target up to hundreds of genes; collectively, miRNAs are predicted to target >60% of the transcriptome, establishing them as potential regulators of complex gene networks. miRNAs have been shown to regulate processes that are pivotal to brain development and function, including neurogenesis, neuronal maturation and synaptic plasticity.

To assess the potential role of miRNAs in ASD, we performed genome-wide miRNA expression profiling in post-mortem brains from individuals with ASD and controls. We found a shared pattern of miRNA dysregulation in a majority of ASD-affected brains. The targets of ASD-associated miRNAs were enriched in ASD risk genes. Using bioinformatic and gene network analyses, we were able to link these perturbations with transcriptomic changes in ASD-affected brain.

RESULTS
Differential expression of miRNAs in ASD brain

We profiled miRNAs in 242 post-mortem brain tissue samples from 55 ASD cases and 42 controls (CTL) (Fig. 1a and Supplementary Table 1) using Illumina small RNA sequencing (sRNA-seq; Online Methods). Up to three brain regions from each individual were assessed: frontal cortex (FC, Brodmann area (BA) 9), temporal cortex (TC, BA41/42/22) and cerebellar vermis (Fig. 1a), all of which have been implicated in ASD. After quality control (Supplementary Fig. 1a,b; Online Methods), mature miRNAs documented in miRBase release 20 (http://www.mirbase.org) were identified and quantified using the miRDeep2 algorithm (Online Methods). We also included in our analysis previously unknown miRNAs that were identified in a recent study based on 94 human sRNA-seq data sets and supported by experimental evidence, as well as previously unknown miRNAs predicted from our sequencing data with high confidence using two different methods (Online Methods and Supplementary Table 2).

699 miRNAs (552 in miRBase 20 and 147 previously unknown) were detected (Online Methods and Supplementary Table 2).

1Program in Neurobehavioral Genetics, Semel Institute, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, California, USA.
2Center for Autism Research and Treatment, Semel Institute, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, California, USA.
3Interdepartmental Program in Neuroscience, University of California, Los Angeles, Los Angeles, California, USA.
4Program in Neurogenetics, Department of Neurology, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, California, USA.
5Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, California, USA.
6Present address: Verge Genomics, San Francisco, California, USA.

Received 16 January; accepted 29 July; published online 29 August 2016; doi:10.1038/nn.4373
The miRNA expression profiles were very similar between the frontal and temporal cortex, but were distinct in the cerebellum (Supplementary Fig. 2a-f), consistent with previous observations for mRNAs. We therefore combined 95 covariate-matched samples (47 samples from 28 ASD cases and 48 samples from 28 controls; Supplementary Fig. 1c and Supplementary Table 1) from the FC and TC for differential gene expression (DGE) analysis, comparing ASD and CTL using a linear mixed-effects regression framework to control for potential confounders (Online Methods). We identified 58 miRNAs showing significant (false discovery rate (FDR) < 0.05) expression changes between ASD and CTL: 17 were downregulated and 41 were upregulated in ASD cortex (Fig. 1b and Supplementary Table 2). The fold changes for the differentially expressed miRNAs were highly concordant between the FC and TC (Pearson correlation coefficient \( R = 0.96, P < 2.2 \times 10^{-16} \); Fig. 1c).

To ensure the robustness of the signal, we performed resampling (Online Methods), finding that the fold changes for all miRNAs were highly concordant between the resampled and the original sample sets (Pearson's \( R = 0.93–0.97, P < 2.2 \times 10^{-16} \); Supplementary Fig. 3a). To confirm that the result was not biased by a small number of samples with low RNA Integrity (RIN) or high post-mortem interval (PMI), or from ASD cases with chromosome 15q11-13 duplication syndrome, we also performed DGE analysis after removing these samples and found that the expression changes were concordant with those observed in all samples combined (Pearson's \( R = 0.99, 0.98 \) and 0.99, \( P < 2.2 \times 10^{-16} \), for samples with RIN ≥ 5, PMI ≤ 30 h and after removal of 15q11-13 duplication samples, respectively; Supplementary Fig. 3b–d).

Hierarchical clustering based on the top differentially expressed miRNAs (FDR < 0.05, \( |\log_2(\text{fold change})| \geq 0.3 \)) revealed distinct clustering for the majority of ASD cortex samples; 37 of 47 ASD samples from 23 of 28 cases grouped together, and confounders such as age, sex, brain region, seizures, medication, RIN and brain bank did not drive the clustering (Fig. 1d), suggesting a shared miRNA dysregulation signature among the majority of ASD samples.

We further validated ten differentially expressed miRNAs using quantitative reverse transcription PCR (qRT-PCR) and confirmed sequencing-detected changes (Supplementary Fig. 3f,g). Together, these results support the robustness and reproducibility of our data.

In addition, we performed DGE analysis for 47 covariate-matched cerebellum samples (21 samples from ASD cases and 26 samples from controls; Supplementary Fig. 1e, Supplementary Table 1 and Online Methods), and observed a similar trend in differential expression compared with the cortex (Pearson's \( R = 0.86, P < 2.2 \times 10^{-16} \) for miRNAs differentially expressed in the cortex; Supplementary Fig. 3e; 16 miRNAs differentially expressed at FDR < 0.05 in both cortex and cerebellum).

**Perturbation of miRNA coexpression modules in ASD brain**

To further gain a systems-level understanding of the relationship between miRNA expression changes and disease status, we next applied weighted gene coexpression network analysis (WGCNA) using 109 cortex samples and a method that is robust to outliers (Supplementary Figs. 1d and 4a; Online Methods) to assign individual miRNAs to coexpression modules. We identified 11 modules, summarized by their first principal component (PC1), or module eigengene (ME)\(^4\) (Fig. 2a and Supplementary Table 2). We then assessed ME relationship to ASD status and identified four modules that were significantly correlated with disease status (Pearson's correlation, FDR < 0.05) and not any of the technical confounders: two downregulated (brown and salmon) and two upregulated (yellow and magenta) in ASD samples (Fig. 2b–j).

WGCNA permits direct assessment of the relationship of modules to important experimental covariates, such as age, sex, brain region and technical confounders, including RIN, PMI and brain bank\(^13–15\). The yellow and magenta modules showed significant correlations with age (Fig. 2b). Further inspection of the saloon module revealed that it could be driven by younger non-aged-matched samples in cases compared with controls (Fig. 2c,d, Supplementary Fig. 1d and Online Methods), so it was excluded from subsequent analysis.

To determine whether the coexpression structure is similar between ASD and CTL samples, we constructed networks using ASD or CTL samples only and performed module preservation tests\(^16\), observing that all three ASD-associated modules were preserved across the ASD and CTL networks (Supplementary Fig. 4c,e). Similar analysis revealed that the ASD-associated modules were also preserved across the FC and TC (Supplementary Fig. 4g). Furthermore, these coexpression relationships were also observed in two independent data sets from human brain (Online Methods and Supplementary Fig. 4h,i). These results support the robustness and reproducibility of the three ASD-associated modules.

Coexpression among miRNAs may arise from co-regulation by common transcription factors (TFs) and/or chromatin regulators (CRs). To test this possibility, we obtained genome-wide high-confidence binding sites for 61 human TFs and CRs expressed in the cerebral cortex\(^17\), and identified miRNAs that are potentially regulated by each TF or CR (Online Methods). We then assessed enrichment of these potential targets in the three ASD-associated miRNA modules, observing distinct patterns of TF and CR enrichment in each module, consistent with their differential co-regulation (Supplementary Table 3). Notably, the magenta module was enriched (Fisher’s exact test, FDR < 0.05) for potential targets of SMARCC1, a core component of the BAF complex that has been implicated in ASD via mutational and network analysis of brain gene expression data\(^15,18\) (Fig. 2h and Supplementary Table 3).
Enrichment for ASD risk genes among miRNA targets

Previous transcriptomic analyses have revealed convergent molecular pathology in ASD, which is characterized by downregulation of genes involved in neuronal and synaptic function, concomitant with upregulation of genes involved in immune-inflammatory response\textsuperscript{11}. We hypothesized that miRNAs that are differentially expressed in ASD may contribute to these perturbations by repressing their mRNA targets. Alternatively, they might function

\[ \text{log}_2(\text{fold change}) \text{ in frontal cortex} \]

\[ \text{log}_2(\text{fold change}) \text{ in temporal cortex} \]

\[ R = 0.96 \]

\[ P < 2.2 \times 10^{-16} \]

242 RNA samples from 3 brain regions (131 samples from 56 ASD cases; 111 samples from 42 controls)

Illumina small RNA library preparation and 50 bp single-end sequencing

Known and novel miRNA identification and quantification

Differential expression analysis

Weighted gene coexpression network analysis

Target prediction, gene list enrichment analyses, integration with RNA-seq data

Experimental validation in human neural progenitor cells

Diagnosis

Age

Sex

Region

Seizure

Medication

RIN

Brain bank

© 2016 Nature America, Inc., part of Springer Nature. All rights reserved.
as compensatory mechanisms that mitigate the existing mRNA dysregulation.

