Identification of *CDH11* as an ASD risk gene by matched-gene co-expression analysis and mouse behavioral studies

Nan Wu[1], Yue Wang[2], Yi-Hsuan Pan[1*], Xiao-Bing Yuan[1,3*]

[1]Key Laboratory of Brain Functional Genomics of Shanghai and Ministry of Education, Institute of Brain Functional Genomics, School of Life Science and the Collaborative Innovation Center for Brain Science, East China Normal University, Shanghai 200062, People's Republic of China

[2]Hussman Institute for Autism, Baltimore, MD 21201, USA

[3]Department of Anatomy and Neurobiology, University of Maryland School of Medicine, Baltimore, MD 21201, USA

*Correspondence and requests for materials should be addressed to Dr. Xiao-Bing Yuan (email: xbyuan@brain.ecnu.edu.cn) or Dr. Yi-Hsuan Pan (email: yihsuanp@gmail.com)

Abstract

In the study of autism spectrum disorder (ASD) by gene co-expression analysis (GCA), we found that four gene features, including gene size, mRNA length, mRNA abundance, and guanine-cytosine content, profoundly affect gene co-expression profiles. To circumvent the potential interference of these confounding factors on GCA, we developed the "matched-gene co-expression analysis" (MGCA) to investigate gene co-expression relationships. This method demonstrated the convergent expression profile of high confidence ASD risk genes and effectively revealed convergent molecular pathways of ASD risk genes. Application of MGCA to two ASD candidate genes *CDH11* and *CDH9* showed association of *CDH11*, but not *CDH9*, with ASD. Mouse behavioral studies showed that *Cdh11*-null mice, but not *Cdh9*-null mice, have multiple autistic-like behavioral alterations. This study confirmed that *CDH11* is an important ASD risk gene and demonstrated the importance of considering matched gene features in the analysis of gene co-expression.

Introduction

Autism spectrum disorder (ASD) is a group of heterogeneous neurodevelopmental conditions with a complex genetic basis [1,2]. A large number of susceptibility genes whose mutations or copy number variations (CNV) may be associated with ASD have been identified by genetic linkage analyses, genome-wide association studies (GWAS), whole-exome sequencing (WES), or whole-genome sequencing (WGS) [3,4]. However, the functions of most of these risk genes in developing brains remain unknown, and a causal relationship between their variations and autism traits has not been established. In order to prioritize investigation of genes and signaling pathways of high relevance to ASD, a method for efficient prediction of the functional importance of a large group of risk genes is vital.

The highly diverse ASD risk genes are believed to functionally converge in several common molecular pathways closely relevant to autism, such as the Wnt signaling pathway, the mammalian target of rapamycin (mTOR) pathway, and dendrite development and synaptic remodeling pathways [3,5]. Consistent with the hypothesis of functional convergence of ASD risk genes, results of several studies suggest the

convergence of developmental expression profiles of a large group of risk genes [6,7]. It is generally believed that genes with similar expression profiles are co-regulated or have related functions [6,8]. Although the co-expression of individual genes may be a coincident event, the co-expression of a group of functionally related genes is unlikely to be a random occurrence. Co-expression of genes within a biological pathway is a strong indication of their shared functions [8]. Based on this concept, computational analyses of various brain transcriptomes have been conducted to identify potential co-expression networks of ASD risk genes and to discover brain circuits that may be affected by the risk genes [6,7,9,10]. In these studies, the correlation coefficient (CC) of a pair of genes is calculated based on their expression levels in different brain regions and/or developmental stages. Genome-wide gene co-expression networks are constructed by setting an empirically determined threshold of CC [6]. A major limitation in most of these studies is lack of consideration of potential effects of confounding factors such as the size, expression level (mRNA abundance), and guanine-cytosine (GC) content of genes on the result of GCA [11]. Most ASD risk genes are large genes with a higher expression level in the brain than in other tissues [12]. It is unclear whether the size or expression level of an ASD gene affect its co-expression with other genes. It is also unknown whether the convergent pattern of developmental expression profiles is specific to ASD risk genes or a common property of genes with similar features, such as large gene size and high mRNA abundance [11].

Some ASD risk genes code for adhesion molecules, such as members of neurexin and neuroligin families, which mediate pre- and post-synaptic adhesion, respectively, in ASD-related brain circuits [13,14]. Genetic variants of several other adhesion molecules, including classical and non-classical cadherin family members, are also frequently found to be associated with ASD by GWAS [15] and WES studies [16-19]. Cadherin family members play important roles in multiple developmental processes, including cell proliferation, polarization, neuronal migration, axon projection, dendrite arborization, and synapse assembly, by mediating homophilic and heterophilic cell-cell interactions [20-24]. It is unclear as to which cadherin family members are functionally important in ASD and which ASD-relevant brain areas are affected by cadherin mutations.

In this study, we discovered that four gene features, including mRNA abundance, genomic DNA (gDNA) size, mRNA size, and GC content of the coding region of a gene, profoundly affect gene co-expression profiles in the brain. We developed a novel method called “matched-gene co-expression analysis” (MGCA) to examine whether a gene exhibits significant co-expression with a group of high-confidence ASD risk genes (hcASDs). This was accomplished by statistically comparing the co-expression level of a gene with the hcASD gene set to that of this gene with a large number of permuted gene sets of matched features (see method). Compared with the method without considering these gene features, MGCA was found to be more effective in identifying functionally convergent molecular pathways of ASD risk genes. With MGCA, the cadherin gene *CDH11*, but not *CDH9*, was found to be associated with ASD. This finding was corroborated by mouse behavioral studies using *Cdh11* and *Cdh9* knockout mice.

Results

Effects of gene features on gene co-expression profiles

The potential effects of the four gene features, including mRNA abundance, mRNA size, gDNA size, and GC content of the coding region, on gene co-expression profiles were first analyzed. The BrainSpan human brain transcriptome dataset was used for this analysis. This dataset contains transcriptomes of human (both gender) brain tissues from 16 different brain regions of various developmental stages and ages (from PCW8 to 40Y). A total of 15,942 genes with information on all 4 gene features were identified and used for analyses. These genes were placed in ascending order of mRNA abundance, mRNA size, gDNA size, or GC content as gene lists (***Figure 1—source data 1***). The correlation coefficient (CC) of each gene pair was calculated to reflect the co-expression level of the two genes, and the results were displayed in pseudo color-coded matrices. In each of the CC matrices (***Figure 1a***), these 15,942 genes were placed in ascending order on both x and y axes from the lowest mRNA abundance and GC content or the smallest gDNA and mRNA size to the highest mRNA abundance and GC content or the largest gDNA and mRNA size. All four CC matrices

were found to exhibit uneven color intensity in different areas with higher intensity corresponding to higher CC values. The overall color intensity was the highest in areas corresponding to medium mRNA abundance, medium to high gDNA or mRNA size, and low GC content (***Figure 1a***). This result suggests that all four gene features may affect gene co-expression profiles.

Most hcASDs are large genes with medium to high mRNA abundance levels, but with no clear bias in GC content (***Figure 1—figure supplement 1***). To determine whether each of these four gene features affects the co-expression of a gene with the hcASD gene set as a whole, the co-expression coefficient (CEC, mean CC between a gene and each of the hcASD genes) of each of the 15,942 genes with the entire hcASD gene set was calculated (blue dots in ***Figure 1b***; ***Figure 1—source data 1***). In each of the 4 panels (***Figure 1b***), the 15,942 genes were placed in ascending order (x-axis). A noise-reduced (by data averaging) CEC distribution curve was then generated by plotting the average CEC of a gene with its neighboring 20 (10 above and 10 below; ±10), 50 (±25), 100 (±50), or 200 (±100) genes on the gene lists under each gene ranking condition. Results showed a bell-shaped curve when genes were ranked by mRNA abundance, suggesting that genes with medium expression levels are more likely to co-express with the hcASD gene set (***Figure 1b, left panel***). There was an overall positive correlation between gDNA or mRNA size of a gene and its CEC with the hcASD gene set (***Figure 1b, middle two panels***). The CEC curve peaked at genes with approximately 40% GC content (x-axis between 1000-2000; y-axis between 0.28-0.31) and gradually declined with increasing GC content (***Figure 1b***, right panel***; Figure 1—source data 1***).

With cubic regression, each noise-reduced CEC distribution curve was found to have an R^2 value > 0.9 (***Figure 1—figure supplement 2; Figure 1—source data 2***), indicating a very high correlation between each of these gene features and the tendency of co-expression of a gene with the hcASD gene set. When the 15,942 genes were placed in stochastic (random) orders, CECs were evenly distributed, and the noise-reduced CEC distribution curves were flat (***Figure 1c***).

