

Trans-genetic effects of circular RNA expression quantitative trait loci and potential causal mechanisms in autism

Te-Lun Mai

Genomics Research Center, Academia Sinica <https://orcid.org/0000-0002-9356-6198>

Chia-Ying Chen

Genomics Research Center, Academia Sinica, Taipei

Tai-Wei Chiang

Academia Sinica

Trees-Juen Chuang (✉ trees@gate.sinica.edu.tw)

Genomics Research Center, Academia Sinica, Taipei

Article

Keywords: Autism spectrum disorder, Circular RNA, circQTL, trans-eQTL, Regulatory network, Causative biology

Posted Date: February 5th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-155043/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Genetic risk variants and transcriptional expression changes in autism spectrum disorder (ASD) have been widely identified, but their causal relationship is largely unknown. Circular RNAs (circRNAs) are abundant in brain and often serve as upstream regulators of mRNAs. By integrating RNA-sequencing with genotyping data from autistic brains, we accessed expression quantitative trait loci (eQTL) of circRNAs (circQTLs) that influenced expression of distant genes (*trans*-eGenes) and constructed 43,372 circQTL-*trans*-eGene pairs. Mediation test suggested that 19,393 pairs were significantly *cis*-mediated by expression of circRNAs near the circQTLs; meanwhile, causal inference test (CIT) suggested 1,000 pairs influencing 708 *trans*-eGenes, wherein *trans*-eGene expression mediated *trans*-eQTL effects on ASD diagnosis. The 708 *trans*-eGenes were enriched for ASD risk genes and overrepresented in neurons in the upper neocortical layer. Integration of mediation test- and CIT-passing pairs further constructed 257 circQTL-circRNA-*trans*-eGene-ASD propagation paths. These findings increase our understanding of causative biology and cryptic regulatory mechanisms underlying ASD.

Introduction

Autism spectrum disorder (ASD) is a highly pervasive neurodevelopmental and heritable complex disorder, which are characterized by limited social communication, restricted and repetitive interests or behaviors¹. Previous large-scale genomic studies have identified a variety of genomic variants associated with the etiology of ASD^{2,3}, providing valuable biological insights into this disorder. However, the contribution of genetic factors to this complex disease is highly heterogeneous and not yet well understood. On the other hand, through high-throughput RNA sequencing (RNA-seq) data of postmortem brains from people with idiopathic ASD and non-ASD controls, numerous studies identified valuable differential expression patterns in mRNAs^{4,5}, long non-coding RNAs⁵, microRNAs (miRNAs)^{6,7}, RNA editing⁸, and circular RNAs (circRNAs)^{9,10}. While these results increased our understanding of the underlying molecular mechanisms in ASD pathophysiology, the causal relationships between genetic sequence variants and transcriptional expression changes in ASD remains mostly unknown.

By integrating genome-wide genotyping with transcriptome profiling, expression quantitative trait loci (eQTL) analysis identifies single nucleotide polymorphisms (SNPs) that affect the expression levels of local (*cis*-eQTL) or distant (*trans*-eQTL) genes, providing a connection between genetics and regulatory mechanisms of gene expression¹¹. *cis*-eQTL SNPs often reside within or close to promoter regions of their nearby genes (*cis*-eGenes) and mediate gene expression through directly affecting the corresponding transcription factor binding process¹². In contrast to frequent investigation of *cis*-eQTLs, the regulatory mechanism underlying *trans*-eQTLs are less known because of the relatively weaker effects on gene expression for *trans*-actions than for *cis*-actions and the multiple-testing burden^{13,14}. However, compared with *cis*-effects on gene expression, *trans*-genetic effects were reported to explain more than two times the variability in gene expression¹⁴ and be more cell type-specific¹⁵. Some disease-causing variants involved in *trans*-effect analyses may miss in *cis*-effect analyses¹⁶. A recent study

further indicated that at least 70% of complex trait heritability is driven by *trans*-eQTL effects¹⁷. A possible regulatory mechanism for *trans*-eQTLs is that *trans*-eQTLs indirectly affect the expression of distant genes (*trans*-eGenes) through regulating the expression of genes near the *trans*-eQTLs (*cis*-eGenes)^{13,18,19}. In such cases, a *trans*-eQTL is also a *cis*-eQTLs, wherein the corresponding *cis*-eGene acts as a mediator of the *trans*-eQTL. Investigation of *trans*-genetic effects and the causal inference for the corresponding eQTL-associated networks will allow us to better understand complex trait genetics.

CircRNAs are an emerging class of RNAs, which are formed by pre-mRNA back-splicing with a structure of covalently closed loops and endogenously expressed as single-strand, non-polyadenylated circular molecules^{20,21}. Although circRNAs are generally expressed at a much lower level compared with their corresponding co-linear mRNA isoforms, they are more stable than other types of RNAs^{20,22}. CircRNAs were observed to be especially abundant in the brain^{23,24} and often play important roles in development of nervous system^{24,25}. As for the layer of gene regulatory networks, circRNAs often serve as an upstream regulator of mRNAs. For example, they may regulate genes in *trans* through mediating the activities of miRNAs or RNA binding proteins (RBPs), with the common miRNA/RBP target site(s) of circRNAs and the target genes^{20,21}. Hence, we are curious about whether SNPs that influence expression of nearby circRNAs (referred to as circQTLs) may also affect expression of remote genes, whereby circRNAs serve as an intermediate regulator bridging circQTLs and *trans*-eGenes. If this is possible, we can generate circQTL-circRNA-*trans*-eGene propagation paths in autistic brains and shed light on the functional consequences of the genetic loci and the etiology of this complex disease.

There were two challenges to achieve the abovementioned goals. First, eQTL analysis requires transcriptomic and genomic data from the same samples with a large sample size. Second, large-scale circRNA study is often hampered by the limitation of the intrinsic circRNA characteristic of lacking polyadenylated tails. The newly released Synapse database^{5,26} comprises a large human brain sample size of both genotyping data and RNA-seq data (rRNA-depleted RNAs from total RNAs without poly(A)-selection) from ASD cases and controls, offering an unprecedented opportunity for us to decipher the effects of genetic variation on circRNA expression in ASD brain. The data from the Synapse database thus allowed us to identify a new reference of circQTLs/*trans*-eQTLs and ASD-associated *trans*-eQTL hotspots in the brains and infer *cis*-mediators of the identified circQTLs (i.e., such circQTLs are also *trans*-eQTLs) and causal effects between circQTLs and ASD diagnosis. Our findings help to unveil the gap between genetic variation and phenotypic changes in ASD and provides a framework for future analyses of eQTLs in neuropsychiatric disease or other complex traits.

Results

Identification of circQTLs in ASD brain

We first collected rRNA-depleted paired-end RNA-seq data⁵ and genotyping data²⁶ of postmortem samples from frontal cortex (FC) (Brodmann area 9), temporal cortex (TC) (Brodmann area 22, 41, and

42), and cerebellar vermis (CV) from individuals with ASD and non-ASD controls (the Synapse database) (Table 1). To identify local genetic effects that regulate circRNA expression, we considered the RNA-seq data and the corresponding genotyping data, both of which were derived from the same individuals (Fig. 1a, Table 1, and Supplementary Table 1). The examined circRNAs (1,060 circRNAs; Supplementary Table S1) were extracted from our previous study¹⁰ based on the RNA-seq data used in this study. The expression levels of circRNAs were calculated using the number of supporting circRNA junction reads per million uniquely mapped reads (RPM)²⁷ and adjusted for covariates such as the corresponding host gene expression, brain region, diagnosis, and other biological/technical confounding factors (see Methods and Supplementary Table 1). For accuracy, the genotyped and imputed SNPs should be satisfied imputation score ³ 0.8 and estimated minor allele frequency ³ 0.01. To focus on *cis*-effects of circQTLs, we limited our analysis to SNPs in \pm 200 kb nucleotides of each back-splice site. CircQTLs were evaluated by testing the correlations between the imputed genotype dosages and covariate-adjusted circRNA expression using Matrix eQTL²⁸. *P* values were adjusted across all circQTLs using false positive rate (FDR) correction (Methods). Significant circQTLs were determined if FDR values were less than 0.05. We thus identified 989 circQTLs associated with 81 circRNAs (Supplementary Table 2).

Recent transcriptomics studies of ASD showed that expression profiles of transcripts (including circRNAs¹⁰, miRNAs⁶, and mRNAs^{29,30}) and allele-specific genes²⁶ were highly similar between the two cortical regions (FC and TC) but were quite distinct in the CV. We calculated the circQTL effects for the 989 circQTLs in the three brain regions, respectively. We found that the circQTL effect size and direction were highly concordant between the two cortical regions and relatively weakly concordant (or even not concordant) between the cortex and CV (Fig. 1b). The differences in circQTL effect size between the cortex and CV were more significant than that between the two cortical regions (Fig. 1c). We also performed principal component analysis (PCA) and observed the similar results for the expression profiles of the 81 circRNAs (Fig. 1d). Previous studies have reported that the ASD pathophysiology was present in the cortex^{31,32} and transcriptional dysregulation in ASD were shown to be consistently stronger in the cortex than in the cerebellum⁵. These observations suggest that the cortical regions are more selectively vulnerable to transcriptomic changes than the cerebellum. Therefore, we focused on the RNA-seq data and the corresponding genotyping data from the cortex (FC/TC) samples (105 samples from 53 ASD cases and 52 controls; Table 1) for the subsequent circQTL analyses. For accuracy, we only considered the identified circQTL SNPs (666 circQTLs associated with 48 circRNAs; Supplementary Table 2) in which the number of individuals with the minor allele (heterozygotes and minor allele homozygotes) was larger than three.