We bioinformatically predicted the mRNA targets of top differentially expressed miRNAs and hub miRNAs in the ASD-related modules (Online Methods), applying the well-established and widely used algorithm TargetScan, which searches for miRNA target sites in mRNA 3’ UTRs and evaluates the targeting efficacy and evolutionary conservation (while controlling for background conservatism) of the target sites6,19-21 (Online Methods). We then selected the top targets expressed in the temporal and frontal cortex (N.N.P., V. Swarup, T.G.B., M. Irinia, G. Ramaswami, unpublished observations) using two different criteria: (1) the strongest targets, which have the highest predicted targeting efficacy and are shared by two or more miRNAs (Online Methods), and (2) the most conserved target sites, which are more likely to have conserved physiological roles, but may not include newly evolved targets with species-specific functions (Online Methods). Overall, these two methods identified comparable numbers of targets, many overlapping, and some that were unique (Supplementary Table 4 and Online Methods).

To test the validity of the bioinformatic predictions, we overexpressed a well-studied miRNA, hsa-miR-21-5p, in human neural progenitor cells (hNPCs; Online Methods). We observed significant (one-sided t test, P < 2.2 × 10^-16) downregulation of hsa-miR-21-5p miRNA targets predicted by our methods, with a magnitude comparable to those supported by previous experimental evidence (miRTarBase release 4.5, http://mir tarbase.mbc.nctu.edu.tw22, Supplementary Fig. 5 and Supplementary Table 5).

To explore the relationship of the ASD-affected miRNAs to genes that have been previously implicated in ASD, we systematically assessed whether targets of the differentially expressed miRNAs or ASD-associated miRNA modules are enriched for these genes. We first tested enrichment for a set of ASD risk genes from the Simons Foundation Autism Research Initiative (SFARI) AutDB database23 (ASD SFARI; Online Methods), which have been implicated via common variant association, candidate gene studies, copy number variation (CNV) and genetic syndromes. For comparison, we also examined genes implicated in monogenic forms of intellectual disability.13 We found that top targets of the differentially expressed miRNAs, as well as the brown and magenta modules, showed significant (Fisher’s exact test, FDR < 0.05; Online Methods) enrichment for SFARI ASD genes, but not for intellectual disability genes (Fig. 3a and Supplementary Table 6), suggesting that targets of ASD-affected miRNAs are enriched for genes that are causally connected with ASD, but less so for genes that are related solely to intellectual disability.

We further examined additional classes of ASD-relevant genes: genes whose transcripts are bound by the fragile X mental retardation protein (FMRP)24,25, genes encoding postsynaptic density (PSD) proteins24,26, genes encoding chromatin modifiers24 and genes expressed preferentially during embryonic brain development24,27. Enrichment for FMRP targets and embryonically expressed genes (Fisher’s exact test, FDR < 0.05; Online Methods) was observed for most target groups (Fig. 3a and Supplementary Table 6). The most conserved targets of the upregulated miRNAs and miRNA modules significantly (Fisher’s exact test, FDR < 0.05; Online Methods) overlapped with PSD genes (Fig. 3a and Supplementary Table 6). The most conserved targets of the upregulated miRNAs were also enriched (Fisher’s exact test, FDR < 0.05; Online Methods) for chromatin modifiers (Fig. 3a and Supplementary Table 6).

Recent whole-exome sequencing studies have identified high-confidence, likely-gene-disrupting (LGD, including nonsense, splice site and frame shift) de novo variants (DNVs) in ASD1,2. We next asked whether these independently identified ASD risk genes are over-represented in the targets of ASD-affected miRNAs. For comparison, we also examined LGD DNVs in unaffected siblings, missense and synonymous DNVs, and LGD DNVs in individuals with schizophrenia24 (Online Methods). Given that DNV frequency has been shown to be linearly correlated with gene length,24 we applied logistic regression that incorporates gene coding region length covered in exome sequencing25 to assess enrichment (Online Methods). Notably, the top targets of the ASD-affected miRNAs and miRNA modules showed specific enrichment (FDR < 0.05; Online Methods) for genes harboring LGD DNVs in ASD probands, but little enrichment for genes with LGD DNVs in unaffected siblings or individuals with schizophrenia, or missense/synonymous DNVs (Fig. 3b and Supplementary Table 6). We also observed enrichment (FDR < 0.05; Online Methods) for a shorter list of genes hit by severe recurrent mutations (two or more LGD mutations, small indels, and mutations that remove start or stop codons24; Fig. 3b and Supplementary Table 6). We also found that the top targets of the upregulated miRNAs and miRNA modules were enriched (Fisher’s exact test, FDR < 0.05; Online Methods) for a list of 65 ASD risk genes that have been implicated through rare and de novo variations (ASD rare variants; Fig. 3a)28. The observation that the targets of the downregulated miRNAs were enriched for ASD risk genes affected by LGD DNVs (Fig. 3b), which are expected to disrupt protein function, suggests that these miRNAs might be part of a compensatory, or adaptive, mechanism, as their downregulation would favor target mRNA upregulation.

We next asked whether targets of the ASD-affected miRNAs are enriched for common genetic variants associated with ASD in genome-wide association studies (GWAS)29-31. We applied the INRICH method32,33 to assess the overlap between ASD-associated, linkage disequilibrium-based independent genomic intervals and the miRNA target genes, observing enrichment (multiple testing-corrected P < 0.10 via nested permutation) for ASD GWAS signals near the strongest targets of the differentially expressed miRNAs (Fig. 3c and Online Methods). Notably, the most substantial
enrichment was in the targets of the downregulated miRNAs, further suggesting that these might be compensatory changes. As a control, we performed the same analysis using GWAS data for Alzheimer's disease and schizophrenia and did not observe significant enrichment (Fig. 3c). The lack of enrichment in the control data sets was unlikely a result of fewer single-nucleotide polymorphisms (SNP) or...
The empirical and multiple-testing corrected LGD DNVs in individuals with schizophrenia. (c) prbF, female probands) and unaffected siblings (sib). Severe_recurMutation, genes targeted by protein-disrupting recurrent mutations; DNV_LGDs_SCZ, including LGD, missense, synonymous, and recurrent (recurMutation) mutations, in ASD-affected probands (prb, all probands; prbM, male probands; prbF, female probands) and unaffected siblings (sib). Severe_recurMutation, genes targeted by protein-disrupting recurrent mutations; DNV_LGDs_SCZ, including LGD, missense, synonymous, and recurrent (recurMutation) mutations, in ASD-affected probands (prb, all probands; prbM, male probands; prbF, female probands) and unaffected siblings (sib).

Figure 3: Enrichment of ASD risk genes among the top targets of ASD-affected miRNAs and miRNA modules. (a) Heat map showing enrichment (Fisher’s exact test) of ASD risk genes from SFARI or implicated by rare variants (ASD rare variants), intellectual disability genes (ID all), genes encoding transcripts bound by FMRP (FMRP targets), genes encoding proteins in the postsynaptic density (PSD), genes expressed preferentially in human embryonic brains (Embryonic), and genes encoding chromatin modifiers. ASD/ID overlap, the overlap between the ASD SFARI and ID all sets; ASD only and ID only, non-overlapping ASD SFARI and ID genes, respectively. (b) Heat map showing enrichment (logistic regression) of genes affected by DNVs, including LGD, missense, synonymous, and recurrent (recurMutation) mutations, in ASD-affected probands (prb, all probands; prbM, male probands; prbF, female probands) and unaffected siblings (sib). Severe_recurMutation, genes targeted by protein-disrupting recurrent mutations; DNV_LGDs_SCZ, LGD DNVs in individuals with schizophrenia. (c) Enrichment for overlap with linkage disequilibrium–based independent genomic regions associated with ASD (from Autism Genetic Resource Exchange (AGRE) or Psychiatric Genomics Consortium (PGC)), Alzheimer’s disease or schizophrenia in GWAS among the strongest miRNA targets. The empirical and multiple-testing corrected P values calculated using the INRICH program are shown where the corrected P < 0.10. (d) Heat map showing enrichment (Fisher’s exact test) for ASD-associated developmental gene coexpression modules in human cortex. In a, b and d, enrichment odds ratios (OR) and FDR corrected P values (Online Methods) are shown for enrichments with FDR < 0.05.
Figure 4  Relationship between miRNA and mRNA expression changes in post-mortem ASD cortex.  (a–c) Correlations between the PC1s of differentially expressed miRNAs (FDR < 0.05, log₂(fold change) ≥ 0.3) and differentially expressed mRNAs (FDR < 0.05) that are predicted targets (after regressing out disease status) in 101 cortex samples. (a) All differentially expressed miRNAs versus all differentially expressed mRNAs. (b) Upregulated miRNAs versus downregulated mRNAs. (c) Downregulated miRNAs versus upregulated mRNAs. Pearson correlation coefficients (R) and P values are shown below the plots. (d) Negative correlations between the PC1s of top miRNAs in the ASD-associated miRNA modules and their predicted targets in the ASD-associated mRNA models. mRNA modules are represented with network plots showing the top 20 most connected module genes. Pearson correlation coefficients (R) and P values are indicated. Correlations for the magenta and yellow miRNA modules were calculated using 45 younger samples (ages between 15 and 30 years), given their stronger disease association at younger ages relative to older ages (>30 years).
intervals tested, as all data sets included comparable numbers of SNPs and intervals (Supplementary Table 6).

A previous study examining the temporal expression trajectories of ASD genes in the developing human cortex identified gene coexpression modules that are enriched for ASD risk genes\(^1\). We therefore tested for over-representation of these modules in the targets of ASD-affected miRNAs. This analysis highlighted module M16, which is upregulated during early cortical development and enriched for genes that have been implicated in neural development and synaptic function\(^1\); genes in M16 were over-represented (Fisher’s exact test, FDR < 0.05; Online Methods) in most miRNA target groups (Fig. 3d and Supplementary Table 6).