Similar genome-wide gene co-expression profiles of the hcASD gene set were observed in transcriptomes of early (8PCW-2Y) and late (4Y-40Y) stages (***Figure 1—***

figure supplement 3a), both gender, and different brain regions (***Figure 1—figure supplement 3b, c***). These findings suggest that the co-expression profile of hcASD genes is affected by all four gene features, regardless of developmental stages, gender, and brain areas.

The genome-wide co-expression profile of the hcASD gene set was then compared to those of gene sets with an equal number (64) of genes in the top, middle, or bottom positions of the gene lists corresponding to highest, medium, and lowest in mRNA abundance and GC content or largest, medium, and smallest in mRNA and gDNA sizes (***Figure 1—source data 3; Figure 1d***). When genes were ranked by mRNA abundance, the noise-reduced CEC distribution curve of the hcASD gene set largely overlapped with that of the gene set of median mRNA abundance levels (Middle in ***Figure 1d***); this result is consistent with the observation that most hcASDs are genes of moderate mRNA abundance. The noise-reduced CECs of the top mRNA abundance gene set (Top in ***Figure 1d***) were positively correlated with mRNA abundance, whereas those of the bottom mRNA abundance gene set (Bottom in ***Figure 1d***) were negatively correlated with mRNA abundance. When genes were ranked by gDNA or mRNA size, the noise-reduced CEC distribution curve of the hcASD gene set was most similar to, and higher than, that of the gene set of largest genes (Top in ***Figure 1d***). This result is also consistent with the fact that most hcASDs are large genes. When genes were ranked by GC content, the noise-reduced CEC distribution curve of the hcASD gene set greatly deviated from those of the Top, Middle, and Bottom gene sets (***Figure 1d***), consistent with the lack of correlation with the GC content of hcASD genes.

Similar co-expression profiles of feature-matched gene sets

The genome-wide gene co-expression profile of the hcASD gene set was then compared to the profiles of 200 non-hcASD gene sets, each comprised equal number (64) of randomly selected and feature-matched non-hcASD genes under the four different gene ranking conditions (***Figure 2a***). These gene sets were named “matched random” (mRand) gene sets (see methods). In general, the genome-wide CEC distribution of hcASDs was similar to that of each of the 200 mRand gene sets under all four gene ranking conditions. These findings suggest that gene sets with matched gene

features have similar genome-wide co-expression profile as the hcASD gene set. However, genes with moderate mRNA abundance had higher noise-reduced CECs with the hcASD gene set than with any of the 200 mRand gene sets. In contrast, both low and high mRNA abundance genes (< 1 and > 30 RPKM; < 3000 and >14000 on x-axis) had lower noise-reduced CECs with the hcASD gene set than with most mRand gene sets. Moreover, genes with medium to large sizes had higher noise-reduced CECs with the hcASD gene set than with most size-matched mRand gene sets. Most genes, except those with highest GC content, had higher noise-reduced CECs with the hcASD gene set than with most GC content-matched mRand gene sets.

Co-expression of ASD risk genes

To determine whether hcASDs exhibit a significant tendency of co-expression with each other, the mean CEC of each of the 64 hcASDs with the hcASD gene set as a whole (hcASD-hcASD, see method) was compared to that of a large number of permuted gene sets, each comprised equal number of feature-matched non-hcASD genes (mRand-mRand) or randomly selected non-hcASD genes (Rand-Rand) and to the CEC between hcASD and mRand (hcASD-mRand) or Rand (hcASD-Rand) gene sets. Two hundred each of mRand and Rand gene sets were first analyzed. Results showed that feature-matched gene sets (mRand) had overall higher CECs than random gene sets (Rand) under all four matched conditions (@@@ in ***Figure 2b***), suggesting that genes of similar features tend to co-express with each other. The CEC of hcASD-hcASD (dashed line in ***Figure 2b***) was 1.5 times higher than the interquartile range [Q3 + 1.5 x (Q3-Q1), upper fence] of the CECs of mRand-mRand, Rand-Rand, hcASD-mRand, and hcASD-Rand gene sets. This result suggests that hcASDs have a significantly greater co-expression tendency with each other than other feature-matched non-hcASD genes or randomly selected genes. Results of the Grubbs' test confirmed that this tendency (*** in ***Figure 2b***). To corroborate this finding, permutation test was further conducted with 100,000 permuted sets of genes with matched or non-matched features (see method). The CEC of hcASD-hcASD was still found to be significantly larger (permutation *p*-value < 0.001) than that of hcASD-mRand, mRand-mRand,

hcASD-Rand, or Rand-Rand (### in ***Figure 2b***), indicating a significant co-expression tendency of hcASDs.

Significant co-expression of hcASDs was also observed in transcriptomes of brain tissues from both early (8PCW – 2Y) and late (4Y – 40Y) stages (***Figure 2—figure supplement 1a, b***), both gender (***Figure 2—figure supplement 1c, d***), and different brain regions (***Figure 2—figure supplement 2a-d***). These results indicate a highly conserved mechanism for co-expression of hcASDs. Combined ranking of -log10 *p*-values of the Grubbs' test under all four different matched conditions was then performed to determine the relative significance level of co-expression of hcASDs with each other in different brain regions (***Figure 2—figure supplement 2e***). The top five brain regions with the highest significance levels were cerebellum (CB), dorsal frontal cortex (DFC), orbital frontal cortex (OFC), primary sensory cortex (S1C), and striatum (STR); these are the brain regions previously implicated in ASD [25-31]. These results suggest that hcASDs play important roles in the development and function of these ASD-relevant brain regions.

ASD-relevant pathways identified by MGCA

A gene whose CEC with hcASDs was significantly higher than its CECs with permuted sets of feature-matched genes ($p < 0.001$) was considered as significantly co-expressed with hcASDs. Results of this matched-gene co-expression analysis (MGCA) showed that 3931, 3330, 5629, and 5854 genes were significantly co-expressed with hcASDs under each of the four matched conditions, respectively (***Figure 3a***, ***Figure 3—resource data 1***), with a false calling rate below 5% (FDR < 0.05). The false discovery rate (FDR) of a gene was determined by the frequency of this gene significantly co-expressed ($P < 0.001$) with 10,000 mRand gene sets determined by MGCA.

Altogether, 2515 genes were found to significantly co-express with hcASDs under all four matched conditions with an estimated FDR of each gene below 6.25×10^{-6} (0.05^4); this gene set was named TetraM-2515 (***Figure 3a***, ***Figure 3—resource data 1***). TetraM-2515 was then compared with 2515 genes that had the highest CECs with the hcASD gene set (referred to as Top-2515 gene set, ***Figure 3—resource data 1***). TetraM-2515 and Top-2515 gene sets had 1500 genes in common (Overlapped) and

each had 1015 non-overlapped genes; these two non-overlapped gene sets were named TetraM-only and Top-only, respectively (***Figure 3b***, ***Figure 3—resource data 1***). Most Top-2515 genes had a medium mRNA abundance level, a large gene size, and a high CEC value (> 0.4), whereas TetraM-2515 genes had a broad range of mRNA abundance, gene size, and CEC values (***Figure 3b***). TetraM-2515 genes and Top-2525 genes had 34 and 32 genes overlapped with hcASD genes, respectively.

Gene ontology (GO) enrichment analysis of the TetraM-2515 gene set showed significant over representation of genes in molecular pathways closely related to ASD, including “covalent chromatin modification”, “protein polyubiquitination”, “homophilic cell adhesion”, “axon guidance”, “negative regulation of dendrite development”, “synapse assembly”, “Wnt signaling pathway”, and “RNA splicing”. The Top-2515 gene set also showed significant enrichment of genes in several pathways relevant to ASD, including “covalent chromatin modification”, “mRNA slicing”, “protein ubiquitination”, and “Wnt signaling pathway” (***Figure 3c***, ***Figure 3—resource data 2***).

To investigate the functional relationship between hcASD and TetraM-2515 or Top-2515 gene sets, an integrated GO enrichment network of multiple gene sets was constructed [32]. Genes of TetraM-only, Top-only, Overlapped, and the hcASD gene sets were divided into 130 nodes based on GO and KEGG annotations. Nodes were connected based on the similarities between node pairs (Kappa similarity > 0.3). These 130 nodes formed 20 networks. Based on the number of genes from each of the four gene sets in each node, some networks were found to be dominated by genes from one of these four gene sets (***Figure 3d***, ***Figure 3—resource data 3***). None of hcASDs were in networks dominated by Top-only genes. The hcASD-dominated network was found to connect with networks dominated by TetraM-only genes or Overlapped genes (***Figure 3d***, ***Figure 3—resource data 3***), suggesting that hcASDs have a closer functional relationship with TetraM-2515 genes than with Top-2515 genes.