We then studied the relationship between circQTLs and circRNA formation. We first showed that the identified circQTLs were preferentially located in the flanking sequences close to the back-splice donor/acceptor sites (Fig. 1e). The majority of circQTLs (90.5%; 603 out of 666; Supplementary Table 2) were located in the flanking sequences of back-splice sites. These observations were consistent with previous studies for distance distribution analysis of circQTLs in other human samples^{33,34}. Next, back-splicing was reported to be facilitated by reverse complementary sequences (RCSs) residing in the introns

flanking circularized exons^{27,35,36}. We found that circQTLs residing in the flanking sequences of back-splice sites were significantly enriched in RCSs compared with background (empirical $P < 10^{-4}$; see Methods) and non-circQTL SNPs ($P = 3.2 \times 10^{-7}$ by two-sided Fisher's exact test; Fig. 1f). Of note, the examined non-circQTL SNPs were also located in the flanking sequences of the examined back-splice sites and not determined as circQTLs. Moreover, it was shown that RNA binding proteins (RBPs) can regulate back-splicing by binding to the flanking sequences^{20,21}. Indeed, the circQTLs located in the flanking sequences were significantly enriched in RBP binding sites compared with background (empirical $P < 10^{-4}$) and non-circQTL SNPs ($P = 0.01$; Fig. 1g). These observations support that circQTLs located in the flanking sequences of back-splice sites may contribute to circRNA formation.

Mediation effects of *trans*-eQTLs (circQTLs) via circRNA expression

We next examined whether the identified circQTLs were also *trans*-eQTLs affecting the expression of distant genes (*trans*-eGenes) using Matrix eQTL. By definition, such circQTL SNPs and *trans*-eGenes should be located on different chromosomes or the same chromosome separated by a distance greater than 5Mb (from the SNP site to the transcription start site of the *trans*-eGene). Since some SNPs in high linkage disequilibrium (LD) were associated with different *trans*-eGenes, non-pruned circQTLs were considered in this study. With controlling for potential biological/technical confounding factors and multiple testing correction (Methods), 546 circQTLs (associated with 47 circRNAs) were identified to be also *trans*-eQTLs, which affected the expression of 7,165 *trans*-eGenes and constructed 43,372 circQTL-*trans*-eGene pairs (Fig. 2a; Supplementary Table 3). Meanwhile, these circQTL-*trans*-eGene pairs also resulted in 13,379 circRNA-*trans*-eGene pairs, where the expression levels of circRNAs and *trans*-eGenes were both regulated through the same circQTLs (Fig. 2a). We then explored the causes for the *trans*-effect of circQTLs on the expression of *trans*-eGenes. In addition to the direct effect of circQTLs on *trans*-eGene expression, we hypothesized that in some cases the expression of *trans*-eGenes were indirectly regulated through the expression of circRNAs near the circQTLs (i.e., the "mediation effect"; Fig. 2a). For the scenario of mediation effects, the circRNAs served as *cis*-mediators of *trans*-eQTLs. We first showed that the expression levels of circRNAs and *trans*-eGenes in the 13,379 circRNA-*trans*-eGene pairs were more likely to be correlated with each other than expected (empirical $P < 10^{-4}$; see Supplementary Fig. 1 and Methods). To test the mediation effects of *trans*-eQTLs, we performed mediation analyses³⁷ for all 43,372 circQTL-*trans*-eGene pairs to identify the proportion of association between a circQTL and the *trans*-eGene that was caused by the effect of the circQTL on the corresponding circRNA expression. We found that 19,393 out of the 43,372 (45%) circQTL-*trans*-eGene pairs passed the mediation test (Supplementary Table 3), which were significantly mediated (P and FDR < 0.05 with adjusting for related covariates; see Methods) by expression of circRNAs near the circQTLs. The majority (67%, 13,023 pairs) of the 19,393 pairs exhibited that more than 10% of the circQTL-*trans*-eGene association were mediated by the expression of circRNAs; in some cases (2,099 pairs), the proportion of mediation of the circQTL-*trans*-eGene association by the circRNAs was even more than 50% (Fig. 2b).

Regarding the 19,393 circQTL-*trans*-eGene pairs (or 19,393 circQTL-circRNA-*trans*-eGene axes), we observed that the percentages of circQTL-*trans*-eGene pairs passing the mediation test markedly increased with increasing significance levels of *trans*-eQTL effects for circQTL-*trans*-eGene pairs (Fig. 2c). Such a percentage even reached more than 70% if FDR values of *trans*-eQTL effects were smaller than 10^{-8} . We further examined the relationship between mediation effects and the significance levels of circRNA-*trans*-eGene correlation of expression profile. We found 55% of the circRNA-*trans*-eGene pairs passing the mediation test if the circRNA expression and *trans*-eGene expression were significantly (Spearman's $P < 0.05$) correlated with each other (Fig. 2d), whereas such a percentage was significantly reduced from 55% to 40% (odds ratio=0.56 and $P < 2.2 \times 10^{-16}$ by Fisher's exact test) if the significant correlation between circRNA and *trans*-eGene expression disappeared (Fig. 2d). The percentages of the circQTL-*trans*-eGene pairs passing the mediation test markedly increased with increasing significance levels of the correlation of expression profile between circRNAs and *trans*-eGenes (Fig. 2d). Such a percentage even reached 90% at Spearman's P value < 0.0005 . Similar results were observed in the analyses based on the CV samples (Supplementary Fig. 2 and Supplementary Table 3). These results reveal that the mediation effects of the circQTLs (*trans*-eQTLs) on circRNA expression are correlated with the magnitudes of *trans*-eQTL effect and circRNA-*trans*-eGene correlation of expression profile.

Causal effects between circQTLs and ASD diagnosis

Regarding the identified 43,372 circQTL-*trans*-eGene pairs, we then conducted causal inference test^{38,39} (CIT; see Methods) to examine whether the *trans*-effects of circQTLs on *trans*-eGene expression might explain the association between the circQTLs and diagnosis status (ASD vs. non-ASD) and infer the direction of association between circQTLs, *trans*-eGene expression, and ASD diagnosis (Fig. 3a, top). At P and FDR < 0.05 , we identified 1,000 circQTL-*trans*-eGene pairs with a propagation path from circQTL SNP to ASD diagnosis via *trans*-eGene expression (Fig. 3a, bottom; Supplementary Table 4). We speculated that the *trans*-eGenes (708 genes; CIT-passing *trans*-eGenes) involved in the 1,000 circQTL-*trans*-eGene pairs may be implicated in ASD. We examined enrichment analysis for genes previously implicated in ASD from Simons Foundation Autism Research Institute (SFARI)⁴⁰. We observed that the CIT-passing *trans*-eGenes were significantly enriched for the SFARI genes, but not for genes implicated in monogenetic forms of other brain disorders (Fig. 3b). These CIT-passing genes were also enriched for other classes of ASD-relevant genes, including genes encoding postsynaptic density (PSD) proteins⁴¹, genes whose transcripts were bound by the RBPs of FMR1⁴², RBFox1⁴³, and ELAVL1⁴⁴, and differentially expressed genes (DEGs) in ASD⁵ that were derived from the same cortex samples used in this study (Fig. 3b). In contrast, these trends were not observed in the *trans*-eGenes not passing CIT (Fig. 3b). Moreover, a previous study has presented a genome-wide prediction of ASD risk genes and provided an estimated probability of ASD-association for each gene⁴⁵. On the basis of the probabilities of ASD risk, we observed the CIT-passing genes indeed had a significantly higher probability of ASD risk compared with background gene set (empirical $P < 0.001$; Supplementary Fig. 3) and the *trans*-eGenes not passing CIT (all $P < 10^{-8}$ by two-sided Fisher's exact test; Fig. 3c). We further found that the CIT-passing *trans*-eGenes were more intolerant of a loss function mutation (measured by gene variant intolerance

(pLI) scores⁴⁶) compared with the background gene set (Supplementary Fig. 3) and the *trans*-eGenes not passing CIT (Fig. 3d), reflecting previous observations that genes implicated in ASD tended to be subject to stronger selective constraints than the other genes^{45,47}. Regarding DEGs in a cell type-specific manner, the CIT-passing genes were most overrepresented in vasoactive intestinal polypeptide (VIP)-expressing interneurons and L2/3 and L4 excitatory neurons. This trend was generally consistent with the previous observation for SFARI genes⁴⁸. Gene Ontology (GO) analysis revealed that the CIT-passing genes were enriched in GO terms related to chemical synaptic transmission, synaptic signaling, neurogenesis, synapse, neuron projection, axon, and so on (Fig. 3f and Supplementary Table 4), also reflecting their enrichment for PSD. These observations support the relevance of the CIT-passing *trans*-eGenes to ASD pathophysiology.

Furthermore, the above mediation analysis has identified 19,393 circQTL-circRNA-*trans*-eGene axes with significant mediation effects of circQTLs on *trans*-eGene expression through circRNA expression (Fig. 2b and Fig. 4a, middle). By integrating the CIT-passing circQTL-*trans*-eGene pairs with the 19,393 circQTL-circRNA-*trans*-eGene axes, we determined 257 circQTL-circRNA-*trans*-eGene-ASD diagnosis propagation paths, where the circRNA-*trans*-eGene axes may act as causal mediators for the circQTL-ASD diagnosis associations (Fig. 4a, left; Supplementary Table 4). The 257 ASD-associated circQTL-circRNA-*trans*-eGene regulatory axes involved 33 circQTLs (*trans*-eQTLs), 8 circRNAs, and 124 *trans*-eGenes (Figs. 4a and 4b). We found that 97 of 124 (78%) *trans*-eGenes were ASD-relevant genes and 217 of 257 (84%) circQTL-circRNA-*trans*-eGene axes were associated with the 97 ASD-relevant genes (Fig. 4b). Particularly, 10 of the 97 genes were SFARI genes and five were top SFARI genes with score = 1 or 2, suggesting the regulatory role of the corresponding circQTL-circRNA-*trans*-eGene axes in ASD brain.