**Figure 5** Enrichment for ASD-affected mRNAs and mRNA modules in the top targets of ASD-affected miRNAs. (a) Heat map showing enrichment (Fisher’s exact test) for ASD-affected mRNAs and mRNA modules in the top targets of ASD-related miRNA modules. P values were FDR corrected across six target groups for each mRNA group. Odds ratios and FDRs are shown for enrichments with FDR < 0.05. (b) Model for the role of miRNA dysregulation in ASD molecular pathology. The upregulated miRNAs and mRNA modules may have a contributory role by repressing ASD risk genes and neuronal/synaptic genes downregulated in ASD. The downregulated miRNAs and miRNA module may contribute to the upregulation of immune/inflammatory genes in ASD, but might also have a compensatory role given the enrichment of their targets for rare protein-disrupting and common genetic variants associated with ASD. (c,d) Enrichment (Fisher’s exact test) for ASD-affected mRNAs and mRNA modules in the strongest (c) or the most conserved (d) targets of individual candidate miRNAs.
Because the TargetScan algorithm searches for miRNA target sites in the 3' UTR regions of mRNAs, we also used logistic regression to assess enrichment while controlling for 3' UTR length and observed a similar enrichment for ASD-related genes (Supplementary Fig. 6a–c). Together, these results suggest a functional involvement of miRNAs perturbed in ASD in its molecular pathology.

We also assessed shared functions among the targets of ASD-affected miRNAs by enrichment for Gene Ontology (GO) terms (GO-Elite software; Online Methods). Targets of the upregulated miRNAs and the magenta module were enriched (FDR < 0.10) for genes related to neural development and several signaling pathways (Supplementary Fig. 7a,b). The top GO terms (P < 0.01) for the targets of the downregulated miRNAs and mRNA module concerned both neuronal processes and immune function (Supplementary Fig. 7a,b).

**Relationship between miRNA and mRNA expression changes**

To directly assess the role of miRNA dysregulation in ASD-associated mRNA level alterations, we next examined the relationship between miRNA and mRNA expression changes. We used mRNA expression data generated in a separate study that investigated mRNA expression changes in post-mortem ASD brain using RNA-seq in 163 cortex (frontal and temporal) samples from ASD cases and controls (N.N.P., V.Swarup, T.G.B., M. Irimia, G. Ramaswami, unpublished observations), 101 of which overlapped with our study. We evaluated the relationship...
between miRNA and mRNA expression changes in the same set of individuals and samples, while at the same time recognizing that ASD-associated mRNA level perturbations are likely driven by several regulatory mechanisms, including TFs and epigenetic changes.

DGE analysis for miRNAs using 106 covariate-matched samples identified 1,156 genes that were differentially expressed (FDR ≤ 0.05) between ASD and CTL cortex: 574 upregulated and 582 downregulated (N.N.P., V. Swarup, T.G.B., M. Irimia, G. Ramaswami, unpublished observations). Consistent with previous findings, the downregulated set was enriched for genes involved in neuronal and synaptic function, and the upregulated set was enriched for microglia- and astrocyte-related genes involved in immune-inflammatory function. We first compared the fold changes of differentially expressed miRNAs that are predicted targets of the differentially expressed miRNAs or mRNA modules to those of non-targets. Overall, we observed a trend consistent with a negative effect of miRNAs on target mRNA levels; miRNAs targeted by the upregulated miRNAs or modules showed lower (P < 0.05) fold changes than the non-targets, whereas miRNAs targeted by the downregulated miRNAs or module showed higher (P < 0.05) fold changes than the non-targets (Supplementary Fig. 8a–h). We further assessed the relationship between mRNA and miRNA differential expression signatures by examining the correlations between the PC1s of differentially expressed miRNAs and differentially expressed mRNAs that were predicted targets. We observed significant (Pearson’s correlation, P < 0.005) negative correlations (Fig. 4a–c), suggestive of negative regulation of the ASD-affected mRNAs by the ASD-associated miRNAs in the brain. Notably, it is unlikely that the result was driven by separate association of the miRNAs and mRNAs to ASD status, as we regressed out disease status from the expression data when deriving the PC1s (Fig. 4a–c) or computed the correlations in CTL samples or ASD samples alone (Supplementary Fig. 9a–c).

In a separate study, we identified gene coexpression modules significantly correlated with ASD status that were enriched for genes differentially expressed between ASD and control (N.N.P., V. Swarup, T.G.B., M. Irimia, G. Ramaswami, unpublished observations). Consistent with previous findings, three modules (M4, M10, M16) that were downregulated in ASD were related to neuronal and synaptic function, as revealed by GO analysis, whereas two modules (M9, M19) that were upregulated in ASD were related to immune-inflammatory response and enriched for genes highly expressed in astrocytes and microglia (N.N.P., V. Swarup, T.G.B., M. Irimia, G. Ramaswami, unpublished observations). To explore the potential regulatory relationship between ASD-affected miRNA and mRNA modules, we first asked whether the top miRNAs in the ASD-associated miRNA module were negatively correlated with their predicted targets in the ASD-associated mRNA modules. We observed
The strongest miRNA targets predicted by TargetScan were assessed using one-sided enrichment (FDR < 0.05) for the upregulated mRNAs and M9 mRNA module, whereas targets of several upregulated miRNAs showed enrichments (Fisher’s exact test, P < 0.05) for the downregulated mRNAs and the M4 and M16 mRNA modules (Fig. 5c,d and Supplementary Table 6).

**hsa-miR-21-3p targets neuronal genes downregulated in ASD**

We next experimentally tested the predicted downregulation of target mRNAs by miRNAs in ASD via *in vitro* perturbation of several dysregulated miRNAs in hNPCs. We first focused on hsa-miR-21-3p, which was the second most upregulated miRNA in ASD cortex (1.5-fold increase, FDR < 0.05; Supplementary Fig. 10a) and whose predicted targets exhibited prominent enrichment (Fisher’s exact test, P < 0.01) for downregulated mRNAs and the M16 mRNA module (Fig. 5c,d and Supplementary Table 6). hsa-miR-21-3p is conserved across vertebrates (Supplementary Fig. 10b) and widely expressed in different human brain regions throughout development (Supplementary Fig. 10d,e). However, its role in the CNS has not yet been explored. We overexpressed hsa-miR-21-3p in hNPCs and examined the consequent mRNA changes by RNA-seq (Supplementary Table 5). We found that predicted hsa-miR-21-3p target genes showed significant downregulation (one-sided t test, P < 2.2 × 10^{-16}) as a group compared with non-targets (Fig. 6a), again validating the bioinformatic predictions.

We next performed a series of enrichment analyses to characterize the overlap between validated hsa-miR-21-3p targets and with the miRNAs in ASD. We found that the targets of several downregulated miRNAs were enriched (Fisher’s exact test, P < 0.05) for the upregulated mRNAs and M9 mRNA module, whereas targets of several upregulated miRNAs showed enrichments (Fisher’s exact test, P < 0.05) for the downregulated mRNAs and the M4 and M16 mRNA modules (Fig. 5c,d and Supplementary Table 6).

**hsa-miR-21-3p targets neuronal genes downregulated in ASD**

We next experimentally tested the predicted downregulation of target mRNAs by miRNAs in ASD via *in vitro* perturbation of several dysregulated miRNAs in hNPCs. We first focused on hsa-miR-21-3p, which was the second most upregulated miRNA in ASD cortex (1.5-fold increase, FDR < 0.05; Supplementary Fig. 10a) and whose predicted targets exhibited prominent enrichment (Fisher’s exact test, P < 0.01) for downregulated mRNAs and the M16 mRNA module (Fig. 5c,d and Supplementary Table 6). hsa-miR-21-3p is conserved across vertebrates (Supplementary Fig. 10b) and widely expressed in different human brain regions throughout development (Supplementary Fig. 10d,e). However, its role in the CNS has not yet been explored. We overexpressed hsa-miR-21-3p in hNPCs and examined the consequent mRNA changes by RNA-seq (Supplementary Table 5). We found that predicted hsa-miR-21-3p target genes showed significant downregulation (one-sided t test, P < 2.2 × 10^{-16}) as a group compared with non-targets (Fig. 6a), again validating the bioinformatic predictions.