MGCA was performed to further analyze an expanded set of ASD risk genes containing 1166 non-redundant ASD risk genes from ten different sets of previously reported ASD risk genes [6,17,33-39] (***Figure 3—resource data 4***). An integrated GO enrichment network of this combined ASD gene set (cASD) along with TetraM-only, Top-only, and Overlapped gene sets was constructed. Results showed that all gene

sets were converged separately in a subset of functional pathways, however the pathway patterns of TetraM-only and Top-only gene sets were largely complementary to each other (***Figure 3—figure supplement 1a***). The pathways of TetraM-only genes were related to brain developmental processes, such as “dendrite development”, “synapse development”, and “neuronal projection morphogenesis” and that of Top-only genes were related to “mRNA processing” and “covalent chromatin modification”. (***Figure 3—figure supplement 1a, resource data 5***). More co-localized nodes of cASD and TetreM-only genes than that of cASD and Top-only genes were seen in the network graph (***Figure 3—figure supplement 1b, resource data 5***). These results suggest that cASD genes have a closer functional relationship with TetraM-only genes than with Top-only genes.

Co-expression of cadherin genes with hcASDs

Consistent with previous findings [7], MGCA revealed that “homophilic cell adhesion” is the most significantly over-represented pathway of TetraM-2515 genes (***Figure 3c, d***; ***Figure 3—resource data 2, 3***). Some cadherin family members, such as CDH2, in the TetraM-2515 gene set are known to be high risk genes of ASD (***Figure 3—resource data 6***) that play important roles in brain circuit development [24]. Several cadherin family members were also found in the TetraM-2515 gene set, including many members of the protocadherins β gene cluster and Dachsous Cadherin-related 1 (*DCHS1*), suggesting that these genes also participate in the development and function of ASD-relevant brain circuits. Some cadherin genes were not significantly co-expressed with hcASDs under any of the matched conditions; these genes were referred to as tetra-negative genes (TetraN; ***Figure 3—resource data 6***). Several recent genetic studies have implicated two type II cadherins, *CDH11* and *CDH9*, in ASD and other psychiatric diseases [40-44]. As *CDH11* and *CDH9* belonged to the TetraM and TetraN gene sets, respectively, we hypothesized that *CDH11*, but not *CDH9*, is more likely to be an authentic ASD risk gene.

Autistic-like traits of *Cdh11*-null mice

To assess the functional relevance of *CDH11* and *CDH9* to ASD, the behaviors of *Cdh11* knockout (KO) and *Cdh9* KO mice were investigated. In the open field test (OFT), both male and female *Cdh11*-null mice spent a longer time exploring the central area of the open field arena than wild type (WT) littermates (***Figure 4a, d***). Heterozygous littermates showed a similar but less significant pattern. Total locomotion and average moving speed of *Cdh11* KO mice were slightly reduced compared to WT littermates (***Figure 4b, c***). Both male and female *Cdh9*-null mice were largely normal in the OFT (***Figure 4e-g***).

In the elevated plus maze test, female *Cdh11*-null mice visited the open arm more frequently and spent a significantly longer time there. Heterozygous females spent a slightly but not statistically significant more time in the open arm (***Figure 4h, i***). The increased time and frequency of open arm exploration by female *Cdh11*-null mice is consistent with the results of a previous study using the same mouse line of mixed gender [45]. Male *Cdh9*-null mice showed longer exploration of the open arm, but female *Cdh9*-null mice did not, although female heterozygotes showed an increased frequency of open arm entry (***Figure 4j, k***).

Individuals with ASD often have a weaker grip strength than age-matched controls [46]. The gripping strength test and the horizontal bar test showed that both male and female *Cdh11*-null mice exhibited significantly shorter hanging duration than WT littermates (***Figure 5a, b***), indicating reduced gripping strength and/or impaired motor coordination. The gripping strength of *Cdh9*-null mice was normal (***Figure 5c***).

The rotarod test was conducted to evaluate motor-related functions of KO mice. Since female and male mutant mice displayed similar behaviors in most of the above behavioral tests, only female mice were analyzed in this test. Compared to WT littermates, *Cdh11*-null mice, but not *Cdh9*-null mice, stayed longer on the rotarod and endured a higher rotation speed in the initial trial (***Figure 5d-g***). In subsequent trials, *Cdh11*-null mice did not display significant improvement in performance (***Figure 5d, e***), indicating impaired motor learning. The enhanced performance of *Cdh11*-null mice in the initial trial was very similar to the phenotype of several other well-characterized ASD mouse models and suggested increased repetitive motion of these mutant mice [47].

Repetitive behaviors were then evaluated by measuring the duration and frequency of self-grooming within a 10-minute period, during which mice were placed in a novel or a relatively familiar environment. As shown in figure 5h-i, during the first 10 minutes of exploring a novel chamber, *Cdh11*-null mice exhibited a significantly greater frequency of self-grooming than WT littermates, indicating elevated repetitive behavior in a novel environment. *Cdh11*-null mice also showed a significantly higher frequency of self-grooming than WT littermates during the second 10-minute period (***Figure 5j, k***), indicating elevated repetitive behavior even in a relatively familiar environment. No such behavioral alteration was observed in *Cdh9*-null mice (***Figure 5l, m***).

The modified three-chamber social preference test was conducted to evaluate the sociability of mutant mice. One main modification was an enlargement of the area for housing social partner mice in order to reduce their potential stress and anxiety. Another major modification to the protocol was using three mice instead of a single mouse as social partners. This was done to increase the availability of social cues and reduce the variability of test results caused by differences in the sociability of individual social partners (***Figure 6a***). In addition, the two side chambers were covered on the top to slow the diffusion and mixing of odorant cues. Results showed that female *Cdh11*-null mice exhibited a significant preference to social partner mice than to an object and to novel partners than to familiar ones (***Figure 6b, c***). However, compared to WT littermates, mutant mice spent significantly longer time in the middle chamber but significantly shorter time interacting with partner mice (***Figure 6b, c***), indicating reduced sociability. In contrast, *Cdh9*-null mice did not show any abnormality in this test (***Figure 6d, e***).

Discussion

Gene co-expression analysis (GCA) is a powerful tool to find functionally convergent genes. Several previous ASD studies using the GCA method had considered the potential influence of gene size and GC content on the test of a convergent expression of ASD risk genes[6] but had ignored the potential influence of the gene abundance. None of these studies had taken these factors into account in the construction of gene co-expression network. In the present study, we discovered that

four gene features, including mRNA abundance, mRNA length, gDNA size, and GC content, affected the genome-wide gene co-expression profiles in the brain. Although how these gene features affect gene co-expression profiles is unclear, this finding suggests the importance of considering the influence of these gene features in GCA. Instead of setting a threshold of correlation coefficient (CC) for gene co-expression analysis as in most other studies, we screened for significant co-expression relationships by comparing the co-expression coefficient (CEC) of a gene with the hcASD gene set to that with permuted gene sets of matched gene features. Only genes that had a CEC with the hcASD gene set significantly higher than its CECs with permuted sets of feature-matched genes were considered to be co-expressed with hcASDs. This matched-gene co-expression analysis (MGCA) paradigm allowed demonstration of significant co-expression of hcASDs with each other and avoided the potential bias caused by empirically determined threshold for CC of gene pairs in GCA. By MGCA, we found that TetraM-2515 genes are enriched in several molecular pathways closely related to neuronal morphogenesis and synaptic development, which are most commonly affected in ASD. It is likely that many of these TetraM-2515 genes function synergistically with hcASDs in developmental processes, such as axon projection, dendrite development, or synapse assembly. However, Top-2515 genes are more prominently enriched in pathways related to gene expression regulation, such as epigenetic chromatin regulation and mRNA processing (***Figure 3c, d; Figure 3—figure supplement 1a, b***). Many of these genes are very likely to be either upstream genes that regulate the transcription of hcASDs or downstream target genes whose expression is regulated by hcASDs, and thus display relatively high co-expression score (CEC) with hcASDs. A very significant finding in this study is the association of *CDH11* with ASD determined by MGCA. The importance of *CDH11* in ASD was ignored by previous GCA studies due to a relatively low CC with other genes (***Figure 3—resource data 6***); such genes were excluded from the construction of the gene co-expression network because of CC values being below the empirically determined threshold. Therefore, in the determination of genes that have shared functions with hcASDs during brain developmental, MGCA will be an important compliment to current GCA methods that ignore matched gene features.