Potential mediators for the circRNA-*trans*-eGene interactions

Regarding the 19,393 circQTL-circRNA-*trans*-eGene axes (associated with 5,669 circRNA-*trans*-eGene axes) passing the mediation test, we proceeded to identify potential mediators for the circRNA-*trans*-eGene interactions (Fig. 5a, left). Since circRNAs and mRNAs may carry the common miRNA/RBP target sites, numerous cases of circRNAs were demonstrated to act as an upstream regulator of mRNAs through mediating miRNA/RBP activities^{20,21}. To this end, we utilized crosslinking immunoprecipitation (CLIP)-seq data-supported RNA interactomes (extracted from ENCORI⁴⁹) to search for the common miRNA/RBP target sites of the 5,669 circRNA-*trans*-eGene axes (see Methods). We thus identified 3,141 circRNA-miRNA-*trans*-eGene axes (associated with 958 circRNA-*trans*-eGene axes; Fig. 5a, middle) and 54 circRNA-RBP-*trans*-eGene axes (associated with 30 circRNA-*trans*-eGene axes; Fig. 5a, right), respectively (Supplementary Table 5). Regarding the 257 abovementioned circQTL-circRNA-*trans*-eGene axes that passed both the mediation test and CIT (see Fig. 4a, right and Fig. 5b, left), we further found that 90 axes (Fig. 5b, right) overlapped with the 958 circRNA-*trans*-eGene axes with common miRNA target sites of the circRNAs and *trans*-eGenes and no axis overlapped with the 30 circRNA-*trans*-eGene axes with common RBP target sites of the circRNAs and *trans*-eGenes. Accordingly, 158 circQTL-circRNA-miRNA-*trans*-eQTL axes that might be causally related to ASD diagnosis were generated (Figs. 5b and 5c; Supplementary

Table 5). The 158 axes were associated with 28 circQTLs, 6 circRNA, 17 miRNAs, and 34 *trans*-eGenes (Fig. 5b, right). Of note, 28 of the 34 *trans*-eGenes were ASD-relevant genes and one (*ANK3*) was a top SFARI gene with score = 1 (Fig. 5c). A previous study reported that genetic variations in *ANK3* were associated with autism susceptibility⁵⁰. *ANK3* can regulate the structure and function of glutamatergic synapses⁵¹. Alteration of *ANK3* expression can regulate neuronal microtubule dynamics functions and affect the corresponding role in psychiatric illness⁵². Taken together, we provided a framework for detecting potentially causal relationships between genetics and trait and generating the corresponding regulatory axes in ASD.

Discussion

The majority of eQTL studies to date have focused on investigation of *cis*-genetic effects on gene expression, although *trans*-eQTL effects may contribute most trait heritability¹⁷. One challenge for genome-wide scanning for *trans*-eQTLs is the heavy burden of multiple testing^{13,14}. This study focused on detecting the circQTLs that *cis*-regulated expression of nearby circRNAs and also *trans*-regulated expression of distant genes simultaneously. We first identified circQTLs and then searched for *trans*-eGenes based on these circQTL SNPs (not all SNPs extracted), greatly decreasing the burden of multiple *trans*-eQTL testing for the associations between SNPs and gene expression. We thereby identified 546 circQTLs that were also *trans*-eQTLs and generated 43,372 circQTL-*trans*-eGene associations (Fig. 2a).

Since circRNAs often serve as an upstream regulator of mRNAs^{20,21}, we asked whether in some cases the effects of *trans*-eQTLs on *trans*-eGene expression were mediated by the circRNA expression. The mediation analysis suggested that 45% (19,393) of the 43,372 circQTL-*trans*-eGene associations were mediated by the expression levels of the corresponding circRNAs. Some mediation test-passing pairs exhibited that more than 50% of the circQTL-*trans*-eGene association were mediated by the expression of circRNAs, supporting the importance of the mediation effects for *trans*-eQTLs (Fig. 2b). In addition, we observed that circQTLs residing in the flanking sequences of back-splice sites were enriched in RCSs and RBP binding sites (Figs. 1f and 1g), both of which are important for circularization. These observations imply a possible regulatory mechanism for the circQTLs (*trans*-eQTLs) passing the mediation test, wherein circQTLs affect the formation of their nearby circRNAs and the circRNAs, as an upstream regulator of the corresponding *trans*-eGenes, consequently regulate the *trans*-eGene expression. This is worth further investigation.

To increase our understanding of the association between genetic variants and complex traits/diseases such as ASD, previous genome-wide association studies (GWASs) have provided valuable insights into the role of genetic etiology of ASD^{2,53}. Recent transcriptome-wide association studies (TWASs) further integrated GWAS loci with gene expression data to investigate *cis*-effects on gene expression associated with risk loci for ASD^{54,55,56}. However, TWASs do not clarify the major cause of the association between eQTLs and the expression of the corresponding eGenes because the direction of association between eGene expression and trait is unclear. This leads to three possible models (i.e., eGene mediation model,

reverse causality model, and independent model) to explain the relationships between an eQTL, an eGene, and a trait^{19,57,58}. For the first model, the eGene acts as a causal mediator for the eQTL-trait associations, wherein the eQTL regulates the trait through mediating the eGene expression. For the second model the trait acts as a causal mediator, wherein alteration of the eGene expression is the consequence of the trait. For the third model, the effects of eQTL on eGene expression and trait are independent. This study focuses on identification of eGenes that are causally related to idiopathic ASD (i.e., the first model). To this end, we utilized CIT^{38,39} to examine the 43,372 circQTL-*trans*-eGene pairs and identified 1,000 pairs with *trans*-eGenes as a potentially causal mediator for ASD diagnosis (ASD vs. non-ASD) (Fig. 3a). By definition of CIT, a CIT-passing pair should simultaneously satisfy four conditions (see Methods), especially for the two conditions: circQTLs should be independent of ASD diagnosis after adjusting for *trans*-eGene expression and *trans*-eGene expression should be associated with ASD diagnosis after adjusting for circQTLs. Therefore, upon passing the CIT screening, both reverse causality and independent models are less probable for the explanation of the CIT-passing circQTL-*trans*-eGene associations.

We proceeded to examine the *trans*-eGenes (708 genes) involved in the CIT-passing circQTL-*trans*-eGene pairs and found significant enrichment for ASD-relevant genes such as SFARI genes, genes encoding PSD proteins, targets of FMR1, RBFOX1, and ELAVL1, and DEGs in ASD, but not for genes implicated in monogenetic forms of other brain disorders (Fig. 3b). Like SFARI genes, the CIT-passing *trans*-eGenes had high probabilities of ASD risk (Fig. 3c), high pLI scores (Fig. 3d) and similar patterns of DEG distribution among cell types (Fig. 3e). Particularly, we showed that the CIT-passing *trans*-eGenes exhibited significant enrichment for DEGs in neurons in the upper neocortical layer such as IN-VIP, L4, and L2/3 (Fig. 3e). A previous mouse study had demonstrated that excess upper-layer pyramidal neurons can result in an excitatory-inhibitory neuron imbalance, leading to behavioral abnormality similar to the features in ASD⁵⁹. L2/3 neurons are known to be important for underpinning high-level cognitive functions and implicated in cognitive syndromes of high-level associative dysfunction such as ASD⁶⁰. These observations support the important role of these CIT-passing *trans*-eGenes in ASD susceptibility.

Of the 43,372 circQTL-*trans*-eGene pairs (associated with 546 circQTLs), although the circQTL SNPs did not reach genome-wide significance in the largest ASD GWAS² currently, 217 of 546 (40%) circQTLs were associated with the expression of at least one ASD risk genes (SFARI genes; see Supplementary Fig. 4 and Supplementary Table 6). The 217 circQTLs resided at 40 distinct LD blocks, 133 of which (residing at 30 LD blocks) affected the expression of even more than 50 *trans*-eGenes. The 30 LD blocks can be regarded as *trans*-eQTL hotspots associated with SFARI genes. Of the 133 circQTLs, 25 circQTLs were particularly connected with ASD risk genes, in which the affected *trans*-eGenes were significantly enriched for SFARI genes (all $P < 0.05$ by hypergeometric test; Supplementary Table 6). This result provides a useful resource for further investigating regulatory mechanisms of the corresponding ASD risk genes. Intriguingly, we found that five *trans*-eGenes (*CRHR1*, *XKR6*, *NKX2-2*, *NSF*, and *ARL17B*) were previously prioritized as TWAS-significant genes/transcripts in ASD in two recent studies^{55,56} (Table 2). This revealed that these TWAS-significant genes/transcripts might be regulated by not only *cis*-acting SNPs

but also *trans*-acting ones. Of note, *NSF* was also a *trans*-eGene involved in the CIT-passing rs182920567-*NSF*-ASD diagnosis propagation path, supporting that *NSF* may be causally related to ASD diagnosis. This reflects a previous report that *NSF* may play a role in ASD pathophysiology through mediating serotonin transporter trafficking⁶¹. In addition, some circQTL-*trans*-eGene associations passed the mediation test, wherein the effects on *trans*-eGene expression (*CRHR1*, *XKR6*, and *ARL17B*) were mediated by the circRNA expression (circERC1 and circSPHKAP). Two phenomenon were observed. First, the expression of a TWAS-significant gene may be regulated by multiple circRNAs. We found circERC1-1 and circSPHKAP with more than 50% (or even more than 70%) of mediation effects on the relations of *trans*-eQTLs to *CRHR1*, suggesting multiple upstream regulators of *CRHR1*. Second, two TWAS-significant genes/transcripts (*XKR6* and *ARL17B*) may be under the same regulatory control (rs572758241-circSPHKAP association). Of note, rs572758241 was located at a *trans*-eQTL hotspot and affected the expression of 724 *trans*-eGenes. The 724 genes significantly overlap with SFARI genes ($P = 0.01$ by hypergeometric test; Supplementary Table 6), indicating the relevance of the *trans*-eQTL to ASD. These results thus provided potential upstream regulators of these TWAS-significant genes/transcripts.

By integrating the circQTL-*trans*-eGene pairs that passed CIT with the circQTL-circRNA-*trans*-eGene axes that passed mediation test, we further detected 257 potential circQTL-circRNA-*trans*-eGene-ASD diagnosis propagation paths (Fig. 4a). Meanwhile, we searched for the common miRNA target sites of the corresponding circRNAs and *trans*-eGenes and generated potential regulatory associations between the circRNAs and the *trans*-eGenes. Combined the predicted circRNA-miRNA-*trans*-eGene axes with the 257 potential circQTL-circRNA-*trans*-eGene-trait propagation paths, we finally identified 158 circQTL-circRNA-miRNA-*trans*-eQTL-trait propagation paths (Fig. 5b). Of the 158 axes, it is possible to further utilize the mediation analysis to examine whether the circRNA-*trans*-eGene associations were mediated by the expression of miRNAs, if both mRNA and miRNA expressing data derived from the same samples are available. For example, a circQTL-circRNA-miRNA-*trans*-eGene-trait propagation path that passes CIT and two layers of mediation test was detected according to miRNA expressing data⁶ derived from the same ASD and control samples used in this study (Supplementary Fig. 5).