We next performed a series of enrichment analyses to characterize the overlap between validated hsa-miR-21-3p targets and

![Figure 8](image-url)  
*Figure 8* Experimental validation of targets of other candidate miRNAs. (a–c) Distributions (left) and cumulative distributions (right) of mRNA log2(fold change) in response to overexpression of hsa-miR-103a-3p (a), hsa-miR-143-3p (b) and hsa-miR-23a-3p (c) in hNPCs. Statistical significance between target groups and non-targets was assessed using one-sided t tests assuming unequal variance. (d) Enrichment (Fisher’s exact test) of validated targets of hsa-miR-103a-3p and hsa-miR-143-3p for downregulated mRNAs and the downregulated M16 mRNA module. (e) Enrichment (Fisher’s exact test) of validated targets of hsa-miR-21-3p for ASD SFARI genes and ASD risk genes implicated by rare variants. (f) A partial list of validated target genes.
ASD-related genes. hsa-miR-21-3p targets showed enrichment for genes harboring LGD and recurrent DNVs in ASD-affected individuals (logistic regression, \( P < 0.05 \)), but little enrichment for missense or synonymous DNVs, or LGD DNVs in unaffected siblings or schizophrenia-affected individuals (Fig. 6c,f and Supplementary Table 6). ASD SFARI genes, ASD rare variants, FMRP targets, chromatin modifiers, embryonically expressed genes, and the ASD-related M3 and M16 brain developmental modules\(^{13}\) were also enriched (Fisher’s exact test, \( P < 0.05 \)), but genes associated with intellectual disability were not (Fig. 6b,d and Supplementary Table 6). Enrichment (Fisher’s exact test, \( P < 0.005 \)) was also observed for genes downregulated in post-mortem ASD cortex, particularly the M16 mRNA module (Fig. 6e,f and Supplementary Table 6), the ME of which was negatively correlated with hsa-miR-21-3p (Pearson’s \( R = -0.50 \), \( P = 8.5 \times 10^{-8} \)). Notably, hsa-miR-21-3p overexpression downregulated (one-sided \( t \) test, \( P < 0.05 \)) several hub genes in M16, including PAFAH1B1/LIS1 (which is critical for neuronal migration and is causally linked to lissencephaly\(^{35}\)), DLGAP1 (a PSD scaffold protein that binds to SHANK3, an ASD risk gene\(^{1}, \) and ATTP2B1/PMCA1 (a calcium transporter). In addition, hsa-miR-21-3p repressed the levels of ATTP1B1 (a PSD-localized Na\(^+\)/K\(^+\) transporter harboring ASD-associated LGD DNVs\(^{23}\)), DYNCI1 (which is critical for neuronal migration\(^{23}\)), NEEP21 (which is involved in synaptic transmission\(^{23}\)), SV2B (a neurotransmitter release regulator\(^{23}\)) and several genes in the ubiquitin-proteasome pathway (UCHL5, USP33, USP7, UBE2K, FBXO11 and KIAA0368) (Fig. 6d and Supplementary Table 5). Also of interest, hsa-miR-21-3p overexpression led to a pronounced decrease (38%; one-sided \( t \) test, \( P = 0.0001 \)) in PCDH19, mutations in which cause epilepsy and ASD in females\(^ {38} \). hsa-miR-21-3p and PCDH19 levels were also negatively correlated in post-mortem cortex (Pearson’s \( R = -0.48 \), \( P = 4.0 \times 10^{-7} \)), consistent with the \textit{in vitro} observation. Together, these results implicate hsa-miR-21-3p in regulating the mRNA levels of neuronal and synaptic genes and suggest a role for hsa-miR-21-3p in the downregulation of these genes in ASD.

**hsa_can_1002-m targets genes in the EGFR and FGFR pathways**

Another candidate of particular interest was the predicted miRNA hsa-can_1002-m, one of the most downregulated miRNAs in ASD cerebral cortex (2.5-fold decrease, \( FDR < 0.05 \); Supplementary Fig. 10a) and the top hub miRNA of the brown module (Fig. 2f and Supplementary Table 2). Overexpression of hsa-can_1002-m in hNPCs led to significant decreases (one-sided \( t \) test, \( P < 10^{-7} \)) in its predicted targets (Fig. 7a and Supplementary Table 5), supporting its function. Notably, the predicted hsa-can_1002-m precursor is located in a genomic region that is only present in primates, and not in lower vertebrates (Supplementary Fig. 10c). Moreover, RNA-seq in the mouse cortex did not detect the hsa-can_1002-m sequence. We further performed qRT-PCR using a primer for the human mature sequence, detecting robust expression in human, chimpanzee and rhesus macaque cerebral cortices, but not the mouse cortex (Fig. 7b), confirming our finding from the genome sequence that hsa-can_1002-m is a primate-specific miRNA. RNA-seq data from the BrainSpan project revealed that hsa-can_1002-m is broadly expressed, but is developmentally regulated in the human brain from infancy to adulthood (Supplementary Fig. 10f,g).

Validated hsa-can_1002-m targets did not show enrichment for known ASD risk genes. Nonetheless, the top GO categories (GO-Elite software, \( P < 0.001 \); Online Methods) implicated the epidermal growth factor receptor (EGFR) and the fibroblast growth factor receptor (FGFR) signaling pathways (Fig. 7c–e and Supplementary Table 6), which have been implicated in neuron and glia proliferation, differentiation, and survival in both the developing and adult brain, as well as inflammatory/immune processes\(^ {39,40} \). Notably, all implicated targets are involved in the activation of these pathways, suggesting that hsa-can_1002-m downregulation would lead to an increase in pathway activity. These targets form a more highly connected local protein-protein interaction network than would be expected by chance (DAPPLE software, \( P < 0.05 \); Online Methods), providing independent confirmation of co-regulation at the protein level (Fig. 7d). Of particular interest are EPS8 (a signaling adaptor protein regulating dendritic spine density and synaptic plasticity), ADAM12 (a metalloprotease required for EGFR ligand processing), CHUK (a kinase critical for NF-kappa-B activation) and RUNX1 (a transcription factor essential for immune cell development and activation), all of which were upregulated (\( P < 0.05 \)) in ASD post-mortem cortex (N.N.P., V. Swarup, T.G.B., M. Irimia, G. Ramaswami, unpublished observations) (Supplementary Table 5). In addition, the validated strongest hsa-can_1002-m targets were enriched for genes in the immune-related, upregulated M9 mRNA module (Fig. 7e), consistent with prediction (Fig. 5c). Together, these findings implicate a previously unknown, primate-specific miRNA hsa-can_1002-m in regulating the EGFR and the FGFR signaling pathways, shedding light on its potential role in neuronal and glial development and function, as well as in ASD molecular pathology involving neural-immune interactions.

**Experimental characterization of other candidate miRNAs**

We also experimentally validated the targets of several other candidate miRNAs, including hsa-miR-103a-3p in the yellow module and hsa-miR-143-3p and hsa-miR-23a-3p in the magenta module, the predicted targets of which were enriched (Fisher’s exact test, \( P < 0.05 \)) for downregulated mRNAs and the M16 mRNA module (Fig. 5c,d and Supplementary Table 6). Notably, hsa-miR-23a-3p has also been reported to be upregulated in lymphoblasts in ASD\(^ {41} \), and hsa-miR-143-3p has been recently shown to be regulated by a primate-specific long non-coding RNA and has been implicated in neural progenitor proliferation\(^ {42} \). Overexpression of these miRNAs in hNPCs led to significant reductions (one-sided \( t \) test, \( P < 2.2 \times 10^{-16} \)) in the predicted target mRNAs (Fig. 8a–c and Supplementary Table 5). Consistent with bioinformatic predictions, enrichment analysis revealed that the validated targets of hsa-miR-103a-3p and hsa-miR-143-3p were enriched (Fisher’s exact test, \( P < 0.05 \)) for downregulated mRNAs and/or the M16 mRNA module (Fig. 8d,f and Supplementary Table 6). Validated targets of hsa-miR-23a-3p were enriched (Fisher’s exact test, \( P < 0.05 \)) for ASD SFARI genes and ASD rare variants (Fig. 8e–f and Supplementary Table 6). These data are consistent with a potential functional involvement of these miRNAs in ASD, making them interesting candidates for further functional manipulation in model systems.

**DISCUSSION**

Our genome-wide, integrative analysis provides new insights into the role of miRNAs in ASD pathophysiology. We observed a miRNA differential expression signature that was shared by a majority of ASD cortex samples. Within the targets of the ASD-affected miRNAs and miRNA coexpression modules, we observed enrichment for ASD risk genes that have been implicated by multiple forms of genetic variation, and much less so for variants associated with intellectual disability, schizophrenia or Alzheimer’s disease. This suggests that ASD risk genes are highly dosage sensitive; we surmise that miRNA dysregulation
provides an alternative pathway for gene-disrupting mutations to perturb key transcript levels, thereby potentially contributing to ASD susceptibility (Fig. 5b). This model is supported by the negative correlation between the expression changes of ASD-affected miRNAs and mRNAs, and our experimental validation showing regulation of mRNA targets by several top candidate miRNAs. Collectively, our findings suggest that ASD-associated transcriptomic changes may be partially attributable to miRNA dysregulation, with the upregulated miRNAs potentially contributing to the downregulation of neuronal and synaptic genes and the downregulated miRNAs contributing to the upregulation of immune-inflammatory genes, as well as possible compensatory changes (Fig. 5b).

A few studies have also examined miRNA expression changes associated with ASD, but assessed a limited number of miRNAs (using qRT-PCR or microarray), had a small sample size and/or used non-neuronal tissues and cells41,43,44. Our genome-wide analysis using the most relevant tissue and a better-powered sample, along with integration with multiple gene sets and expression data, provides the most robust and comprehensive assessment of miRNA dysregulation in ASD brain to date. Most of the differentially expressed miRNAs that we identified have not been reported. Notably, several miRNAs, including hsa-miR-107, hsa-miR-106a-5p, hsa-miR-10a-5p, has-miR-136-5p and has-miR-155-5p, overlapped with findings from previous studies and would be interesting candidates for further experimental investigation.

In addition to providing a systems-level view of the miRNA expression landscape in post-mortem ASD brains, we also experimentally characterized the targets of several top candidate miRNAs in hNPCs.