Cadherins have been shown to accumulate in synaptic junctions and regulate dendrite development and synapse maturation [48-51]. Several cadherin family members have been implicated in ASD [16,52-60]. For example, some protocadherins in the *FAT* cadherin subfamily were found by whole-exome sequencing to be associated with ASD [52 16]. A genetic association study of a large cohort of ASD individuals and matched controls revealed genes in the protocadherin α gene cluster (*PCDHA*) as ASD risk genes [53]. Mutations in the *PCDH19* gene have been shown to cause early-onset epilepsy, and many individuals with these mutations also display autistic features [54-56]. Mutations in the cadherin EGF LAG seven-pass G-type receptor 2 gene (*CELSR2*) were speculated to be responsible for the Joubert syndrome, a disease with a high degree of autistic features [61]. It is uncertain whether other cadherins are also high-risk factors. Using MGCA, we found that a group of cadherin superfamily members exhibited a high co-expression with hcASDs, suggesting shared functions with hcASDs and a role in ASD etiology. Among them, several protocadherins, mainly *PCDHB*s, exhibited significant co-expression with hcASDs (***Figure 3—resource data 6***). The functions of these *PCDHB*s in the brain remain to be determined. One of such cadherins identified by MGCA is CDH11. We found *Cdh11*-null mice have significantly increased repetitive activities. The brain regions including neocortex, CB, and STR are known to be involved in the control of repetitive behaviors [62]. It is likely that cadherins, *Cdh11* in particular, play important roles in mediating synapse formation during the wiring of circuits in these brain areas. Consistent with this postulation, our recent work showed *Cdh11* expression in ASD-associated sub-regions in the CB of developing mouse brain [63].

In human studies, partial deletion of *CDH11* was observed in a sporadic case of non-syndromic ASD, mild intellectual disability, and attention deficit hyperactivity disorder (ADHD) [40]. A case-control association study revealed a high prevalence of a homozygous single nucleotide variant rs7187376C/C of *CDH11* in patients with ASD [40]. Several other coding variants of *CDH11* were also discovered in ASD individuals [40]. Behavioral changes that we have observed in *Cdh11*-null mice, including reduced anxiety, increased repetitive behavior, and reduced sociability, are highly consistent with the of the non-syndromic ASD case with partial deletion of *CDH11* [40]. This observation supports the notion that loss-of-function of a single risk gene, such as *CDH11*, is

sufficient to cause several major autism traits. Behavioral phenotypes of ASD are highly heterogeneous. Some individuals with ASD are hypoactive with elevated anxiety, and some have attention deficit hyperactivity disorder (ADHD) but with reduced anxiety [64-67]. The genetic and neurobiological mechanisms underlying this behavioral heterogeneity have not been fully determined. Further investigation with a larger cohort of patient families is needed to determine whether loss-of-function mutations of *CDH11* are associated with ADHD.

Most genetic variants found in patients with ASD are heterozygous. In some behavioral tests, heterozygous *Cdh11* KO mice showed a similar trend of behavioral alterations as homozygous KO mice (***Figure 5j, k, Figure 6c***). As ASD has a complex genetic basis and is affected by environmental factors, it is conceivable that monogenic haplodeficiency of an important risk gene causes a relatively mild behavioral phenotype in mice. It is likely that more severe behavioral deficits may result if the haplodeficiency is combined with other genetic or environmental factors. Our findings suggest that *CDH11* is significantly co-expressed with hcASDs and that its mutations may exert a causal effect in autism traits. *Cdh11* KO mice would be very helpful in dissecting the circuit mechanisms underlying a subgroup of ASD and in screening drugs targeting this subgroup of ASD.

CDH9 plays an important role in the establishment of specific synaptic wiring in both the hippocampus and the retina [68,69]. Its association with ASD has been suggested by several genome-wide association studies (GWAS) [70,71]. The main evidence linking *CDH9* to ASD is the strong association of the single nucleotide polymorphism rs4307059 located in the intergenic region between *CDH10* and *CDH9* with ASD [70]. However, this rs4307059 genotype was not correlated with the expression of either *CDH9* or *CDH10* in adult brains [70,72], and whether a correlation exists in fetal brains is unknown. Recently, an antisense noncoding RNA of the moesin pseudogene 1 (MSNP1AS) was shown to be transcribed from the locus harboring rs4307059. Alterations in this pseudogene were postulated to be responsible for ASD [72-74]. Whether *CDH9* deficiency is a causal factor for ASD remains undetermined. Our MGCA showed that, unlike *CDH11*, *CDH9* was not co-expressed with hcASDs. This is an indication that *CDH9* may not play an important role in the wiring of ASD-relevant circuits. Consistent

with this notion, behavioral tests showed that *Cdh9*-null mice exhibited a very mild behavioral abnormality only in the elevated plus maze test, but not in any other tests. Together with recent findings by other researchers, our results suggest that *CDH9* deficiency may not have a major effect on autism traits.

Materials and Methods

Ethics statement

Animal care and handling were performed according to the guidelines for the Care and Use of Laboratory Animals of the National Institutes of Health. All animal experiments were approved by the Animal Care and Use Committees of Hussman Institute for Autism (06012015D), University of Maryland School of Medicine (0515017), and East China Normal University (m20190236).

Data filtering and computation of correlation coefficient

The human brain transcriptome dataset from BrainSpan (www.brainspan.org) (RNA-Seq Gencode v10) was used for gene co-expression analyses. This dataset contained 256 transcriptomes of 16 different brain regions. The developmental stages ranged from post-conception week 8 (PCW8) to 40 years old (40Y). Normalized mRNA expression values were represented by RPKM (Reads Per Kilobase Per Million Mapped Reads). The average mRNA expression level of each gene in all tissues was considered as the mRNA abundance level of a gene. Gene length and mRNA length were determined based on gene annotations provided by the National Center for Biotechnology Information (NCBI). The GC content in the coding region of a gene was obtained from GCevobase (Ensembl_release_88). Based on statistical analyses of genetic data described previously [75], 64 risk genes that reached a genome-wide significance threshold were used as the hcASD gene set in the present study (high confidence ASD risk gene set, Suppl. Table S1). Genes with an average expression level lower than the lowest expression level of hcASDs were filtered out (Suppl. Table S1). Perl scripts were written to conduct most calculations. Pair-wise Pearson's correlation coefficient (CC) was used to indicate the tendency of co-expression of a gene pair. Heatmaps were

constructed with the software R based on the CC matrix of 1/100 evenly distributed genes. The mean CC was defined as co-expression coefficient (CEC), which indicates the tendency of co-expression of a gene with a specific set of genes ($\mathrm{CEC} = \frac{1}{M}\sum_{i=1}^{M} CC_i$, $i = 1, 2, \ldots, M$; where *M* indicates the total gene number of a gene set) or the tendency of co-expression of two gene sets ($\mathrm{CEC} = \frac{1}{N*M}\sum_{k=1}^{N}(\sum_{i=1}^{M} CC_{k\,i})$, $k = 1, 2, \ldots, N$; $i = 1, 2, \ldots, M$; where *M* and *N* represent the total gene number of two different gene sets, respectively).

Gene set definition

After data filtering, a total of 15942 genes with information on gene length, mRNA length, mRNA level, and GC content available were identified from the BrainSpan human brain transcriptome dataset and used for the study (***Figure 1— source data 1***). The hcASD gene set contains 64 previously identified high confidence ASD risk genes[33]. In addition to the hcASD gene set, the following gene sets were also used: mRand, Rand, TetraM-2515, Top-2515, TetraM-only, and Top-only. Each mRand gene set contained 64 genes with one of the four gene features of each gene matched with that of the corresponding hcASD gene in the hcASD gene set. Each Rand gene set contained randomly selected 64 genes without considering matched gene features. TetraM-2515 was the gene set containing 2515 genes that exhibited significant co-expression with the hcASD gene set under all four matched conditions (see figure 3). The Top-2515 gene set contained top 2515 genes with the highest CEC values with the hcASD gene set. TetraM-only and Top-only gene sets contained non-overlapped genes that were present only in the TetraM-2515 and the Top-2515 gene sets, respectively.

Gene ontology (GO) analysis

GO analysis was performed using DAVID v6.8 (http://david.ncifcrf.gov/tools.jsp), and the human whole-genome genes provided by DAVID were used as the background list. For identification of significantly enriched GO terms, a corrected *P*-value of 0.05 (Benjamini-Hochberg method) was used for filtering.

Enrichment network of multiple gene sets

Metascape (http://metascape.org) was used to sort genes in each gene set into functional groups (nodes) based on GO and KEGG annotations. Gene enrichment networks were visualized by Cytoscape (https://cytoscape.org/) using 'force-directed' layout with edges bundled for clarity. For identifying enrichment of genes in a specific functional group (pathway), a *P*-value of 0.001 and an enrichment factor of 10 were set as the threshold of significance. The 20 most significant (lowest *P*-value) pathways were chosen for heatmap or network construction. In the heatmap (***Figure 3—figure supplement 1a***), pairwise similarities between any two significant terms were computed based on Kappa-test scores. The enriched terms were then hierarchically clustered into a tree with a kappa score of 0.3 as the threshold. Boxes were colored according to their *P*-values. Gray boxes indicate lack of enrichment for a specific GO term. In a network graph (***Figure 3d, Figure 3—figure supplement 1b***), each pie represented a node (a group of genes belonging to a GO term). Within each pie, different slices represented different input gene sets (gene sets used for analysis), coded by different colors, and the area of each slice was proportional to the number of genes in a selected gene set that was associated with the GO term of the node. Edges between nodes indicate a kappa score above the threshold (Kappa similarity > 0.3).