On the basis of a large brain sample size of both RNA-seq data and the corresponding genotyping data from ASD cases and non-ASD controls, this study is the first report, to the best of our knowledge, for systematically investigating *trans*-genetic effects of circQTLs and inferring the corresponding causal relations in diseases. The identified ASD-associated regulatory interactions, particularly the CIT-passing axes involving *trans*-eGenes that were previously implicated in idiopathic ASD (e.g., SFARI genes), may serve as a useful resource for further investigations of more complex regulatory mechanisms underlying ASD. Future studies with larger patient cohorts, including genome-wide transcriptomic and genomic sequencing data, will allow for more precise causal relationships between genetic variants and transcriptional expression changes in ASD. The provided framework can be applied to future analysis of eQTL-trait relations in other complex diseases.

Methods

Identification of circQTLs/*trans*-eQTLs. The RNA-seq data of ASD and non-ASD brain samples⁵ and the corresponding genotyping data²⁶ were obtained from Synapse (<http://www.synapse.org>) with permission under the accession number syn4587609. The circRNAs were extracted from our previous study¹⁰, which were identified by NCLscan⁶² according to the same Synapse RNA-seq data used in this study and the Ensembl annotation (version 90) based on the human reference genome (GRCh38). A sample was excluded in our analysis if the number of identified circRNAs of the sample was below one standard deviation from the mean of the sample set¹⁰. A total of 212 samples, including 73 FC, 61 TC, and 78 CV samples, were retained (Table 1). For accuracy, we only considered the 1,060 circRNAs that were detected in more than 50% of the samples examined. The genotyping data were derived from 79 individuals, including 40 ASD patients and 39 non-ASD-affected controls²⁶. Since eQTL analysis required both RNA-seq and genotyping data from the same individuals, RNA-seq data from 165 samples were examined in the following analyses (Fig. 1a, Table 1, and Supplementary Table 1). The expression levels of circRNAs were calculated using the number of supporting circRNA junction reads per million uniquely mapped reads (RPM)²⁷ and adjusted for biological/technical covariates including the corresponding host gene expression, brain region, diagnosis (ASD/non-ASD), age, sex, RNA quality (RNA integrity number; RIN), post-mortem interval (PMI), sequencing batch, and brain bank (Supplementary Table 1). The expression levels of genes were downloaded from Parikshak et al.'s study⁵, which were measured by log₂ normalized fragments per kilobase million mapped reads (FPKM) values. Imputation genotyping data (39,166,259 SNPs) were downloaded using the Synapse command line client. We excluded SNPs with imputed quality score (R^2) < 0.8 and minor allele frequency (MAF) < 0.01 and thus retained 7,714,475 SNPs for the following eQTL analysis. We limited our analysis to SNPs in ± 200 kb nucleotides of each back-splice site and evaluated circQTLs by testing the correlations between the imputed genotype dosages and covariate-adjusted circRNA expression using Matrix eQTL with an additive linear model²⁸ for adjusting for covariates including sex, age, diagnosis, brain region, RIN, PMI, sequencing batch, brain bank batch. To control for multiple testing, P values will be adjusted across all circQTLs using false positive rate (FDR) correction (i.e., Benjamini-Hochberg correction). At FDR<0.05, we thus identified 989 circQTLs based on the 165 samples from all three brain regions and 666 circQTLs based on the 105 cortex samples, respectively (Supplementary Table 2). To compare the effect sizes for the 989 circQTL-circRNA pairs from different brain regions (TC, FC, or CV), we conducted a Z-test and calculated Z-scores as

$$Z = \frac{\beta_{TC} - \beta_{FC}}{\sqrt{SE_{FC}^2 + SE_{TC}^2}}$$

Here the equation showed an example for TC-FC comparison Z-test. SE represents standard error. We examined whether the 666 circQTLs were also *trans*-eQTLs affecting the expression of *trans*-eGenes using Matrix eQTL with an additive linear model for adjusting for the covariates abovementioned. Such circQTL SNPs and *trans*-eGenes should be located on different chromosomes or the same chromosome separated by a distance greater than 5Mb (from the SNP site to the transcription start site of the *trans*-

eGene). At FDR<0.05, 546 circQTLs were identified to be also *trans*-eQTLs (Fig. 2a; Supplementary Table 3).

Detection of reverse complementary sequences (RCSs). The criteria for detecting RCSs were similar to those presented in a previous study⁶³. For a circRNA, both flanking sequences (± 200 kb nucleotides of the back-splice site) were aligned each other using BLAST⁶⁴ with parameters `-task blastn -word_size 11 -strand minus`. The potential RCSs should be simultaneously satisfied the following rules: `bitscore> 100`, `alignment length>50 bp`, and `identity>80%`.

Determination of miRNA and RBP binding sites. CLIP-supported RBP binding sites on hg19 assembly were downloaded from ENCORI⁴⁹ at <http://starbase.sysu.edu.cn/>. The liftOver tool⁶⁵ was employed to obtain the genomic coordinates of RBP binding sites on the GRCh38 assembly. The miRNA-circRNA, miRNA-mRNA, RBP-circRNA, and RBP-mRNA interactions supported by CLIP-seq data were also extracted from ENCORI. These data were retrieved through ENCORI-provided Web APIs. The used command lines were listed in Supplementary Table 7. Of note, a considered miRNA (or RBP) binding site should be supported by at least one AGO CLIP-seq (or CLIP-seq) experiments. CircRNA-miRNA-mRNA and circRNA-RBP-mRNA interactions were determined according to the common miRNA and RBP target sites of the circRNAs and mRNAs, respectively. A developed package, CircMiMi, for constructing CLIP-seq data-supported circRNA-miRNA-mRNA and circRNA-RBP-mRNA interactions was also provided (<https://github.com/TreesLab/CircMiMi>). In CircMiMi, users can input their identified circRNAs, not limited to ENCORI-provided circRNAs.

Enrichment analyses based on empirical sampling. A sampling procedure was conducted to test whether the 666 circQTL SNPs (associated with 48 circRNAs) had a higher proportion ($proportion_{obs}$) of SNPs located in RCSs residing in the introns flanking circularized exons compared to an empirical distribution of the proportion observed in the 10,000 times of random sampling (see Fig. 1f). For each time, the equivalent number (666) of SNPs were randomly selected from the background set (823,378 SNPs) and the proportion ($proportion_i$) of SNPs located in RCSs was calculated. The 823,378 SNPs were located in the flanking sequences (± 200 kb nucleotides of the back-splice site) for the 48 circRNAs examined. The empirical P value was calculated as

$$\frac{1 + \sum_{i=1}^{10,000} \text{number of } (proportion_i > proportion_{obs})}{10,001}$$

A similar procedure was designed to test whether the 666 circQTL SNPs were predominantly located in RBP binding sites residing in the introns flanking circularized exons for the 48 circRNAs examined (Fig. 1g). Regarding the 43,372 circQTL-*trans*-eGene pairs (associated with 47 circRNAs, 7,165 *trans*-eGenes, and 13,379 circRNA-*trans*-eGene pairs; see Fig. 2a), we conducted a sampling procedure to test whether the 13,379 circRNA-*trans*-eGene pairs had a higher proportion of pairs with a significant correlation ($P<0.05$ by Spearman's correlation test) between expression of the circRNAs and the *trans*-eGenes

compared to an empirical distribution of the proportion observed in the 10,000 times of random sampling. For each time, the equivalent number (13,379) of circRNA-gene pairs was randomly selected from the 47 circRNAs and the 7,165 *trans*-eGenes. To test whether the CIT-passing *trans*-eGenes (708 genes) had a significantly higher probability of ASD risk⁴⁵ than background gene set (20,070 protein-coding genes), an empirical gene enrichment analysis was conducted on the basis of an empirical distribution of the proportion observed in the 10,000 times of random sampling (Supplementary Fig. 3a). The 20,070 genes were selected because they were expressed in the cortex on the basis of the RNA-seq data examined in this study. The equivalent number (708) of genes was randomly selected from the 20,070 genes for each time. A similar procedure was conducted to test whether the CIT-passing *trans*-eGenes were more intolerant of a loss function mutation (measured by pLI scores⁴⁶) compared with the background gene set (Supplementary Fig. 3b). The probability of ASD risk and pLI score for each gene were downloaded from <http://asd.princeton.edu/> and <https://gnomad.broadinstitute.org/> (loss-of-function curation results; v2.1.1), respectively.

Mediation and causal testing. We conducted mediations analysis using the mediation package³⁷ downloaded from the Comprehensive R Archive Network (CRAN) at <https://cran.r-project.org/web/packages/mediation/index.html>. To test the mediation effects of *trans*-eQTLs, the mediation package was performed with circQTL (*trans*-eQTL) as the “exposure,” circRNA expression as the “mediator,” *trans*-eGene expression as the “outcome,” and adjusting for covariates including sex, age, diagnosis, brain region, RIN, PMI, sequencing batch, and brain bank batch. The proportion of mediation for the circQTL was the proportion of causal mediation effect in the total effect (i.e., causal mediation effect + direct effect)³⁷. To test whether the circRNA-*trans*-eGene associations were mediated by the expression of miRNAs, we first extracted miRNA expressing data based on the 63 cortex samples (30 ASD and 33 non-ASD samples) overlapped with the samples examined in our eQTL analysis. The miRNA expressing data were obtained by the request from the authors of the Wu et al.’s study⁶, which were measured by log₂ normalized read counts with adjusting for library size, GC content, RIN, PMI, and batch bank. The mediation package was then performed with circRNA expression as the “exposure,” miRNA expression as the “mediator,” *trans*-eGene expression as the “outcome,” and adjusting for covariates abovementioned. Significant mediation effects were defined at a simulation threshold of $P < 0.05$ (sims=1,000) and FDR (Storey and Tibshirani method⁶⁶) < 0.05 .

The CIT package^{38,39} was downloaded from CRAN at <https://cran.r-project.org/web/packages/cit/index.html>. A CIT-passing circQTL-*trans*-eGene pair (CIT $P < 0.05$ and FDR (q) < 0.05) (Supplementary Table 4) represents that the circQTL has a causal effect on the diagnosis status (ASD vs. non-ASD) through the *trans*-eGene expression, which should simultaneously satisfy the following conditions: (1) the circQTL is associated with the diagnosis status, (2) the circQTL is associated with the *trans*-eGene mediator after adjusting for the diagnosis status, (3) the *trans*-eGene mediator is associated with the diagnosis status after adjusting for the circQTL, and (4) the circQTL is independent of the diagnosis status after adjusting for the *trans*-eGene mediator. The P value of CIT is defined as the maximum of the above four-component test P values. The FDR (q) values were further

estimated using CIT-provided permutation test with default parameters. Illustration of networks was plotted by the Cytoscape package (<https://cytoscape.org/>).