We found that transcripts regulated by hsa-miR-21-3p, an upregulated miRNA of unknown function in the nervous system, showed enrichment for ASD candidate genes and genes downregulated in ASD cortex. Its connection with the M16 mRNA module, which is enriched for neuronal and synaptic genes, suggests a role in regulating neuronal development and synaptic function and a link with the neuronal and synaptic defects in ASD.

We also found that hsa_can_1002-m, a previously unknown, primate-specific miRNA that is downregulated in ASD, regulates several transcripts involved in the EGFR and the FGFR signaling pathways. This is intriguing from an evolutionary point of view given the important roles of these pathways in regulating neural stem cell proliferation in the brain, as a rapid increase in brain size resulting from increased neural stem cell division has been suggested as a critical step in primate brain evolution45. In addition, early postnatal brain overgrowth followed by relative growth normalization has been repeatedly observed in ASD-affected children and is thought to reflect abnormal early neurodevelopment46. It is possible that negative regulatory mechanisms, such as miRNAs, have evolved to restrict increased cell proliferation in the primate brain, as uncontrolled proliferation would disrupt brain development and function.

The observation that the cerebellum showed a similar trend of miRNA differential expression to the cortex was somewhat unexpected, given previous findings11 that very few mRNAs are differentially expressed in the cerebellum. Our recent RNA-seq analysis (N.N.P., V. Swarup, T.G.B., M. Irimia, G. Ramaswami, unpublished observations) revealed that some mRNAs showed similar trends of differential expression in the cerebellum, albeit of substantially smaller magnitude, compared with the cortex (N.N.P., V. Swarup, T.G.B., M. Irimia, G. Ramaswami, unpublished observations).

We speculate that this might be a result of the differences in cell types or some other aspect of the molecular milieu of the cerebellum that renders it resilient to the mRNA changes observed in the cortex.

One limitation of gene expression studies using post-mortem brains is that, although ASD likely arises from abnormalities during early brain development, the majority of available samples are from adults. Although there was no clear association of miRNA or mRNA11 (N.N.P., V. Swarup, T.G.B., M. Irimia, G. Ramaswami, unpublished observations) perturbations with medication history, some observed changes almost certainly reflect the consequences (rather than the causes) of having the disorder or compensatory responses, whereas some early gene expression perturbations that have a causal role may not be captured in adult brain. In this regard, it is interesting to note that the two miRNA modules upregulated in ASD showed stronger disease association at younger ages than at older ages, suggesting that they might be more related to early pathogenic features. Future studies using brain samples from younger patients and patient-derived brain organoids that model early brain development47 can provide more insights into early transcriptomic perturbations. Another limitation is the cellular heterogeneity of post-mortem tissue; cell-type-specific mRNA expression data in the human brain are still lacking. The observed expression changes could occur in a single or multiple cell types, or reflect changes in cell composition. Future studies using transcriptomic profiling of isolated cell types and single-cell sequencing48 could help to resolve these possibilities.

Collectively, our genome-wide, integrative analysis provides a framework for assessing the functional involvement of miRNA in ASD. By integrating several ASD candidate gene sets and correlating with mRNA expression data, we provide multiple lines of evidence for a functional role of miRNA dysregulation in ASD, either as contributory or compensatory factors. These analyses also identify a rich set of ASD-associated candidate miRNAs for further study.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. Raw small RNA-seq andRNA-seq data from brain samples from ASD cases and controls have been deposited to the PsychENCODE Knowledge Portal (http://dx.doi.org/10.7303/syn4587609).

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We would like to acknowledge the BrainSpan Consortium principal investigator N. Sestan (Yale University) for providing BrainSpan microRNA data and I. Ou for excellent technical assistance. This work was supported by US National Institutes of Health grant R01MH094714 to D.H.G. and is part of the PsychEncode Consortium.

AUTHOR CONTRIBUTIONS

N.N.P. and T.G.B. performed brain sample dissections. Y.E.W. performed the other experiments and data analyses. N.N.P. provided code for differential gene expression, coexpression and enrichment analyses. D.H.G. provided guidance and oversight on all experiments and analyses. Y.E.W. and D.H.G. wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Bourgeron, T. From the genetic architecture to synaptic plasticity in autism spectrum disorder. Nat. Rev. Neurosci. 16, 551–563 (2015).
2. Geschwind, D.H. & Stale, M.W. Gene hunting in autism spectrum disorder: on the path to precision medicine. Lancet Neurol. 14, 1109–1120 (2015).
3. Gaugler, T. et al. Most genetic risk for autism resides with common variation. Nat. Genet. 46, 881–885 (2014).
1. Geschwind, D.H. & Flint, J. Genetics and genomics of psychiatric disease.
Science 349, 1489-1494 (2015).
2. Ha, M. & Kim, V.N. Regulation of microRNA biogenesis. Nat. Rev. Mol. Cell Biol. 15, 509-524 (2014).
3. Friedman, R.C., Farh, K.K.H., Burge, C.B. & Bartel, D.P. Most mammalian mRNAs are conserved targets of microRNAs. Genes Dev. 19, 92-105 (2009).
4. Darnell, J.C. et al. A primate lncRNA mediates Notch signaling during neuronal development by sequestering miRNA. Nature 511, 421-427 (2014).
5. Geschwind, D.H. & Flint, J. Genetics and genomics of psychiatric disease.
6. Friedman, R.C., Farh, K.K.H., Burge, C.B. & Bartel, D.P. Most mammalian mRNAs are conserved targets of microRNAs. Genes Dev. 19, 92-105 (2009).
7. Ha, M. & Kim, V.N. Regulation of microRNA biogenesis.
8. Oldham, M.C. et al. Functional organization of the transcriptome in human brain. Nat. Neurosci. 11, 1271-1282 (2008).
9. Parikshak, N.N. et al. Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. Cell 155, 1008-1021 (2013).
10. Zhang, B. & Horvath, S. A general framework for weighted gene co-expression network analysis. Stat. Appl. Genet. Mol. Biol. 4, Article17 (2005).
11. Langfelder, P., Luo, R., Oldham, M.C. & Horvath, S. Is my network module preserved and reproducible? PLoS Comput. Biol. 7, e1001057 (2011).
12. Arbizu, L. et al. Genome-wide inference of natural selection on human transcription factor binding sites. Nat. Genet. 45, 723-729 (2013).
13. Ronan, J.L., Wu, W. & Crabtree, G.R. From neural development to cognition: unexpected roles for chromatin.
14. Lewis, B.P., Burge, C.B. & Bartel, D.P. Conserved seed pairing, often flanked by 25. Darnell, J.C. MicroRNAs in neuronal function and autism.
15. Lewis, B.P., Burge, C.B. & Bartel, D.P. Conserved targets of microRNAs. Genes Dev. 19, 92-105 (2009).
16. Ronan, J.L., Wu, W. & Crabtree, G.R. From neural development to cognition: unexpected roles for chromatin.
17. Lewis, B.P., Burge, C.B. & Bartel, D.P. Conserved seed pairing, often flanked by 25. Darnell, J.C. MicroRNAs in neuronal function and autism.
18. Ronan, J.L., Wu, W. & Crabtree, G.R. From neural development to cognition: unexpected roles for chromatin.
19. Lewis, B.P., Burge, C.B. & Bartel, D.P. Conserved targets of microRNAs. Genes Dev. 19, 92-105 (2009).
20. Grimson, A. et al. MicroRNA targeting specificity in mammals: determinants beyond seed pairing. Mol. Cell 27, 91-105 (2007).
21. Garcia, D.M. et al. Weak seed-pairing stability and high target-site abundance decrease the proficiency of lso-6 and other microRNAs. Nat. Struct. Mol. Biol. 18, 1139-1146 (2011).
22. Hsu, S.-D. et al. miRTarBase update 2014: an information resource for experimentally validated miRNA-target interactions. Nucleic Acids Res. 42, D78-D85 (2014).
23. Basu, S.N., Kollu, R. & Banerjee-Basu, S. AutDB: a gene reference resource for autism research. Nucleic Acids Res. 37, D832-D836 (2009).
24. Jossifov, I. et al. The contribution of de novo coding mutations to autism spectrum disorder. Nature 515, 216-221 (2014).
25. Darnell, J.C. et al. FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. Cell 146, 247-261 (2011).
26. Bayés, A. et al. Characterization of the proteome, diseases and evolution of the human postsynaptic density. Nat. Neurosci. 14, 19-21 (2011).
27. Kang, H.J. et al. Spatio-temporal transcriptome of the human brain. Nature 478, 483-489 (2011).
28. Sanders, S.J. et al. Insights into autism spectrum disorder genomic architecture and biology from 71 risk loci. Neuron 87, 1215-1233 (2015).
29. Arney, R. et al. Individual common variants exert weak effects on the risk for autism spectrum disorders. Hum. Mol. Genet. 21, 4781-4792 (2012).
30. Wang, K. et al. Common genetic variants on 5p14.1 associate with autism spectrum disorders. Nature 459, 528-533 (2009).
31. Cross-Disorder Group of the Psychiatric Genomics Consortium. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. Lancet 381, 1371-1379 (2013).
32. Lee, P.H., O'Dushlaine, C., Thomas, B. & Purcell, S.M. INRICH: interval-based enrichment analysis for genome-wide association studies. Bioinformatics 28, 1797-1799 (2012).
33. Rijpe, S. et al. Biological insights from 108 schizophrenia-associated genetic loci. Nature 511, 421-427 (2014).
34. Lambert, J.-C. et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer’s disease. Nat. Genet. 45, 1452-1458 (2013).
35. Hirokawa, N., Niwa, S. & Tanaka, Y. Molecular motors in neurons: transport mechanisms and roles in brain function, development, and disease. Neuron 68, 610-638 (2010).
36. Lasiecka, Z.M. & Winckler, B. Mechanisms of polarized membrane trafficking in neurons -- focusing in on endosomes. Mol. Cell. Neurosci. 48, 278-287 (2011).
37. Wan, Q.-F. et al. SV2 acts via presynaptic calcium to regulate neurotransmitter release. Neuron 66, 884-895 (2010).
38. Redies, C., Hertel, N. & Höbner, C.A. Cadherins and neuropsychiatric disorders. Brain Res. 1470, 130-144 (2012).
39. Hong, R.W.C. & Guillaud, L. The role of epidermal growth factor and its receptors in mammalian CNS. Cytokine Growth Factor Rev. 15, 147-156 (2004).
40. Turner, C.A., Akil, H., Watson, S.J. & Evans, S.J. The fibroblast growth factor system and mood disorders. Biol. Psychiatry 59, 1128-1135 (2006).
41. Meier, D. et al. The emerging role of microRNAs in schizophrenia and autism spectrum disorders. Front. Psychiatry 3, 39 (2012).
42. Rani, N. et al. A primate IncRNA mediates Notch signaling during neuronal development by sequestering miRNA. Nature 90, 1174-1188 (2016).
43. Mundali Vasu, M. et al. Serum microRNA profiles in children with autism. Mol. Autism 5, 40 (2014).
44. Mor, N., Nardone, S., Sams, D.S. & Elliott, E. Hypomethylation of miR-142 promoter and upregulation of microRNAs that target the oxytocin receptor gene in the autism prefrontal cortex. Mol. Autism 6, 46 (2015).
45. Geschwind, D.H. & Rakic, P. Cortical evolution: judge the brain by its cover. Nature 501, 633-647 (2013).
46. Geschwind, D.H. Advances in autism. Annu. Rev. Med. 60, 367-380 (2009).
47. Mariani, J. et al. FOXG1-dependent dysregulation of GABA/glutamate neuron differentiation in autism spectrum disorders. Cell 162, 375-390 (2015).
48. Gawad, C., Koh, W. & Quake, S.R. Single-cell genome sequencing: current state of the science. Nat. Rev. Genet. 17, 175-188 (2016).
ONLINE METHODS