Animals

Cdh9 KO mice [69] were provided by Dr. Joshua R. Sanes at Harvard University. *Cdh11* KO mice [76] were obtained from the Jackson Labs (*Cdh11*^{tm1Mta}/HensJ, https://www.jax.org/strain/023494). All mice were housed in groups of five with free access to food and water and kept on a 12-hour light/dark cycle. All behavioral tests were carried out on mice 2-5 months of age. All tests were conducted during daytime. The surface of the apparatus for behavioral tests was cleaned with 50% ethanol between tests. At least 5 min between cleaning and the next test was allowed for ethanol evaporation and odor dissipation.

Genotyping

Genotyping of *Cdh9* KO mice was done by PCR as previously described [69]. The PCR product for the wildtype (WT) *Cdh9* allele was 550 bp amplified with the primer pair *Cdh9*-P1 (CCA CTA CAG GAA ACC TTT GGG TT) and Cdh9-P3 (ATG CAA ACC ATC AGG TAT ACC AAC C), and that of the mutant allele was 430 bp amplified with the primer pair *Cdh9*-P1 and *Cdh9*-P2 (CGT GGT ATC GTT ATG CGC CT). The annealing temperature for *Cdh9* PCRs was 63°C. For genotyping of *Cdh11* KO mice, the primer pair *Cdh11*-P1 (CGC CTT CTT GAC GAG TTC) and *Cdh11*-P2 (CAC CAT AAT TTG CCA GCT CA) was used for amplification of the mutant allele, and the primer pair *Cdh11*-P3 (GTT CAG TCG GCA GAA GCA G) and *Cdh11*-P2 was used for the WT allele. The annealing temperatures for PCR were 63.1^{o}C and 56°C for the mutant and WT alleles, respectively. Sizes of the PCR products for the mutant and WT alleles were 500 bp and 400 bp, respectively.

Behavioral tests

Open field test. The standard open field test was performed to evaluate gross locomotor activity, anxiety level, and repetitive behavior. The test mouse was allowed to freely explore the open field arena (50 cm × 50 cm) for 30 min. The motion of the mouse was videoed and tracked by an automated tracking system (EthoVision XT 11.5), which also recorded rearing, hopping, turning, self-grooming, moving time, total moving distance, and time spent in the center of the arena (1/2 of total size).

Elevated plus maze test. The standard elevated plus maze (EPM) apparatus consisted of two open and two closed arms, 12 x 2 inches each, connected by a central platform (2 x 2 inches). The maze was 20 inches off the ground. The test mouse was gently placed on the central platform with its head facing one closed arm and was allowed to freely explore for 10 min. The time that the mouse stayed in the two open arms and the frequency of open arm entry were recorded.

Grip strength test. The test mouse was placed on a metal grid on top of a transparent chamber. The grid was quickly inverted, and the time for the mouse to drop off the grid was determined. Five consecutive trials were carried out, and the average hanging time

for each mouse was calculated. The maximum hanging time was set for 1 min. After 1 min of hanging, the trial was stopped, and the hanging time was recorded as 1 min.
Horizontal bar test. The mouse was gently placed on a metal wire, with the two forepaws gripping the wire. The length of time which the mouse hung on the wire was measured. The maximum hanging time was set for 1 min. The average hanging time was calculated from 5 consecutive trials.

Rotarod test. Mice were habituated to the rotarod apparatus (Harvard Apparatus 760770) by leaving them on the low-speed rotating rod (4 rpm) for 5 min each day for 3 days and tested on the fourth day on the accelerating rod. The time and the maximum rotation speed that the test mouse was able to maintain the balance on the rotating rod were measured. Five consecutive trials were done for each mouse.

Social preference test. A modified three-chamber apparatus was used. The apparatus comprised 3 rectangular (10 x 15 inches) chambers made of white Plexiglas with a 5-inch gate connecting the two side chambers to the middle chamber. A 3-sided (5 inches wide for each side) fence made of clear Plexiglas was placed inside each side chamber facing the door of the side chambers, creating a 5-inch x 5-inch square area separated from the side chambers but connected to the middle chamber through the door (Fig 6a). The two side chambers were covered by transparent Plexiglas to minimize the diffusion and mixing of odor between chambers. To conduct the test, the test mouse was placed inside the middle chamber and allowed to freely explore the middle chamber and the square zone in each side chamber for 10 min. Three social partner mice were then placed into the fenced area in one side chamber, and the test mouse was allowed to freely explore for another 10 min. Another 3 social partner mice were then placed in the other side chamber, and the behavior of the test mouse was tracked for 10 min. The time that the test mouse spent in each chamber was measured.

Statistical Analysis

Data are presented as mean ± standard error of the mean (SEM). Upper fence test and Grubbs' test were performed to evaluate whether a specific CEC value was significantly

higher than the CEC values of 200 randomly selected sets of feature-matched or non-matched non-hcASD genes. Grubb's test was done using the "grubbs.test" script in R. For the permutation test, 100,000 mRand or Rand gene sets were used; hcASDs were not excluded from the permuted gene sets. Behavioral analyses were performed blind to genotypes. Data were analyzed using one-way ANOVA followed by student's *t*-test as post hoc analysis. Statistical analyses were performed with SPSS (IBM, Armonk, USA) or GraphPad Prism (GraphPad Software, La Jolla, CA, USA).

Availability of data and materials

Perl scripts for data analysis are available upon request to the corresponding authors.

Competing interests

The authors declare that they have no competing interests.

Acknowledgments

We thank Dr. Chao-Hung Lee and Elizabeth Benevides for editing the manuscript. This work was funded by the National Science Foundation of China (31771116, 31871501) and Hussman Foundation (HIAS15006). X.Y. was supported by the Simon Foundation (296143).

Authors' contributions

N.W., Y.W., and Y.P. analyzed data and generated figures. C.W. bred animals and performed genotyping. X.Y. conducted behavioral tests. Y.P. and X.Y. designed the study and wrote the manuscript.

References

1 Berg, J. M. & Geschwind, D. H. Autism genetics: searching for specificity and convergence. *Genome Biol* **13**, 247, doi:10.1186/gb4034 (2012).

2 Jeste, S. S. & Geschwind, D. H. Disentangling the heterogeneity of autism spectrum disorder through genetic findings. *Nat Rev Neurol* **10**, 74-81, doi:10.1038/nrneurol.2013.278 (2014).

3 de la Torre-Ubieta, L., Won, H., Stein, J. L. & Geschwind, D. H. Advancing the understanding of autism disease mechanisms through genetics. *Nat Med* **22**, 345-361, doi:10.1038/nm.4071 (2016).

4 Sullivan, P. F. & Geschwind, D. H. Defining the Genetic, Genomic, Cellular, and Diagnostic Architectures of Psychiatric Disorders. *Cell* **177**, 162-183, doi:10.1016/j.cell.2019.01.015 (2019).
5 Courchesne, E. *et al.* The ASD Living Biology: from cell proliferation to clinical phenotype. *Mol Psychiatry* **24**, 88-107, doi:10.1038/s41380-018-0056-y (2019).
6 Willsey, A. J. *et al.* Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. *Cell* **155**, 997-1007, doi:10.1016/j.cell.2013.10.020 (2013).
7 Parikshak, N. N. *et al.* Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. *Cell* **155**, 1008-1021, doi:10.1016/j.cell.2013.10.031 (2013).
8 Stuart, J. M., Segal, E., Koller, D. & Kim, S. K. A gene-coexpression network for global discovery of conserved genetic modules. *Science* **302**, 249-255, doi:10.1126/science.1087447 (2003).
9 Hormozdiari, F., Penn, O., Borenstein, E. & Eichler, E. E. The discovery of integrated gene networks for autism and related disorders. *Genome Res* **25**, 142-154, doi:10.1101/gr.178855.114 (2015).
10 Mahfouz, A., Ziats, M. N., Rennert, O. M., Lelieveldt, B. P. & Reinders, M. J. Shared Pathways Among Autism Candidate Genes Determined by Co-expression Network Analysis of the Developing Human Brain Transcriptome. *J Mol Neurosci* **57**, 580-594, doi:10.1007/s12031-015-0641-3 (2015).
11 Kopp, N., Climer, S. & Dougherty, J. D. Moving from capstones toward cornerstones: successes and challenges in applying systems biology to identify mechanisms of autism spectrum disorders. *Front Genet* **6**, 301, doi:10.3389/fgene.2015.00301 (2015).
12 Gabel, H. W. *et al.* Disruption of DNA-methylation-dependent long gene repression in Rett syndrome. *Nature* **522**, 89-93, doi:10.1038/nature14319 (2015).
13 Krueger, D. D., Tuffy, L. P., Papadopoulos, T. & Brose, N. The role of neurexins and neuroligins in the formation, maturation, and function of vertebrate synapses. *Curr Opin Neurobiol* **22**, 412-422, doi:10.1016/j.conb.2012.02.012 (2012).
14 Sudhof, T. C. Neuroligins and neurexins link synaptic function to cognitive disease. *Nature* **455**, 903-911, doi:10.1038/nature07456 (2008).
15 Hussman, J. P. *et al.* A noise-reduction GWAS analysis implicates altered regulation of neurite outgrowth and guidance in autism. *Mol Autism* **2**, 1, doi:10.1186/2040-2392-2-1 (2011).
16 Cukier, H. N. *et al.* Exome sequencing of extended families with autism reveals genes shared across neurodevelopmental and neuropsychiatric disorders. *Mol Autism* **5**, 1, doi:10.1186/2040-2392-5-1 (2014).
17 Iossifov, I. *et al.* The contribution of de novo coding mutations to autism spectrum disorder. *Nature* **515**, 216-221, doi:10.1038/nature13908 (2014).
18 Sanders, S. J. *et al.* De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* **485**, 237-241, doi:10.1038/nature10945 (2012).
19 Neale, B. M. *et al.* Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature* **485**, 242-245, doi:10.1038/nature11011 (2012).
20 Seong, E., Yuan, L. & Arikkath, J. Cadherins and catenins in dendrite and synapse morphogenesis. *Cell Adh Migr* **9**, 202-213, doi:10.4161/19336918.2014.994919 (2015).
21 Stocker, A. M. & Chenn, A. The role of adherens junctions in the developing neocortex. *Cell Adh Migr* **9**, 167-174, doi:10.1080/19336918.2015.1027478 (2015).
22 Gartner, A., Fornasiero, E. F. & Dotti, C. G. Cadherins as regulators of neuronal polarity. *Cell Adh Migr* **9**, 175-182, doi:10.4161/19336918.2014.983808 (2015).
23 Luccardini, C. *et al.* N-cadherin sustains motility and polarity of future cortical interneurons during tangential migration. *J Neurosci* **33**, 18149-18160, doi:10.1523/JNEUROSCI.0593-13.2013 (2013).