Functional analysis of the CIT-passing *trans*-eGenes. The SFARI⁴⁰ gene list was downloaded from <https://gene.sfari.org/> (SFARI 08-07-2020 release). The lists of genes encoding PSD proteins and targets of FMRP, RBFOX1, and ELAVL1 were downloaded from Lee et al.'s study²⁶. The list of DEGs in ASD was downloaded from Parikshak et al.'s study⁵. The lists of genes associated with Alzheimer's disease, amyotrophic lateral sclerosis, Huntington's disease, intellectual disability, and Parkinson's disease were downloaded from Wang et al.'s study⁶⁷. The schizophrenia gene list was downloaded from a GWAS report⁶⁸. The list of DEGs in a cell type-specific manner was downloaded from Velmeshev et al.'s study¹⁵. For gene set enrichment analyses in Figure 3b, we used the abovementioned 20,070 protein-coding genes as the background set. We took the analysis of SFARI gene enrichment for the CIT-passing *trans*-eGenes as an example. We created a two-way contingency table with rows containing numbers of SFARI and non-SFARI genes and columns containing numbers of CIT-passing *trans*-eGenes and genes that were not CIT-passing *trans*-eGenes. The statistical significance and odds ratio were determined using a one-tailed Fisher's exact test with the *fisher.test* R function. *P* values were FDR adjusted across 13 target groups for each gene list using Bonferroni correction. GO analysis was performed using the ToppFun module of ToppGene Suite software⁶⁹ at <https://toppgene.cchmc.org/enrichment.jsp> (version 2020-Sep-08 01:39). False discovery rate (FDR) adjusted *P* values were calculated using Bonferroni correction.

Hypergeometric testing. For a circQTL (*trans*-eQTL), we conducted a hypergeometric test to evaluate the significance of overlap of the *trans*-eGenes regulated by the circQTL and SFARI genes. Only the circQTLs that affected the expression of more than 50 *trans*-eGenes were considered for the test (Supplementary Table 6). We used 20,070 protein-coding genes as the background set, which involved 951 SFARI genes. The statistical significance was determined using the *phyper* R function. The coordinates of LD blocks were downloaded from the DistILD database⁷⁰ at <http://distild.jensenlab.org>.

Data availability. The CircMiMi package for constructing circRNA-miRNA-mRNA and circRNA-RBP-mRNA interactions is available at <https://github.com/TreesLab/CircMiMi>. The identified circQTLs/*trans*-eQTLs and the related data/codes were deposited in Supplementary Tables 1-7 or GitHub at https://github.com/TreesLab/circQTL_ASD.

Declarations

Acknowledgements

This work was supported by Genomics Research Center (GRC), Academia Sinica, Taiwan; the Ministry of Science and Technology (MOST), Taiwan, under the contract MOST 108-2311-B-001-020-MY3 (TJC); and the Taipei, Taichung, Kaohsiung Veterans General Hospital, Tri-Service General Hospital, Academia Sinica Joint Research Program under the contract AS-VTA-109-03 (TJC).

Author contributions

T.-L.M. performed the bioinformatics assays, analyzed the data, and helped writing the manuscript. C.-Y.C. performed the statistical assays and analyzed the data. T.-W.C. developed and maintained the CircMiMi package. T.-J.C. conceived the study, analyzed the data, and wrote the article.

Competing interest: The authors declare no competing financial interests.

References

1. Ansel A, Rosenzweig JP, Zisman PD, Melamed M, Gesundheit B. Variation in Gene Expression in Autism Spectrum Disorders: An Extensive Review of Transcriptomic Studies. *Frontiers in Neuroscience* **10**, 601 (2016).
2. Grove J, *et al.* Identification of common genetic risk variants for autism spectrum disorder. *Nat Genet* **51**, 431-444 (2019).
3. Satterstrom FK, *et al.* Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism. *Cell* **180**, 568-584 e523 (2020).
4. Gupta S, *et al.* Transcriptome analysis reveals dysregulation of innate immune response genes and neuronal activity-dependent genes in autism. *Nat Commun* **5**, 5748 (2014).
5. Parikshak NN, *et al.* Genome-wide changes in lncRNA, splicing, and regional gene expression patterns in autism. *Nature* **540**, 423-427 (2016).
6. Wu YE, Parikshak NN, Belgard TG, Geschwind DH. Genome-wide, integrative analysis implicates microRNA dysregulation in autism spectrum disorder. *Nat Neurosci* **19**, 1463-1476 (2016).
7. Mor M, Nardone S, Sams DS, 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).
8. Tran SS, *et al.* Widespread RNA editing dysregulation in brains from autistic individuals. *Nat Neurosci* **22**, 25-36 (2019).
9. Gokool A, Anwar F, Voineagu I. The Landscape of Circular RNA Expression in the Human Brain. *Biol Psychiatry* **87**, 294-304 (2020).
10. Chen YJ, *et al.* Genome-wide, integrative analysis of circular RNA dysregulation and the corresponding circular RNA-microRNA-mRNA regulatory axes in autism. *Genome Res* **30**, 375-391 (2020).
11. Rockman MV, Kruglyak L. Genetics of global gene expression. *Nat Rev Genet* **7**, 862-872 (2006).
12. Nica AC, Dermitzakis ET. Expression quantitative trait loci: present and future. *Philos Trans R Soc Lond B Biol Sci* **368**, 20120362 (2013).
13. Pierce BL, *et al.* Mediation analysis demonstrates that trans-eQTLs are often explained by cis-mediation: a genome-wide analysis among 1,800 South Asians. *PLoS Genet* **10**, e1004818 (2014).

14. Grundberg E, *et al.* Mapping cis- and trans-regulatory effects across multiple tissues in twins. *Nat Genet* **44**, 1084-1089 (2012).
15. Liu X, *et al.* Functional Architectures of Local and Distal Regulation of Gene Expression in Multiple Human Tissues. *Am J Hum Genet* **100**, 605-616 (2017).
16. Sun BB, *et al.* Genomic atlas of the human plasma proteome. *Nature* **558**, 73-79 (2018).
17. Liu X, Li YI, Pritchard JK. Trans Effects on Gene Expression Can Drive Omnigenic Inheritance. *Cell* **177**, 1022-1034 e1026 (2019).
18. Yang F, Wang J, Consortium GT, Pierce BL, Chen LS. Identifying cis-mediators for trans-eQTLs across many human tissues using genomic mediation analysis. *Genome Res* **27**, 1859-1871 (2017).
19. Yao C, *et al.* Dynamic Role of trans Regulation of Gene Expression in Relation to Complex Traits. *Am J Hum Genet* **100**, 571-580 (2017).
20. Chen I, Chen CY, Chuang TJ. Biogenesis, identification, and function of exonic circular RNAs. *Wiley Interdiscip Rev RNA* **6**, 563-579 (2015).
21. Chen LL. The expanding regulatory mechanisms and cellular functions of circular RNAs. *Nat Rev Mol Cell Biol*, (2020).
22. Chen LL, Yang L. Regulation of circRNA biogenesis. *RNA Biol* **12**, 381-388 (2015).
23. You X, *et al.* Neural circular RNAs are derived from synaptic genes and regulated by development and plasticity. *Nat Neurosci* **18**, 603-610 (2015).
24. Rybak-Wolf A, *et al.* Circular RNAs in the Mammalian Brain Are Highly Abundant, Conserved, and Dynamically Expressed. *Mol Cell* **58**, 870-885 (2015).
25. Szabo L, *et al.* Statistically based splicing detection reveals neural enrichment and tissue-specific induction of circular RNA during human fetal development. *Genome Biol* **16**, 126 (2015).
26. Lee C, Kang EY, Gandal MJ, Eskin E, Geschwind DH. Profiling allele-specific gene expression in brains from individuals with autism spectrum disorder reveals preferential minor allele usage. *Nat Neurosci* **22**, 1521-1532 (2019).
27. Zhang XO, Wang HB, Zhang Y, Lu X, Chen LL, Yang L. Complementary sequence-mediated exon circularization. *Cell* **159**, 134-147 (2014).
28. Shabalín AA. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics* **28**, 1353-1358 (2012).
29. Oldham MC, *et al.* Functional organization of the transcriptome in human brain. *Nat Neurosci* **11**, 1271-1282 (2008).
30. Voineagu I, *et al.* Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature* **474**, 380-384 (2011).
31. Willsey AJ, *et al.* Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. *Cell* **155**, 997-1007 (2013).
32. Parikshak NN, *et al.* Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. *Cell* **155**, 1008-1021 (2013).

33. Liu Z, Ran Y, Tao C, Li S, Chen J, Yang E. Detection of circular RNA expression and related quantitative trait loci in the human dorsolateral prefrontal cortex. *Genome Biol* **20**, 99 (2019).
34. Ahmed I, Karedath T, Al-Dasim FM, Malek JA. Identification of human genetic variants controlling circular RNA expression. *RNA* **25**, 1765-1778 (2019).
35. Jeck WR, *et al.* Circular RNAs are abundant, conserved, and associated with ALU repeats. *RNA* **19**, 141-157 (2013).
36. Chuang TJ, *et al.* Integrative transcriptome sequencing reveals extensive alternative trans-splicing and cis-backsplicing in human cells. *Nucleic Acids Res* **46**, 3671-3691 (2018).
37. Tingley D YT, Hirose K, Keele L, Imai K. mediation: R Package for Causal Mediation Analysis. *Journal of Statistical Software* **59**, 1-38 (2014).
38. Millstein J, Zhang B, Zhu J, Schadt EE. Disentangling molecular relationships with a causal inference test. *BMC Genet* **10**, 23 (2009).
39. Millstein J, Chen GK, Breton CV. cit: hypothesis testing software for mediation analysis in genomic applications. *Bioinformatics* **32**, 2364-2365 (2016).
40. Abrahams BS, *et al.* SFARI Gene 2.0: a community-driven knowledgebase for the autism spectrum disorders (ASDs). *Mol Autism* **4**, 36 (2013).
41. Iossifov I, *et al.* The contribution of de novo coding mutations to autism spectrum disorder. *Nature* **515**, 216-221 (2014).
42. Darnell JC, *et al.* FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. *Cell* **146**, 247-261 (2011).
43. Weyn-Vanhentenryck SM, *et al.* HITS-CLIP and integrative modeling define the Rbfox splicing-regulatory network linked to brain development and autism. *Cell Rep* **6**, 1139-1152 (2014).
44. Mukherjee N, *et al.* Integrative regulatory mapping indicates that the RNA-binding protein HuR couples pre-mRNA processing and mRNA stability. *Mol Cell* **43**, 327-339 (2011).
45. Krishnan A, *et al.* Genome-wide prediction and functional characterization of the genetic basis of autism spectrum disorder. *Nat Neurosci* **19**, 1454-1462 (2016).
46. Lek M, *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285-291 (2016).
47. Iossifov I, *et al.* Low load for disruptive mutations in autism genes and their biased transmission. *Proc Natl Acad Sci U S A* **112**, E5600-5607 (2015).
48. Velmeshev D, *et al.* Single-cell genomics identifies cell type-specific molecular changes in autism. *Science* **364**, 685-689 (2019).
49. Li JH, Liu S, Zhou H, Qu LH, Yang JH. starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Res* **42**, D92-97 (2014).
50. Bi C, *et al.* Mutations of ANK3 identified by exome sequencing are associated with autism susceptibility. *Hum Mutat* **33**, 1635-1638 (2012).