Brain tissue samples. Brain tissue samples were acquired from the Autism Tissue Program (ATP) brain bank at the Harvard Brain and Tissue Bank, the National Institute for Child Health and Human Development (NICHD) Eunice Kennedy Shriver Brain and Tissue Bank for Developmental Disorders, the UK Brain Bank for Autism and Related Developmental Research (BBA), and the MRC London Neurodegenerative Diseases Brain Bank. Up to three brain regions from each individual were assessed: frontal cortex (FC, Brodmann area (BA) 9), temporal cortex (TC, BA41/42/22), and cerebellar vermis. For some individuals, not all three regions were included due to limited tissue availability. Metadata for all 242 samples, including age, sex, brain region, brain bank, medical history, and sample quality metrics are summarized in Supplementary Table 1. Individuals defined as autistic had either a confirmed ADI-R diagnosis, duplication 15q syndrome with confirmed ASD, or a diagnosis of autism supported by other evidence such as clinical history. Dissections of frozen samples were performed on dry ice in a dehydrated dissection chamber, and randomized for balance of age, sex, brain region and diagnostic status.

RNA extractions, library preparation and small RNA sequencing. Total RNA was extracted from approximately 100 mg of frozen tissue using the miRNeasy kit (Qiagen). RNA integrity number (RIN) of the extracted total RNA was measured using an Agilent Bioanalyzer. For 195 of the 242 samples, rRNAs were depleted from 2 µg total RNA with the Ribo-Zero Gold kit (Illumina). Remaining RNA was size selected with AMPure XP beads (Beckman Coulter). Small RNAs (including miRNAs) in the supernatant (250 µl) were precipitated with 2 µl glycogen (Roche), 2 µl 0.1x Pellet Paint NF Co-precipitant (Mercck Millipore), 25 µl 3 M NaOAc (pH 5.2), and 700 µl 100% ethanol (−20 °C), and resuspended in 5 µl RNase-free water for subsequent library preparation. For the other 47 samples, 0.7 µg total RNA was directly used for library preparation, due to the lack of sufficient material for rRNA depletion. We compared sequencing results on 4 brain tissue samples for which both library preparation methods were used and found the results on the same sample to be highly positively correlated (Pearson correlation coefficients = 0.93-0.97, P < 2.2 × 10−16). Small RNA libraries were prepared in batches of 12–48 samples using the TruSeq Small RNA library Preparation Kits (Illumina) according to the manufacturer's protocol. Library preparation was randomized for balance of diagnostic status (ASD versus CTL), age, sex and brain region. Libraries were then validated using an Agilent Bioanalyzer and quantified with the Qubit dsDNA HS assay (Life Technologies). 19–29 libraries barcoded with Illumina TruSeq small RNA indexes were pooled and sequenced in each lane on a Illumina HiSeq2500 instrument using the high output mode with 50 bp single-end reads. Investigators were blinded during dissection, RNA extraction, library preparation and sequencing to all metadata information about the samples.

Quality control (QC). Sequencing reads were demultiplexed using CASAVA v1.8.2 (Illumina) and sequencing adapters were removed using the fastx_clipper function in the FASTX- Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/index.html). Sequencing quality (including quality scores, GC content, sequence length distribution, duplication levels, overrepresented sequences, and Kmer content) was examined using FastQC v0.10.1 (http://www.bioinformatcs.babraham.ac.uk/projects/fastqc/). All 242 samples sequenced for small RNAs were also sequenced for miRNAs using RNA-seq (N.N.P., V. Swarup, T.G.B., M. Irimia, G. Ramaswami, unpublished observations). Genotypes for sites that are heterozygous or homozygous for the minor allele relative to the reference genome were called from RNA-seq data49 (N.N.P., V. Swarup, T.G.B., M. Irimia, G. Ramaswami, unpublished observations). Genotypes were then coded as NA (homozygous for the minor allele). Pairwise Spearman correlations between samples were calculated, based on which the samples were clustered. Any sample that did not cluster with other samples from the same individual was further examined for possible contamination or sample mix-up and excluded from the downstream analyses if sample mix-up was not resolvable. 20 samples from 18 individuals were removed using this criterion. In addition, 6 samples from 2 individuals who were not diagnosed with ASD, but had other conditions (one with idiopathic epilepsy, one with copy number variation) were also removed to avoid confounding effects. This resulted in a total of 216 samples that passed QC.

miRNA quantification and prediction. For quantification of mature miRNAs documented in miRBase release 20, sequencing reads were first mapped to the hg19 reference genome using the mapper.pl script in the miRDeep2 package50 with reads < 18 nt discarded. The quantifier.pl script was then used with default settings to quantify the number of reads mapped to mature miRNAs, allowing 0 mismatch.

For the prediction of novel miRNAs, two methods were used. First, the miRDeep2.pl script in the miRDeep2 package was run with default settings and mature miRNAs in Pan troglodytes, Gorilla gorilla, Pans paniscus, and Pongo pygmaeus (miRBase release 20) provided as related species. miRDeep2 examines the position and frequency of reads aligned to the genome (signature) with respect to a putative RNA hairpin and scores miRNA candidates employing a probabilistic model based on miRNA biogenesis. The score reflects the energetic stability of the putative hairpin and the compatibility of the observed read distribution with miRNA cleavage. The higher the score, the more likely that the miRNA candidate is a true positive. We only kept predictions with a score ≥ 4 (corresponding to a true positive probability of 78 ± 2%; names of novel miRNAs predicted using this method start with "hsa_chr"). Second, we applied miRanalyzer51,52, which predicts putative mature miRNAs and precursors based on mapped reads and folding energy and employs 5 different Random Forest models to calculate the probability that a candidate is a true miRNA. We only kept candidates with positive predictions in at least 4 out of all 5 models (names of novel miRNAs predicted using this method start with hsa-P). In addition, we included novel miRNAs identified in a recent study, which were predicted using miRDeep2 based on 94 human sRNA-seq data sets49 (names start with hsa_can). Many of these were supported by various levels of experimental evidence, including interaction with Ago1/2, interaction with DGC8R, response to silencing of the miRNA biogenesis machinery, and/or interaction with target miRNAs in CLASH experiments10. Duplicate predictions in the above three novel miRNA sets were collapsed. Predicted miRNAs > 25 nt were removed according to the normal size range of miRNAs. The quantifier.pl script was used to quantify the number of reads mapped to novel mature miRNAs with 0 mismatch.

miRNAs (from miRBase release 20 or predicted) with read counts ≥ 3 in at least 50% of samples in each region, sex, or diagnostic status (ASD versus CTL) group were kept for further analysis. 699 miRNAs (552 in miRBase 20 and 147 predicted) in our data set met this criterion. This step helps remove miRNAs that are supported by only a few reads and likely expressed at very low levels. Such low levels of expression are unlikely to provide reliable, statistically significant differential expression results.