24 Takeichi, M. The cadherin superfamily in neuronal connections and interactions. *Nat Rev Neurosci* **8**, 11-20, doi:10.1038/nrn2043 (2007).
25 Tsai, P. T. *et al.* Autistic-like behaviour and cerebellar dysfunction in Purkinje cell Tsc1 mutant mice. *Nature* **488**, 647-651, doi:10.1038/nature11310 (2012).
26 Khan, A. J. *et al.* Cerebro-cerebellar Resting-State Functional Connectivity in Children and Adolescents with Autism Spectrum Disorder. *Biol Psychiatry* **78**, 625-634, doi:10.1016/j.biopsych.2015.03.024 (2015).
27 Buxhoeveden, D. P. *et al.* Reduced minicolumns in the frontal cortex of patients with autism. *Neuropathol Appl Neurobiol* **32**, 483-491, doi:10.1111/j.1365-2990.2006.00745.x (2006).
28 Green, S. A. *et al.* Neurobiology of Sensory Overresponsivity in Youth With Autism Spectrum Disorders. *JAMA Psychiatry* **72**, 778-786, doi:10.1001/jamapsychiatry.2015.0737 (2015).
29 Green, S. A. *et al.* Overreactive Brain Responses to Sensory Stimuli in Youth With Autism Spectrum Disorders. *Journal of the American Academy of Child & Adolescent Psychiatry* **52**, 1158-1172, doi:https://doi.org/10.1016/j.jaac.2013.08.004 (2013).
30 Di Martino, A. *et al.* Aberrant striatal functional connectivity in children with autism. *Biol Psychiatry* **69**, 847-856, doi:10.1016/j.biopsych.2010.10.029 (2011).
31 Wang, X. *et al.* Altered mGluR5-Homer scaffolds and corticostriatal connectivity in a Shank3 complete knockout model of autism. *Nat Commun* **7**, 11459, doi:10.1038/ncomms11459 (2016).
32 Zhou, Y. *et al.* Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun* **10**, 1523, doi:10.1038/s41467-019-09234-6 (2019).
33 Sanders, Stephan J. *et al.* Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. *Neuron* **87**, 1215-1233, doi:10.1016/j.neuron.2015.09.016 (2015).
34 Werling, D. M., Parikshak, N. N. & Geschwind, D. H. Gene expression in human brain implicates sexually dimorphic pathways in autism spectrum disorders. *Nat Commun* **7**, 10717, doi:10.1038/ncomms10717 (2016).
35 Basu, S. N., Kollu, R. & Banerjee-Basu, S. AutDB: a gene reference resource for autism research. *Nucleic Acids Research* **37**, D832-D836, doi:10.1093/nar/gkn835 (2008).
36 Abrahams, B. S. *et al.* SFARI Gene 2.0: a community-driven knowledgebase for the autism spectrum disorders (ASDs). *Mol Autism* **4**, 36, doi:10.1186/2040-2392-4-36 (2013).
37 Krishnan, A. *et al.* Genome-wide prediction and functional characterization of the genetic basis of autism spectrum disorder. *Nat Neurosci* **19**, 1454-1462, doi:10.1038/nn.4353 (2016).
38 C Yuen, R. K. *et al.* Whole genome sequencing resource identifies 18 new candidate genes for autism spectrum disorder. *Nature Neuroscience* **20**, 602-611, doi:10.1038/nn.4524 (2017).
39 pfeliciano@simonsfoundation.org, S. C. E. a. & Consortium, S. SPARK: A US Cohort of 50,000 Families to Accelerate Autism Research. *Neuron* **97**, 488-493, doi:10.1016/j.neuron.2018.01.015 (2018).
40 Crepel, A. *et al.* Association of CDH11 with non-syndromic ASD. *Am J Med Genet B Neuropsychiatr Genet* **165B**, 391-398, doi:10.1002/ajmg.b.32243 (2014).
41 Gasso, P. *et al.* Association between genetic variants related to glutamatergic, dopaminergic and neurodevelopment pathways and white matter microstructure in child and adolescent patients with obsessive-compulsive disorder. *J Affect Disord* **186**, 284-292, doi:10.1016/j.jad.2015.07.035 (2015).
42 Sokolowski, M., Wasserman, J. & Wasserman, D. Polygenic associations of neurodevelopmental genes in suicide attempt. *Mol Psychiatry*, doi:10.1038/mp.2015.187 (2015).
43 Wang, K. *et al.* A genome-wide association study on common SNPs and rare CNVs in anorexia nervosa. *Mol Psychiatry* **16**, 949-959, doi:10.1038/mp.2010.107 (2011).
44 Anazi, S. *et al.* Clinical genomics expands the morbid genome of intellectual disability and offers a high diagnostic yield. *Mol Psychiatry* **22**, 615-624, doi:10.1038/mp.2016.113 (2017).