51. Smith KR, *et al.* Psychiatric risk factor ANK3/ankyrin-G nanodomains regulate the structure and function of glutamatergic synapses. *Neuron* **84**, 399-415 (2014).
52. Garza JC, *et al.* Disruption of the psychiatric risk gene Ankyrin 3 enhances microtubule dynamics through GSK3/CRMP2 signaling. *Transl Psychiatry* **8**, 135 (2018).
53. Buniello A, *et al.* The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res* **47**, D1005-D1012 (2019).
54. Pain O, *et al.* Novel Insight Into the Etiology of Autism Spectrum Disorder Gained by Integrating Expression Data With Genome-wide Association Statistics. *Biol Psychiatry* **86**, 265-273 (2019).
55. Hall LS, *et al.* Cis-effects on gene expression in the human prenatal brain associated with genetic risk for neuropsychiatric disorders. *Mol Psychiatry*, (2020).
56. Gandal MJ, *et al.* Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science* **362**, (2018).
57. Gusev A, *et al.* Integrative approaches for large-scale transcriptome-wide association studies. *Nat Genet* **48**, 245-252 (2016).
58. Wainberg M, *et al.* Opportunities and challenges for transcriptome-wide association studies. *Nat Genet* **51**, 592-599 (2019).
59. Fang WQ, *et al.* Overproduction of upper-layer neurons in the neocortex leads to autism-like features in mice. *Cell Rep* **9**, 1635-1643 (2014).
60. Fame RM, MacDonald JL, Macklis JD. Development, specification, and diversity of callosal projection neurons. *Trends Neurosci* **34**, 41-50 (2011).
61. Iwata K, *et al.* N-ethylmaleimide-sensitive factor interacts with the serotonin transporter and modulates its trafficking: implications for pathophysiology in autism. *Mol Autism* **5**, 33 (2014).
62. Chuang TJ, Wu CS, Chen CY, Hung LY, Chiang TW, Yang MY. NCLscan: accurate identification of non-co-linear transcripts (fusion, trans-splicing and circular RNA) with a good balance between sensitivity and precision. *Nucleic Acids Res* **44**, e29 (2016).
63. Ivanov A, *et al.* Analysis of intron sequences reveals hallmarks of circular RNA biogenesis in animals. *Cell reports* **10**, 170-177 (2015).
64. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* **215**, 403-410 (1990).
65. Hinrichs AS, *et al.* The UCSC Genome Browser Database: update 2006. *Nucleic Acids Res* **34**, D590-598 (2006).
66. Storey JD, Tibshirani R. Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A* **100**, 9440-9445 (2003).
67. Wang P, Zhao D, Lachman HM, Zheng D. Enriched expression of genes associated with autism spectrum disorders in human inhibitory neurons. *Transl Psychiatry* **8**, 13 (2018).
68. Schizophrenia Working Group of the Psychiatric Genomics C. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421-427 (2014).

69. Chen J, Bardes EE, Aronow BJ, Jegga AG. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res* **37**, W305-311 (2009).
70. Palleja A, Horn H, Eliasson S, Jensen LJ. DistiLD Database: diseases and traits in linkage disequilibrium blocks. *Nucleic Acids Res* **40**, D1036-1040 (2012).

Tables

Table 1. Number of samples of the RNA-seq and genotyping data used in this project.

Data type	Number of samples		
	ASD	Non-ASD	Sum
RNA-seq data ⁵	111 (48)	101 (49)	212 (97)
FC samples	37	36	73
TC samples	36	25	61
CV samples	38	40	78
Genotyping data ²⁶	40	39	79
RNA-seq data with the corresponding genotyping data			
FC+TC+CV samples	81 (38)	84 (37)	165 (75)
FC+TC samples	53 (35)	52 (34)	105 (69)

Note. The numbers of the examined individuals are shown in parentheses. FC, frontal cortex (Brodmann Area 9). TC, temporal cortex (Brodmann Area 22, 41, and 42). CV, cerebellar vermis. The data were downloaded from the PsychENCODE Consortium Knowledge Portal at <https://www.synapse.org/#!/Synapse:syn4587609> under accession number syn458609.

Table 2. The observed *trans*-eGenes that were also previously identified as TWAS-prioritized genes/transcripts in ASD.

<i>Trans</i> -eGene	This study			TWAS reference	
	circQTL/ <i>trans</i> -eQTL SNP	circRNA	Passing CIT?*		Passing mediation test?*
<i>CRHR1</i>	rs79543065, rs80018118, rs117601704	circERC1-1	No	Yes (52-54%)	Gandal et al. ⁵⁶
	rs572758241, rs145151802, rs188227047, rs146232238, rs148453911, rs147867900, rs141400333, rs150595674, rs182196454	circSPHKAP	No	Yes (51-72%)	
	rs118122108, rs79177473, rs117315428	circERC1-2	No	No	
	rs182290762	circZNF236	No	No	
<i>XKR6</i>	rs572758241	circSPHKAP	No	Yes (58%)	
<i>NKX2-2</i>	rs182920567	circACVR2A	No	No	
<i>NSF</i>	rs182920567	circACVR2A	Yes	No	Hall et al. ⁵⁵
<i>ARL17B</i>	rs572758241	circSPHKAP	No	Yes (33%)	

* “Passing CIT” and “passing mediation test” represent that the circQTL-*trans*-eGene-ASD diagnosis axes and the circQTL-circRNA-*trans*-eGene axes pass CIT (see also Supplementary Tables 4) and mediation test (see also Supplementary Tables 3), respectively. The proportions of mediation of the circQTL-*trans*-eGene associations mediated by the corresponding circRNA expression were denoted in parentheses.

Figures

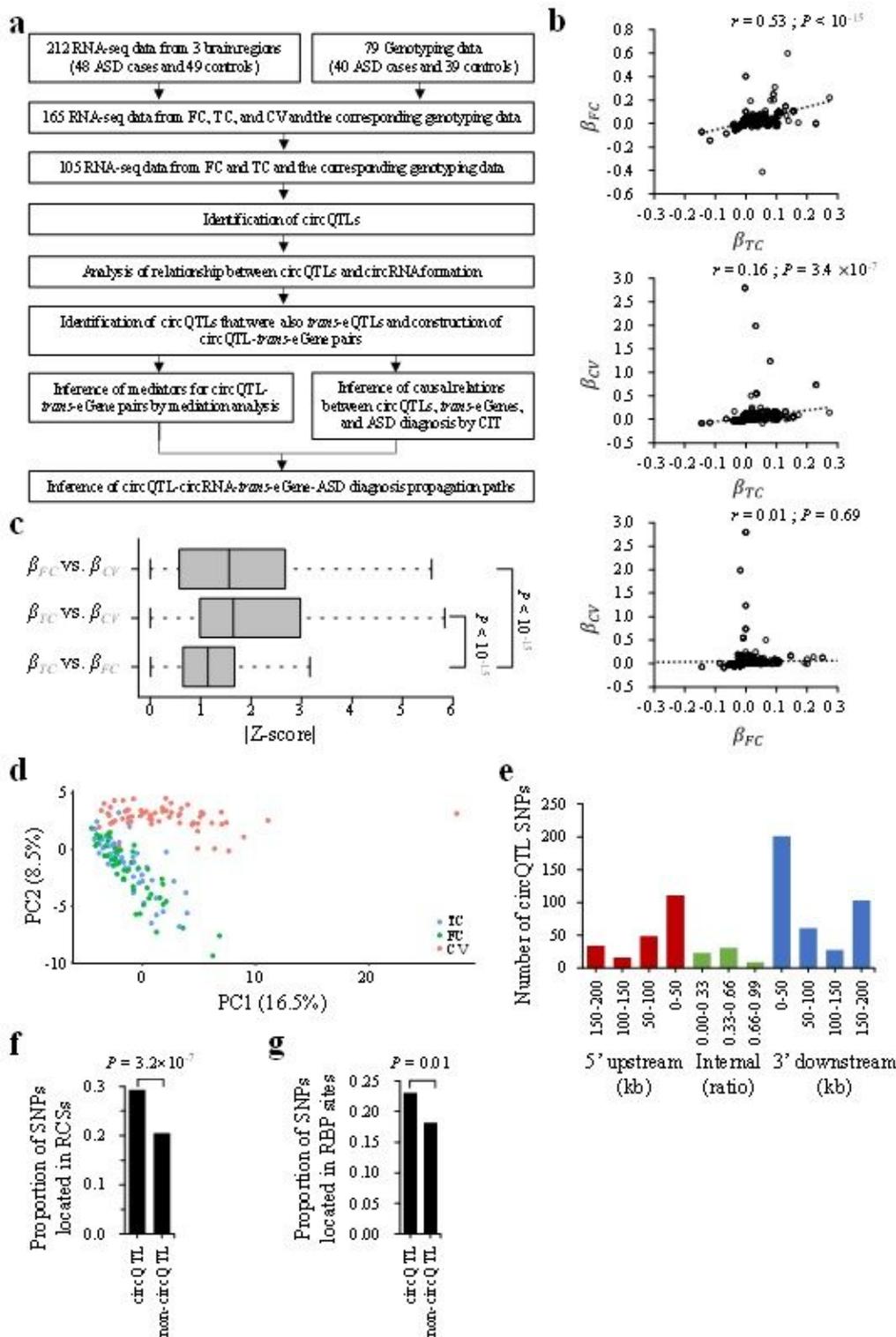


Figure 1

Identification and analysis of circQTLs based on the brain samples from people with ASD and controls. (a) Flowchart of the overall approach. (b, c) Comparisons of effect sizes (b) for circQTL-circRNA associations observed in samples from frontal cortex (FC), temporal cortex (TC), and cerebellar vermis (CV) using Pearson correlation analysis (b) and Z-test (c). For c, P values of differences in absolute values of Z-scores were determined using two-tailed Wilcoxon signed-rank test. (d) Principal component