Expression value normalization and adjustment of covariates. Raw expression data for all 242 samples were normalized for library size using the estimateSizeFactors function in the DESeq R package53. For each miRNA in each sample, a ratio is calculated by dividing the read count by the geometric mean across all samples. A scaling factor for each sample is then calculated as the median of this ratio for all miRNAs in the sample. The raw read counts in each sample are then divided by the scaling factor to generate the library size-normalized data. The effect of GC content was normalized using the CQN R package44. Library preparation batch effects were normalized using the ComBat function in the sva R package45. The effects of other technical covariates, including INI, PMI, and brain bank, were normalized together using a linear model. Additionally, miRNAs prepared from the same total RNA samples used in this study were also analyzed using RNA-seq in a separate study (N.N.P., V. Swarup, T.G.B., M. Irimia, G. Ramaswami, unpublished observations). We observed that the proportion of miRNA reads mapped to exons also showed significant correlation with the miRNA expression data, as it likely reflected the ratio between cytoplasmic (where miRNAs are in the processed, mature form) and nuclear (where miRNAs are in the precursor form) RNAs. Therefore we also included this technical variable in the linear regression model along with other technical variables. The log-transformed expression data normalized for library size and technical covariates were used for subsequent differential gene expression analysis and weighted gene coexpression network analysis.

Definition of sample sets for analysis. In exploratory data analysis, we performed principle component analysis (PCA) and hierarchical sample clustering for 216 samples that passed QC (180 samples prepared using the rRNA depletio

DOI: 10.1038/nn.4373
method and 36 samples prepared using total RNA) to examine the relationship between the expression data and different covariates. We observed that library preparation method (rRNA depletion versus total RNA) and brain region (cortex versus cerebellum) had major effects on the expression profile, both having strong correlations with PC1 and PC2 (Supplementary Fig. 2a,b). Accordingly, samples prepared with the two different methods or from different brain regions (cortex versus cerebellum) were clustered into distinct branches in hierarchical clustering (Supplementary Fig. 2c). In addition, brain bank, age, RIN, and PMI also showed significant correlations with PC1–5 (Supplementary Fig. 2a).

To avoid the confounding effects of library preparation method and brain region, we divided our samples into different subsets based on these two covariates, and first focused on cortex samples prepared using the rRNA depletion method (116 samples) for our main analysis. We performed outlier detection using a standard method. Specifically, we calculated the connectivity between samples based on the biweight midcorrelation of miRNA expression using the fundamentalNetworkConcepts function in the WGCNA R package, and removed samples with connectivity more than 3 s.d. away from the mean as outliers. This process was then repeated on the remaining samples until no more outliers were detected. Using this method, we removed 2 samples as outliers. We also removed 5 samples (from three individuals) for which PMI information was not available, resulting in 109 samples, which were used for the main WGCNA analysis. For the main DGE analysis, to avoid the potential confounding effect of age, we further removed 14 ASD samples for which there were no age-matched control samples, including 13 samples with age ≤ 11 years and 1 sample with age of 67 years (ages of ASD and CTL groups were both between 15 - 60 years after removing these samples), resulting in 95 samples. The WGCNA analysis relies on correlation between gene expression levels across samples and is not as critically affected by unmatched covariates as the DGE analysis. It also permits direct assessment of the relationship of modules to experimental variables or confounders. Therefore we used all 109 cortex samples to maximize the robustness of the WGCNA analysis. We combined samples from the frontal cortex and the temporal cortex, as PCA analysis and hierarchical sample clustering indicated that miRNA expression profiles were very similar between these two regions, but distinct in the cerebellum (Supplementary Fig. 2a–f). The cortex samples prepared using the total RNA method (31 samples) were imbalanced in brain bank between ASD and CTL: 10 of 16 ASD samples were from UK BBA while none of the 15 CTL samples was from this brain bank. To avoid potential confounding effects, this set was not used for DGE analysis, but instead for evaluating miRNA module preservation, which relies on miRNA coexpression relationships across samples and should not be critically affected by unmatched covariates.

For DGE analysis in the cerebellum, we selected 47 samples prepared using the rRNA depletion method as follows. From 64 samples that passed QC, four samples were removed as outliers as described above. We also removed three samples with no PMI information, one sample with very high PMI (50 h), three young ASD samples (age ≤ 5 years) with no age-matched controls, and six ASD samples with RINs lower than the lowest RIN in CTL.

Differential gene expression (DGE) analysis. Differential expression between ASD and CTL was calculated for each miRNA in the cortex using a linear mixed-effects (LME) model using the R package nlme, as more than one brain region from the same individuals were included (Supplementary Code 1). The model treated diagnosis, age, sex, and region as fixed effects (numeric or factor variables) and individual brain ID as a random effect: \text{logit}(p) = \beta_0 + \beta_1\text{region} + \beta_2\text{age} + \beta_3\text{sex} + \beta_4\text{diagnosis} + \epsilon.

Prediction of miRNA targets. For prediction of miRNA targets, the stand-alone version of TargetScan (release 6.2) was used. 3' UTR sequences of RefSeq genes were downloaded from the TargetScan website (http://www.targetscan.org). For each miRNA target site, a branch length score, which evaluates target site conservation while controlling for 3' UTR conservation, and a context+ score, which measures target efficacy irrespective of conservation, were calculated. To select top differentially expressed miRNAs and top miRNAs in ASD-associated.
Gene ontology analysis was performed using TargetScan. In a few cases where two miRNAs are 3′ isoforms, we used only one isoform for target prediction to prevent duplicate predictions. Using these criteria, 10 downregulated miRNAs, 24 upregulated miRNAs, 5 brown module miRNAs, 4 magenta module miRNAs, and 7 yellow module miRNAs were selected for target prediction.

We used two approaches to select top targets as recommended by the developers of the TargetScan algorithm6,20,21. First, we identified all predicted miRNA target sites in a given mRNA 3′UTR and calculated the summed context+ score for all sites, as miRNA targeting at different sites has been shown to have non-cooperative effects in most cases60. We then selected the strongest targets that are hit by two or more miRNAs in each group (±2 for downregulated miRNAs or miRNAs in the brown, magenta, and yellow modules, ±4 for upregulated miRNAs due to the larger number of miRNAs in this group) and have a summed context+ score of ≤ −0.2 (the more negative the context+ score, the stronger the predicted targeting efficiency). Second, we selected the top 25% most conserved target sites (based on branch length scores) with context+ score ≤ −0.1. For individual miRNAs, we define “the strongest” targets as those with a summed context+ score of ≤ −0.1, and “the most conserved” targets as those with a branch length score in the top 25% and a context+ score of ≤ −0.05.

Gene set enrichment analysis. Gene set enrichment analyses were performed using two-sided Fisher’s exact tests with the fisher.test R function, except for enrichment for de novo variants (DNVs). For DNVs, a previous study showed a near linear relationship between gene length and de novo mutation frequency52,53. Therefore, for assessing enrichment of genes affected by DNVs in miRNA targets, we applied logistic regression, in which the probability of a gene being hit by a specific category of DNVs was coded as a function of gene coding region length covered in exome sequencing and whether the gene belongs to a certain miRNA target group. In addition, we also applied logistic regression to assess enrichment while controlling for gene 3′ UTR length (and also for gene coding region length for DNVs) in Supplementary Figure 6. P-values were FDR adjusted across 10 target groups for each gene list using Benjamini-Hochberg correction.

The background gene lists were defined as follows: (1) for the strongest targets of differentially expressed miRNAs and ASD-associated mRNA modules, the intersection between (a) protein-coding genes expressed in the cortex and (b) the targets of all 699 miRNAs that are hit by two or more miRNAs and have a summed context+ score of ≤ −0.2; (2) for the most conserved targets of differentially expressed miRNAs and ASD-associated mRNA modules, the intersection between (a) protein-coding genes expressed in the cortex and (b) the targets of all 699 miRNAs with a branch length ≥ the lowest branch length for the targets of the ASD-related miRNA groups and a summed context+ score of ≤ −0.1; (3) for the strongest targets of individual candidate miRNAs, the intersection between (a) protein-coding genes expressed in the cortex and (b) the targets of all 699 miRNAs with a branch length ≥ the lowest branch length for the targets of the ASD-related miRNA groups and a summed context+ score of ≤ −0.1; (4) for the most conserved targets of individual candidate miRNAs, the intersection between (a) protein-coding genes expressed in the cortex and (b) the targets of all 699 miRNAs with a branch length in the top 25% for each miRNA and a context+ score of ≤ −0.05; (5) for the strongest or the most conserved targets of individual candidate miRNAs that were validated in the hNPCs, the background used in (3) or (4) intersected with all genes expressed in the hNPCs as detected by our RNA-seq analysis.

Enrichment of common variants from genome-wide association studies. GWAS data for ASD, schizophrenia, and Alzheimer’s disease were obtained from the Autism Genome Resource Exchange/Children’s Hospital Philadelphia (AGRE/CHOP), the Psychiatry Genomics Consortium (PGC), and the International Genomics of Alzheimer’s Project (IGAP), respectively. Linkage disequilibrium (LD)-based SNP clamping was performed using PLINK (version 1.07) with the following parameters: --clump-p1 0.001 --clump-p2 0.05 --clump-r2 0.50 --clump-kb

--clump-p1 0.001 --clump-p2 0.05 --clump-r2 0.50 --clump-kb

250. LD information was obtained from AGRE or HapMap (release 23). Overlap between disease-associated SNP clumps and miRNA targets plus 20 kb flanking regions was then assessed using INRICH (ref. 32) with the following parameters: -w 20 -r 10000 -q 5000 -d 0.1 -p 0.1. INRICH takes a genomic permutation approach that accounts for linkage disequilibrium, SNP number and density, and gene density to calculate empirical P-values for each gene set31. It performs multiple testing correction via a second, nested round of permutation to assess the null distribution of the minimum empirical P value across all tested gene-sets32. We used the same background gene lists as in gene set enrichment analysis.