45 Manabe, T. *et al.* Loss of cadherin-11 adhesion receptor enhances plastic changes in hippocampal synapses and modifies behavioral responses. *Mol Cell Neurosci* **15**, 534-546, doi:10.1006/mcne.2000.0849 (2000).
46 Travers, B. G. *et al.* Brainstem White Matter Predicts Individual Differences in Manual Motor Difficulties and Symptom Severity in Autism. *J Autism Dev Disord* **45**, 3030-3040, doi:10.1007/s10803-015-2467-9 (2015).
47 Rothwell, P. E. *et al.* Autism-associated neuroligin-3 mutations commonly impair striatal circuits to boost repetitive behaviors. *Cell* **158**, 198-212, doi:10.1016/j.cell.2014.04.045 (2014).
48 Bekirov, I. H., Nagy, V., Svoronos, A., Huntley, G. W. & Benson, D. L. Cadherin-8 and N-cadherin differentially regulate pre- and postsynaptic development of the hippocampal mossy fiber pathway. *Hippocampus* **18**, 349-363, doi:10.1002/hipo.20395 (2008).
49 Friedman, L. G. *et al.* Cadherin-8 expression, synaptic localization, and molecular control of neuronal form in prefrontal corticostriatal circuits. *J Comp Neurol* **523**, 75-92, doi:10.1002/cne.23666 (2015).
50 Bian, W. J., Miao, W. Y., He, S. J., Qiu, Z. & Yu, X. Coordinated Spine Pruning and Maturation Mediated by Inter-Spine Competition for Cadherin/Catenin Complexes. *Cell* **162**, 808-822, doi:10.1016/j.cell.2015.07.018 (2015).
51 Tsai, N. P. *et al.* Multiple autism-linked genes mediate synapse elimination via proteasomal degradation of a synaptic scaffold PSD-95. *Cell* **151**, 1581-1594, doi:10.1016/j.cell.2012.11.040 (2012).
52 Butler, M. G., Rafi, S. K., Hossain, W., Stephan, D. A. & Manzardo, A. M. Whole exome sequencing in females with autism implicates novel and candidate genes. *Int J Mol Sci* **16**, 1312-1335, doi:10.3390/ijms16011312 (2015).
53 Anitha, A. *et al.* Protocadherin alpha (PCDHA) as a novel susceptibility gene for autism. *J Psychiatry Neurosci* **38**, 192-198, doi:10.1503/jpn.120058 (2013).
54 Cappelletti, S. *et al.* Cognitive development in females with PCDH19 gene-related epilepsy. *Epilepsy Behav* **42**, 36-40, doi:10.1016/j.yebeh.2014.10.019 (2015).
55 van Harssel, J. J. *et al.* Clinical and genetic aspects of PCDH19-related epilepsy syndromes and the possible role of PCDH19 mutations in males with autism spectrum disorders. *Neurogenetics* **14**, 23-34, doi:10.1007/s10048-013-0353-1 (2013).
56 Camacho, A. *et al.* Cognitive and behavioral profile in females with epilepsy with PDCH19 mutation: two novel mutations and review of the literature. *Epilepsy Behav* **24**, 134-137, doi:10.1016/j.yebeh.2012.02.023 (2012).
57 Sanders, S. J. *et al.* Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. *Neuron* **70**, 863-885, doi:10.1016/j.neuron.2011.05.002 (2011).
58 Pagnamenta, A. T. *et al.* Rare familial 16q21 microdeletions under a linkage peak implicate cadherin 8 (CDH8) in susceptibility to autism and learning disability. *J Med Genet* **48**, 48-54, doi:10.1136/jmg.2010.079426 (2011).
59 Moya, P. R. *et al.* Rare missense neuronal cadherin gene (CDH2) variants in specific obsessive-compulsive disorder and Tourette disorder phenotypes. *Eur J Hum Genet* **21**, 850-854, doi:10.1038/ejhg.2012.245 (2013).
60 Redies, C., Hertel, N. & Hubner, C. A. Cadherins and neuropsychiatric disorders. *Brain Res* **1470**, 130-144, doi:10.1016/j.brainres.2012.06.020 (2012).
61 Holroyd, S., Reiss, A. L. & Bryan, R. N. Autistic features in Joubert syndrome: a genetic disorder with agenesis of the cerebellar vermis. *Biol Psychiatry* **29**, 287-294 (1991).

62 Kim, H., Lim, C. S. & Kaang, B. K. Neuronal mechanisms and circuits underlying repetitive behaviors in mouse models of autism spectrum disorder. *Behav Brain Funct* **12**, 3, doi:10.1186/s12993-016-0087-y (2016).
63 Wang, C., Pan, Y. H., Wang, Y., Blatt, G. & Yuan, X. B. Segregated expressions of autism risk genes Cdh11 and Cdh9 in autism-relevant regions of developing cerebellum. *Mol Brain* **12**, 40, doi:10.1186/s13041-019-0461-4 (2019).
64 Gadow, K. D., Devincent, C. J., Pomeroy, J. & Azizian, A. Comparison of DSM-IV symptoms in elementary school-age children with PDD versus clinic and community samples. *Autism* **9**, 392-415, doi:10.1177/1362361305056079 (2005).
65 Simonoff, E. *et al.* Psychiatric disorders in children with autism spectrum disorders: prevalence, comorbidity, and associated factors in a population-derived sample. *J Am Acad Child Adolesc Psychiatry* **47**, 921-929, doi:10.1097/CHI.0b013e318179964f (2008).
66 Greene, R. W. *et al.* Toward a new psychometric definition of social disability in children with attention-deficit hyperactivity disorder. *J Am Acad Child Adolesc Psychiatry* **35**, 571-578, doi:10.1097/00004583-199605000-00011 (1996).
67 Grzadzinski, R., Dick, C., Lord, C. & Bishop, S. Parent-reported and clinician-observed autism spectrum disorder (ASD) symptoms in children with attention deficit/hyperactivity disorder (ADHD): implications for practice under DSM-5. *Mol Autism* **7**, 7, doi:10.1186/s13229-016-0072-1 (2016).
68 Williams, M. E. *et al.* Cadherin-9 regulates synapse-specific differentiation in the developing hippocampus. *Neuron* **71**, 640-655, doi:10.1016/j.neuron.2011.06.019 (2011).
69 Duan, X., Krishnaswamy, A., De la Huerta, I. & Sanes, J. R. Type II cadherins guide assembly of a direction-selective retinal circuit. *Cell* **158**, 793-807, doi:10.1016/j.cell.2014.06.047 (2014).
70 Wang, K. *et al.* Common genetic variants on 5p14.1 associate with autism spectrum disorders. *Nature* **459**, 528-533, doi:10.1038/nature07999 (2009).
71 Ma, D. *et al.* A genome-wide association study of autism reveals a common novel risk locus at 5p14.1. *Ann Hum Genet* **73**, 263-273, doi:10.1111/j.1469-1809.2009.00523.x (2009).
72 Kerin, T. *et al.* A noncoding RNA antisense to moesin at 5p14.1 in autism. *Sci Transl Med* **4**, 128ra140, doi:10.1126/scitranslmed.3003479 (2012).
73 DeWitt, J. J. *et al.* Impact of the Autism-Associated Long Noncoding RNA MSNP1AS on Neuronal Architecture and Gene Expression in Human Neural Progenitor Cells. *Genes (Basel)* **7**, doi:10.3390/genes7100076 (2016).
74 DeWitt, J. J. *et al.* Transcriptional Gene Silencing of the Autism-Associated Long Noncoding RNA MSNP1AS in Human Neural Progenitor Cells. *Dev Neurosci* **38**, 375-383, doi:10.1159/000453258 (2016).
75 Sanders, S. J. *et al.* Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. *Neuron* **87**, 1215-1233, doi:10.1016/j.neuron.2015.09.016 (2015).
76 Horikawa, K., Radice, G., Takeichi, M. & Chisaka, O. Adhesive subdivisions intrinsic to the epithelial somites. *Dev Biol* **215**, 182-189, doi:10.1006/dbio.1999.9463 (1999).

Figure Legends

Figure 1. Effect of gene features on genome-wide gene co-expression profiles. A total of 15,942 genes with information on all 4 gene features were identified, placed in ***Figure 1—source data 1*** as gene lists, and used for analyses. **a**: Heatmaps of CCs of genome-wide gene pairs. Genes were ranked according to mRNA abundance, gDNA size, mRNA size, or GC content. The correlation coefficient (CC) of each gene with all genes was plotted and displayed in pseudo color-coded matrices. **b**: Genome-wide distribution of CECs of each gene with the hcASD gene set under four different gene ranking conditions. In each of the 4 matrix panels, the 15,942 genes were placed in ascending order on the x-axis with 1 being the lowest mRNA abundance and GC content or shortest in gDNA or mRNA size and 15,942 being the highest in mRNA abundance and GC content or the longest in gDNA and mRNA size. Each blue dot represents the CEC of a gene with the hcASD gene set. Purple, yellow, green, and red dots represent noise-reduced (average) CEC of a gene with its neighboring 20 (10 above and 10 below; ±10), 50 (±25), 100 (±50), or 200 (±100) genes on the gene lists, respectively. Rods at the bottom of each panel show locations of hcASD genes on the ranked gene lists. **c**: Genome-wide distribution of CECs of each gene with the hcASD gene set when genes are placed in stochastic (random) orders. **d**: Comparison of distribution curve of noise-reduced CECs of the hcASD gene set with that of the top, middle, and bottom gene sets (corresponding to highest, medium, and lowest in mRNA abundance and GC content or longest, medium, and shortest in mRNA and gDNA length) of equal number (64) of genes on the ranked gene lists under different gene ranking conditions. Colored rods show the locations of top, middle, and bottom genes on the ranked gene lists.

Figure 1—figure supplement 1. Distribution of hcASDs in gene rank matrices. Genes in the whole genome were ranked by gDNA size, mRNA size, GC content (horizontal axis), and mRNA abundance level (vertical axis) and plotted in matrices. Each blue dot

represents a single gene, and each red dot represents an hcASD gene. Values of both horizontal and vertical axes are ranks (orders) of the 15,942 genes under 4 different gene ranking conditions.

Figure 1—figure supplement 2. Fitting (regression) of CEC distribution curves under four different gene ranking conditions.

Figure 1—figure supplement 3. CEC distribution curves for different developmental stages, gender, and brain regions.

Figure 1—source data 1. List of total genes and hcASDs included in this study.

Figure 1—source data 2. Parameters of fitted curves (regression curves) in figure supplement 2.

Figure 1—source data 3. List of Top, Middle, and Bottom genes under each of the four different gene ranking conditions.