plots of circRNA expression profiles of the 81 circRNAs associated with the identified circQTL-circRNA pairs in samples from TC, FC, and CV. PC1/PC2, the first and second principal components. (e) Distance distribution of the circQTL SNPs to their nearest back-spliced sites. For internal circQTL SNPs, the distance was calculated by the ratio of the distance between the SNP and the donor-to-acceptor back-spliced site. (f, g) Comparison of proportion of circQTL and non-circQTL SNPs located in (f) RCS and (g) predicted RBP binding sites. The non-circQTL SNPs were also located in the flanking sequences (± 200 kb nucleotides) of the examined back-splice sites. P values were determined using two-tailed Fisher's exact test.

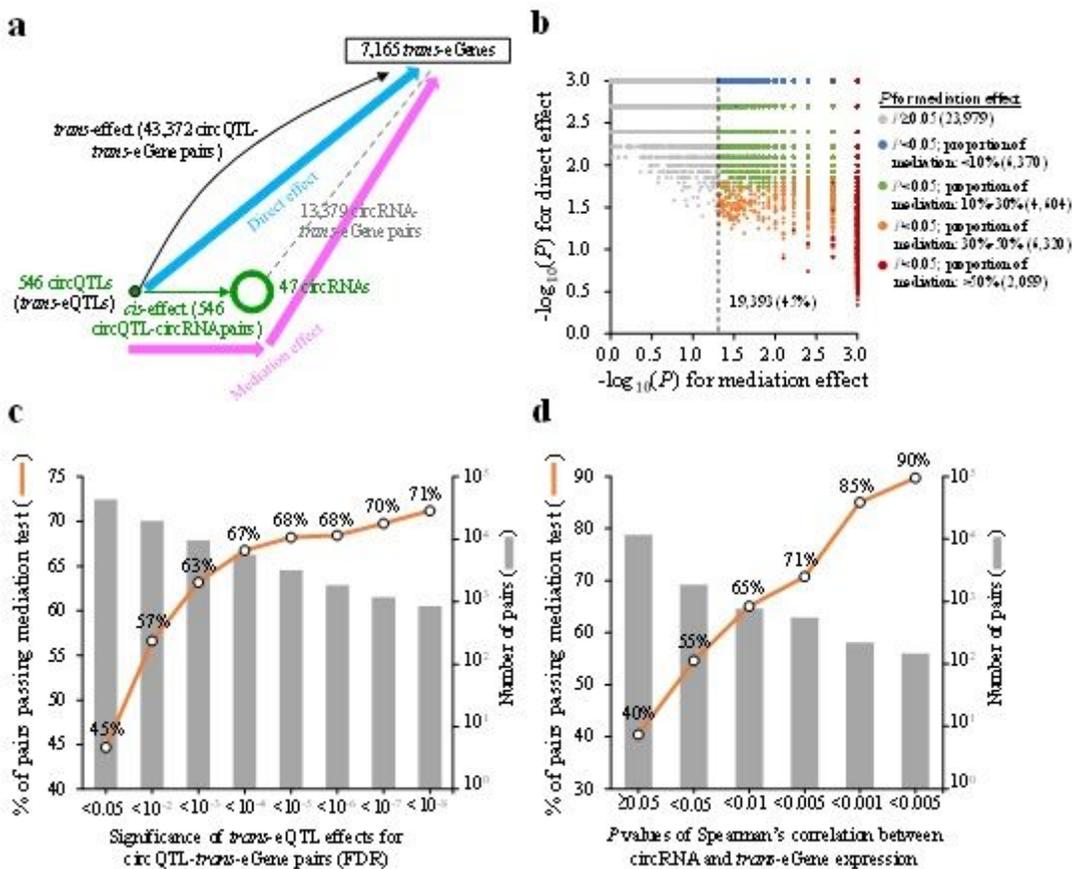


Figure 2

Identification and analysis of circQTLs based on the brain samples from people with ASD and controls. (a) Flowchart of the overall approach. (b, c) Comparisons of effect sizes (b) for circQTL-circRNA associations observed in samples from frontal cortex (FC), temporal cortex (TC), and cerebellar vermis (CV) using Pearson correlation analysis (b) and Z-test (c). For c, P values of differences in absolute values of Z-scores were determined using two-tailed Wilcoxon signed-rank test. (d) Principal component plots of circRNA expression profiles of the 81 circRNAs associated with the identified circQTL-circRNA pairs in samples from TC, FC, and CV. PC1/PC2, the first and second principal components. (e) Distance distribution of the circQTL SNPs to their nearest back-spliced sites. For internal circQTL SNPs, the distance was calculated by the ratio of the distance between the SNP and the donor-to-acceptor back-spliced site. (f, g) Comparison of proportion of circQTL and non-circQTL SNPs located in (f) RCS and (g)

predicted RBP binding sites. The non-circQTL SNPs were also located in the flanking sequences (± 200 kb nucleotides) of the examined back-splice sites. P values were determined using two-tailed Fisher's exact test.

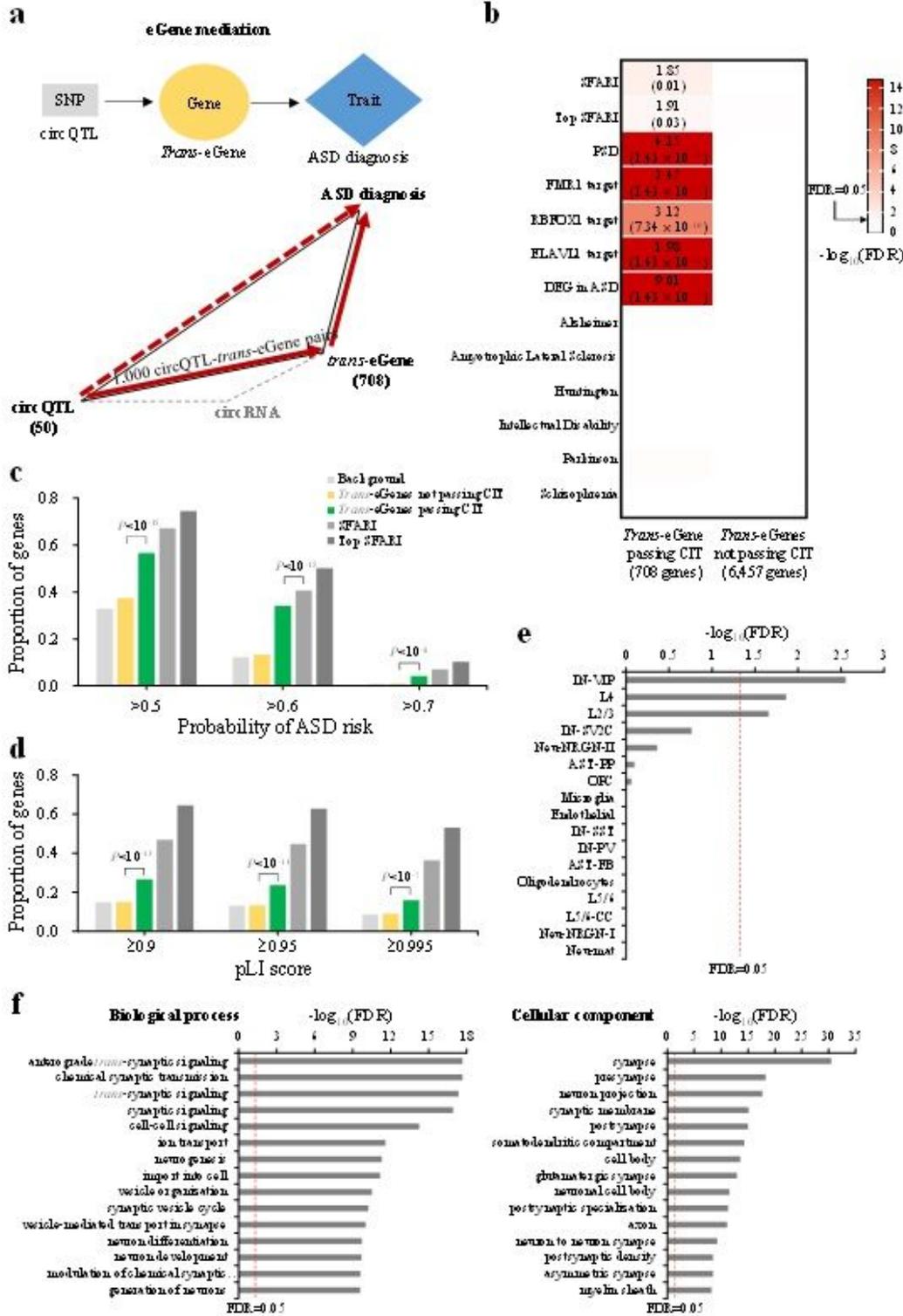


Figure 3

Inference of causal relations between circQTLs (trans-eQTLs), trans-eGenes, and ASD diagnosis by the causal inference test (CIT)38,39. (a) The circQTL-trans-eGene pairs (1,000 pairs) passing CIT with the

propagation path from circQTL SNPs (50 circQTLs) to ASD diagnosis via trans-eGene (708 genes) expression. (b) Heat map showing enrichment of ASD-relevant genes and genes implicated in monogenetic forms of other brain disorders for trans-eGenes passing CIT and trans-eGenes not passing CIT. ASD-relevant genes included ASD risk genes from SFARI, genes encoding postsynaptic density (PSD) proteins, genes encoding transcripts bound by FMR1 (FMR1 target), RBFOX1 (RBFOX1 target), and ELAVL1 (ELAVL1 target), and differentially expressed genes (DEGs) in ASD. Top SFARI genes represented SFARI genes with score = 1 or 2. The enrichment odd ratios and FDR Bonferroni-corrected P values (in parentheses) were shown for enrichment with FDR < 0.05. (c, d) Comparisons of the proportions of trans-eGenes passing CIT and trans-eGenes not passing CIT with (c) different probabilities of ASD risk and (d) different gene variant intolerance (pLI) scores. P values were determined using two-tailed Fisher's exact test. Background, the background gene set of 20,070 protein-coding genes. (e) Enrichment analysis of DEGs in a cell type-specific manner for trans-eGenes passing CIT. (f) GO analysis for trans-eGenes passing CIT. For b and e, P values were determined using one-tailed Fisher's exact test. For e and f, the red dashed lines represented FDR Bonferroni-corrected P values for enrichment with FDR = 0.05.