Gene ontology analysis. Gene ontology analysis was performed using GO-Elite (version 1.2.5), which uses a Z-score approximation of the hypergeometric distribution to assess term enrichment, with default settings and 5000 permutations62. False-discovery rate adjusted P values were calculated using Benjamini-Hochberg correction. We used the same background gene lists as in gene set enrichment analysis.

Quantitative RT-PCR. 200 ng total RNA treated with RNase-free DNase I (Qiagen) was reverse-transcribed using the miScript II RT Kit (Qiagen). Real-time PCR was performed using the miScript Primer Assays (Qiagen) and miScript SYBR Green PCR Kit (Qiagen) on a Roche LightCycler 480 instrument. Human RNU6B was used as internal control.

miRNA overexpression in hNPCs and RNA-seq. Primary human neural progenitor cells (hNPCs) were generated in a previous study and cultured as described63. The cells were free of mycoplasma contamination based on DAPI staining. At the fourth passage, hNPCs were seeded in 6-well plates at 250,000 cells per well. 24 h later, mimics of mature miRNAs (GE Healthcare) were transfected at a final concentration of 50 nM using the HiPerFect Transfection Reagent (Qiagen). The miRIDIAN microRNA Mimic Negative Control 1 (GE Healthcare), which is based on cel-mir-67 and has minimal sequence identity with human miRNAs, was used as negative control. Transfection for each miRNA mimic was performed in triplicate. 48 h after transfection, total RNA was extracted using the miRNeasy kit (Qiagen). RNA integrity number (RIN) was measured using an Agilent Bioanalyzer and all samples had a RIN > 9. Overexpression of miRNAs was confirmed with quantitative RT-PCR. 1.5 μg total RNA was then converted to miRNA libraries using the Illumina TruSeq Stranded miRNA Library Preparation Kit with poly-A selection. ERCC ExFold RNA Spike-In Mixes (Life Technologies) were added as internal controls. Libraries were then validated on an Agilent 2200 TapeStation system and quantified with the Quant-IT PicoGreen assay (Life Technologies). 12 libraries barcoded with Illumina TruSeq indexes were pooled into one lane and sequenced 3 times on an Illumina HiSeq2500 instrument using the rapid run mode with 69-bp paired-end reads. After demultiplexing with CASAVA v1.8.2 (Illumina), reads were mapped to the GRCh37.75 reference genome using TopHat2 (ref. 64). Sequencing quality was then examined using Picard Tools version 1.128 (commands CollectMultipleMetrics, CollectRnaSeqMetrics, and CollectGcBiasMetrics) and the flagstat command in SAMtools (version 1.2)65. 41.0–77.9 million QC-passed reads were mapped to the reference genome, with 84.8–87.6% mapped to miRNAs, for each sample. Gene expression levels were then quantified using HTSeq (version 0.6.1 beta)66 with a union exon model. Genes with 10 or more counts in at least 2 samples in any miRNA overexpression or negative control group were kept for further analysis. Gene expression levels in samples within the same group were highly correlated (R² ≥ 0.99). The expression data were then normalized using the DESeq R package for library size53 and/or the CQN R package for GC content44. Differential gene expression analysis was performed using one-sided t test assuming unequal variance. Uncorrected P values were used to define differentially expressed genes, as the sample size was small (n = 3 for each group).

Protein-protein interaction (PPI) analysis. PPI analysis was performed using DAPPLE (ref. 67) v2.0 which uses the InWeb database68 and applies a within-degree within node permutation methods. 10,000 permutations were used.

Summary of statistical methods. Blindness. Tissue dissection, RNA extraction, library preparation, and sequencing were performed blind to all metadata
information about the samples. Data analysis was not performed blind to metadata information about the samples.

Randomization. Tissue dissection, library preparation, and sequencing were randomized for balance of diagnostic status, age, sex, and brain region.

Sample sizes. No statistical methods were used to pre-determine sample sizes as effect sizes were not known a priori, but our sample sizes are larger than those reported in previous publications that detected miRNA changes.41,43,44

Parametric tests. For DGE analyses using linear mixed-effects and linear models (Fig. 1b,c, Supplementary Fig. 3a–e and Supplementary Table 2), normality was not formally tested for each miRNA. For the main DGE analysis of 95 cortex samples, we also computed permutation-based P values and found that the P value rankings of miRNAs were highly concordant with those observed in the original sample set (Pearson's R = 0.99, P < 2.2 × 10−16). For calculation of Pearson correlations (Figs. 1c, 2b–d and 4a–d and Supplementary Figs. 2a,d, 3a–e and 9a–c), normality was not formally tested. One-sided t-tests were used for Figures 6a, 7a and 8a–c, and Supplementary Figure 5 because the distributions are approximately normal and sample sizes were reasonably large (n = 88-11695). For Supplementary Figure 3f,g, normality was tested by the Shapiro–Wilk test. All groups except for the CTL group for hsa-miR-10a-5p (P = 0.005) and the ASD group for hsa_can_1155-m (P = 0.03) meet the normality assumption (P > 0.05). Two-sided t-tests were used for all groups. For differential gene expression analysis after miRNA overexpression in hNPCs (Supplementary Table 5), one-sided t-tests were used but normality was not formally tested for each gene. For all t-tests, equal variances were not formally tested, and so all tests were performed assuming unequal variance. For DGE analyses using linear mixed-effects and linear models, equal variances were not formally tested for each miRNA.

Non-parametric tests. One-tailed Wilcoxon rank sum tests were performed in Supplementary Figure 8a–h because the distributions do not appear to be normal.

A Supplementary Methods Checklist is available.

Code availability. The R code for the DGE analysis using a linear mixed-effects model and the WGCNA analysis is provided in Supplementary Code 1 and 2.

Data availability. Brain sample metadata, miRNA raw read counts, miRNA DGE analysis data, miRNA WGCNA data, and miRNA DGE analysis data following miRNA overexpression in hNPCs are provided in Supplementary Tables 1, 2 and 5. Raw small RNA-seq and RNA-seq data from brain samples from ASD cases and controls have been deposited to the PsychENCODE Knowledge Portal (http://dx.doi.org/10.7303/syn4587609). Raw RNA-seq data from hNPCs following miRNA overexpression are available from the corresponding author upon request.

49. Quinn, E.M. et al. Development of strategies for SNP detection in RNA-seq data: application to lymphoblastoid cell lines and evaluation using 1000 Genomes data. PLoS One 8, e58815 (2013).
50. Friedländer, M.R. et al. Discovering microRNAs from deep sequencing data using miRDeep. Nat. Biotechnol. 26, 407–415 (2008).
51. Hackenberg, M., Sturm, M., Langenberger, D., Falcón-Pérez, J.M. & Aransay, A.M. miRanalyzer: a microRNA detection and analysis tool for next-generation sequencing experiments. Nucleic Acids Res. 37, W68–W76 (2009).
52. Hackenberg, M., Rodríguez-Espeleta, N. & Aransay, A.M. miRanalyzer: an update on the detection and analysis of microRNAs in high-throughput sequencing experiments. Nucleic Acids Res. 39, W132–W138 (2011).
53. Anders, S. & Huber, W. Differential expression analysis for sequence count data. Genome Biol. 11, R106 (2010).
54. Hansen, K.D., Irizarry, R.A. & Wu, Z. Removing technical variability in RNA-seq data using conditional quantile normalization. Biostatistics 13, 204–216 (2012).
55. Leek, J.T., Johnson, W.E., Parker, H.S., Jaffe, A.E. & Storey, J.D. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics 28, 882–883 (2012).
56. Miller, J.A., Horvath, S. & Geschwind, D.H. Divergence of human and mouse brain transcriptome highlights Alzheimer disease pathways. Proc. Natl. Acad. Sci. USA 107, 12698–12703 (2010).
57. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 9, 559 (2008).
58. Langfelder, P., Zhang, B. & Horvath, S. Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. Bioinformatics 24, 719–720 (2008).
59. Csardi, G. & Nepusz, T. The igraph software package for complex network research. InterJournal Complex Systems 1695 (2006).
60. Quinlan, A.R. & Hall, I.M. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842 (2010).
61. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81, 559–575 (2007).
62. Zambon, A.C., et al. GO-Elite: a flexible solution for pathway and ontology over-representation. Bioinformatics 28, 2209–2210 (2012).
63. Stein, J.L. et al. A quantitative framework to evaluate modeling of cortical development by neural stem cells. Neuron 83, 69–86 (2014).
64. Kim, D. et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 14, R36 (2013).
65. Li, H. et al. The sequence alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079 (2009).
66. Anders, S., Pyl, P.T. & Huber, W. HTSeq–a Python framework to work with high-throughput sequencing data. Bioinformatics 31, 166–169 (2015).
67. Rossin, E.J. et al. Proteins encoded in genomic regions associated with immune-mediated disease physically interact and suggest underlying biology. PLoS Genet. 7, e1001273 (2011).
68. Lage, K. et al. A human phenotype-interactome network of protein complexes implicated in genetic disorders. Nat. Biotechnol. 25, 309–316 (2007).