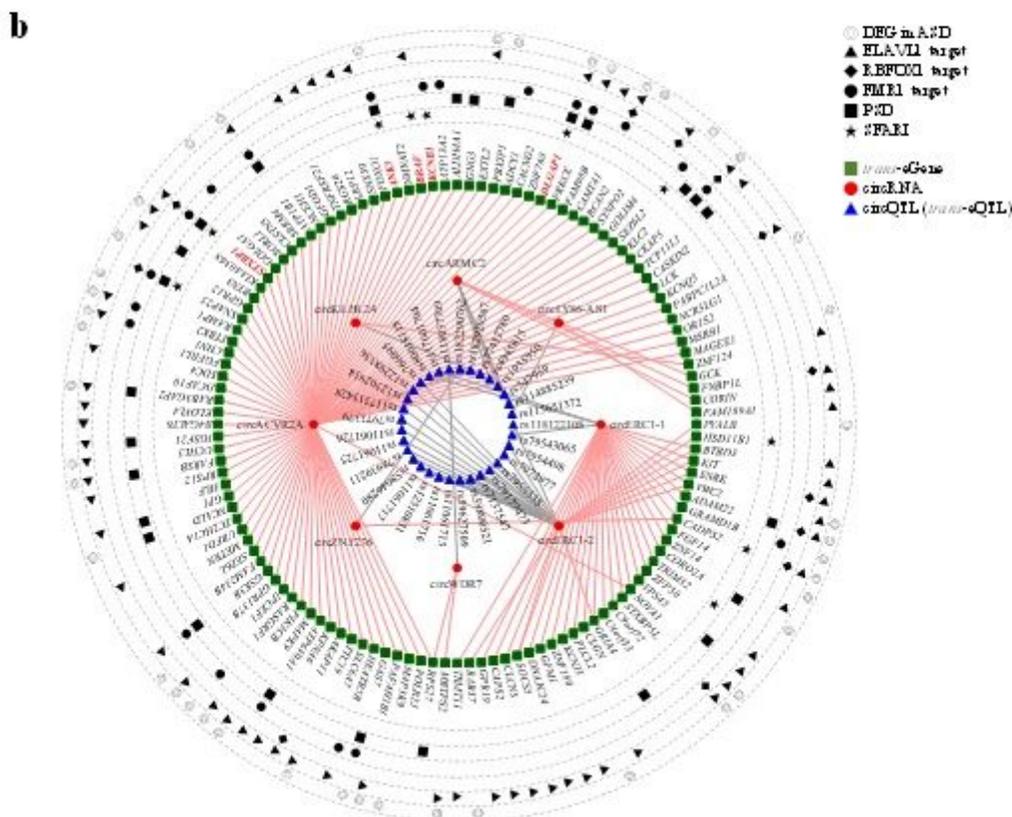
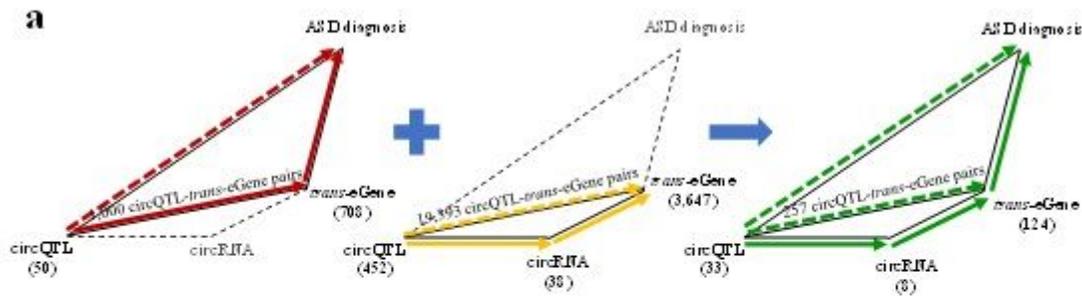


Figure 4

Potential circQTL-circRNA-trans-eGene-trait (ASD diagnosis) propagation paths. (a) Identification of potential circQTL-circRNA-trans-eGene-trait propagation paths (right) by integrating CIT-passing circQTL-trans-eGene pairs (left) with mediation-passing circQTL-circRNA-trans-eGene axes (middle). (b) The 257 identified circQTL-circRNA-trans-eGene-trait propagation networks plotted by the Cytoscape package. The ASD-relevant trans-eGenes were shown, in which the top SFARI genes were highlighted in red.

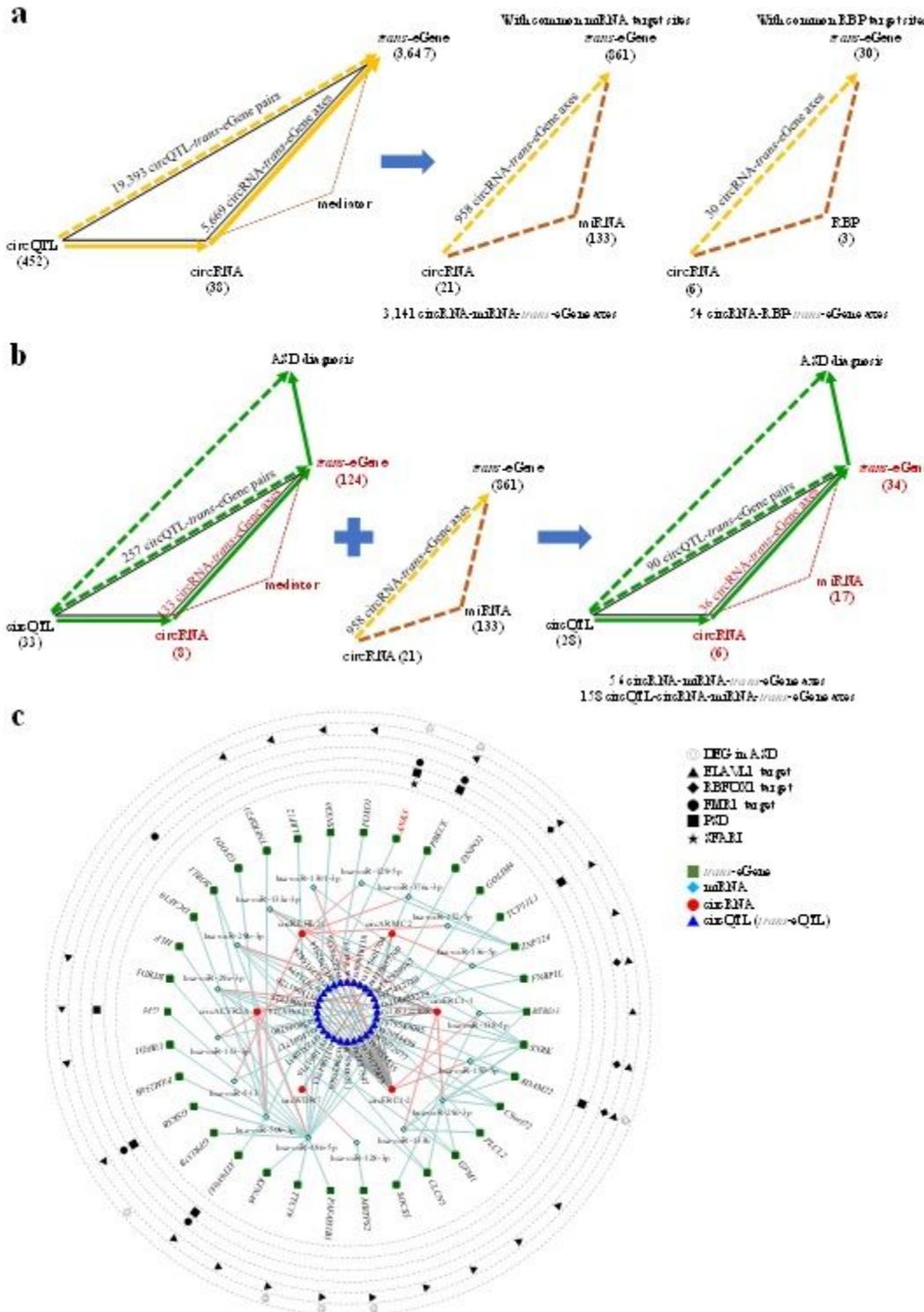


Figure 5

Potential mediators for the circRNA-trans-eGene interactions. (a) Identification of potential circRNA-miRNA-trans-eGene and circRNA-RBP-trans-eGene axes according to the common miRNA and RBP target sites of the circRNAs and trans-eGenes, respectively. (b) Identification of potential circQTL-circRNA-miRNA-trans-eGene-trait propagation paths (right) by integrating the 257 identified circQTL-circRNA-trans-eGene-trait propagation paths (left; see also Fig. 4) with the identified circRNA-miRNA-trans-eGene axes in (a) (middle). (c) The 158 identified circQTL-circRNA-miRNA-trans-eGene-trait propagation networks plotted by the Cytoscape package. The ASD-relevant trans-eGenes were shown, in which the top SFARI genes were highlighted in red.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryFigTable.docx](#)
- [SupplementaryTable1Dataused.xlsx](#)
- [SupplementaryTable2circQTLs.xlsx](#)
- [SupplementaryTable3transmediation.xlsx](#)
- [SupplementaryTable4CITnew20210115.xlsx](#)
- [SupplementaryTable5miRNARBP.xlsx](#)
- [SupplementaryTable6LDblock.xlsx](#)