Figure 2. Convergent expression of hcASD genes determined by MGCA. **a**: Comparison of noise-reduced CEC distribution curves between the hcASD gene set and 200 matched random gene sets (mRand) under different gene ranking conditions. X-axis represents gene ranks. **b**: CECs of hcASD-hcASD, hcASD-mRand, mRand-mRand, hcASD-Rand, and Rand-Rand gene set pairs. 200 each of mRand and Rand gene sets were analyzed. Box plots show ranges of CECs of hcASD-mRand, mRand-mRand, hcASD-Rand, and Rand-Rand gene set pairs. In each box plot, the central rectangles span the first quartile to the third quartile of 200 ranked CEC values. The white bar inside the rectangle shows the median CEC value, and whiskers above and below the box show the maximum and minimum values, respectively. The dotted line represents the CEC among hcASDs (hcASD-hcASD) in each panel. Three statistical methods were used to determine whether the CEC of hcASD-hcASD is significantly higher than that of hcASD-mRand, mRand-mRand, hcASD-Rand, and Rand-Rand. Upper fences test: red triangles stand for the boundaries of significant difference (1.5 x fence). Grubbs' test: * $P < 0.05$, *** $P < 0.001$. Permutation test: ### $P < 0.001$. Student's t-test was used to determine whether the CECs of hcASD-mRand and mRand-mRand are significantly greater than those of hcASD-Rand and Rand-Rand, respectively. @@@ $P < 0.001$.

Figure 2—figure supplement 1. Co-expression of hcASDs in the brain of different developmental stages and gender. **a** and **b** show data from early (8PCW-2Y) and late (3Y-40Y) developmental stages, respectively. **c** and **d** show data from male and female brain tissues, respectively.

Figure 2—figure supplement 2. Co-expression of hcASDs in different brain regions. **a-d:** Box plots show the range of CECs of hcASD-mRand in different brain regions. **e:** Significant scores of co-expression of hcASDs in different brain regions. Score = N-R. N: number of brain regions. R: integrated ranking (average of 4 conditions) of -Log (*P*-value) in the Grubbs' test under four different gene ranking conditions.

Figure 3. Comparison of TetraM-2515 and Top-2515 genes. **a**: A total of 2515 genes significantly co-expressed with hcASDs under all four matched conditions were identified by MGCA. **b**: Three-dimensional distribution of TetraM-2515 and Top-2515 genes based on the CEC with the hcASD gene set, mRNA size (bp), and mRNA abundance (RPKM) of each gene. **c**: Gene Ontology (GO) analysis of TetraM-2515 and Top-2515 genes. **d**: Integrated gene enrichment network of TetraM-only, Top-only, Overlapped genes, and hcASDs. Each node (pie) represents a GO term with genes mapped to the term. Different slices in each pie represent different gene sets, coded by different colors, and the area of each slice is proportional to the number of genes belonging to the represented gene set. 20 networks were identified by Metascape based on the significance and similarity of each node. Molecular pathways relevant to each network are listed on the right.

Figure 3—figure supplement 1. Enrichment of TetraM-2515 genes and Top-2515 genes in ASD-related pathways. **a**: Clustering analysis of –Log(p) values in the GO analysis of TetraM-only, Top-only, Overlapped, and combined ASD gene sets (cASDs). –Log(p) values (0-20) are color coded and shown in the heatmap. **b**: Integrated gene enrichment network of cASDs, TetraM-only, Overlapped, and Top-only genes.

Figure 3—source data 1. The *p*-value and FDR of each gene under different gene ranking conditions and lists of TetraM-2515, Top-2515, TetraM-2515-only, and Top-2515-only genes.

Figure 3—source data 2. GO entries of TetraM-2515 and Top-2515 genes.

Figure 3—source data 3. Node and cluster information in integrated GO enrichment networks of hcASDs, TetraM-only, Top-only, and Overlapped gene sets.
Figure 3—source data 4. Lists of 10 groups of ASD genes and cASDs.
Figure 3—source data 5. Node and cluster information of integrated GO enrichment networks of cASDs, TetraM-only, Top-only, and Overlapped gene sets.
Figure 3—source data 6. List of TetraM- and TetraN-cadherins.

Figure 4. Open field and elevated plus maze tests of *Cdh11* and *Cdh9* mutant mice. **a**: Heatmaps showing cumulated frequency of locations visited by *Cdh11* KO, heterozygote (Het), and WT mice in the open field arena. **b, c, and d**: Moving distance, moving speed, and center exploration time of *Cdh11* KO mice. **e, f, and g**: Moving distance, moving speed, and center exploration time of *Cdh9* KO mice. (male *Cdh11* KO: n=21, Het: n=22, WT: n=14; female *Cdh11* KO: n=21, Het: n=22, WT: n=21; male *Cdh9* KO: n=14, Het: n=15, WT: n=12; female *Cdh9* KO: n=8, Het: n=15, WT: n=12). **j and k**: Time spent in open arm and open arm entry frequency of *Cdh11* KO mice. **h and i**: Time spent in open arm and open arm entry frequency of *Cdh9* KO mice (male *Cdh11* KO: n=14, Het: n=14, WT: n=8; female *Cdh11* KO: n=15, Het n=14, WT n=17; male *Cdh9* KO: n=13, Het: n=15, WT: n=10; female *Cdh9* KO: n=9, Het: n=9, WT: n=10). Data are Mean ± SEM. Statistical difference was determined by one-way ANOVA followed by Student's *t*-test. * $P < 0.05$, ** $P < 0.01$.

Figure 4. Open field and elevated maze tests of *Cdh11* and *Cdh9* mutant mice. **a**: Heatmaps showing cumulated frequency of locations visited by *Cdh11* KO, heterozygote (Het), and WT mice in the open field arena. **b, c, and d**: Moving distance, moving speed, and center exploration time of *Cdh11* KO mice. **e, f, and g**: Moving distance, moving speed, and center exploration time of *Cdh9* KO mice. (male *Cdh11* KO: n=21, Het: n=22, WT: n=14; female *Cdh11* KO: n=21, Het: n=22, WT: n=21; male *Cdh9* KO: n=14, Het: n=15, WT: n=12; female *Cdh9* KO: n=8, Het: n=15, WT: n=12). **j and k**: Time spent in open arm and open arm entry frequency of *Cdh11* KO mice. **h and i**: Time spent in open arm and open arm entry frequency of *Cdh9* KO mice (male *Cdh11* KO: n=14, Het: n=14, WT: n=8; female *Cdh11* KO: n=15, Het n=14, WT n=17; male

Cdh9 KO: n=13, Het: n=15, WT: n=10; female *Cdh9* KO: n=9, Het: n=9, WT: n=10). Data are Mean ± SEM. Statistical difference was determined by one-way ANOVA followed by Student's *t*-test. * $P < 0.05$, ** $P < 0.01$.

Figure 5. Gripping strength and repetitive behaviors of *Cdh11* and *Cdh9* mutant mice. **a** and **b**: Results of gripping test and horizontal bar test for *Cdh11* KO mice (male *Cdh11* KO: n=21, Het: n=23, WT: n=14; female *Cdh11* KO: n=12, Het: n=14, WT: n=12). **c**: Results of gripping test for *Cdh9* KO mice (male *Cdh9* KO mice n=11, male *Cdh9* Het mice n=11, male *Cdh9* WT mice n=7, female *Cdh9* KO mice n=5, female *Cdh9* Het mice n=5, female *Cdh9* WT mice n=4). **d-g**: Latency to fall (**d, f**) and maximum durable speed (**e, g**) in rotarod test for female *Cdh11* and *Cdh9* mutant mice (*Cdh11* KO: n=14, Het: n=14, WT: n=17; *Cdh9* KO: n=10, Het: n=12; WT: n=9). Numbers below the X-axis (1-5) represent different trials of tests. **h-k**: Frequency and duration of self-grooming of female *Cdh11* mutant mice during the first (stage 1, **h, i**) and the second (stage 2, **j, k**) 10 min in the open field arena (*Cdh11* KO: n=9, Het: n=7, WT n=9). **l** and **m,** Frequency and duration of self-grooming of female *Cdh9* mutant mice during the first (stage 1) and second (stage 2) 10 min in the open filed arena (*Cdh9* KO: n=11, Het: n=17, WT: n=14). Data are Mean ± SEM. Statistical difference was determined by one-way ANOVA followed by Student's *t*-test. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, compared to WT littermates. # $P < 0.05$, ## $P < 0.01$, compared to mice of the same genotype.

Figure 6. Modified three-chamber test of female *Cdh11* and *Cdh9* mutant mice. **a**: Schematics of standard and modified three-chamber tests. **b** and **c**: Results of sociability and social novelty preference tests of *Cdh11* mutant mice (*Cdh11* KO: n=9, Het: n=8; WT: n=9). **d** and **e**: Results of sociability and novelty preference tests of *Cdh9* mutant mice (*Cdh9* KO: n=13, Het: n=5, WT: n=10). Data are Mean ± SEM. Statistical difference was determined by one-way ANOVA followed by Student's *t*-test. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, compared to the duration spent in the other side chamber. # $P < 0.05$, compared to WT littermates.

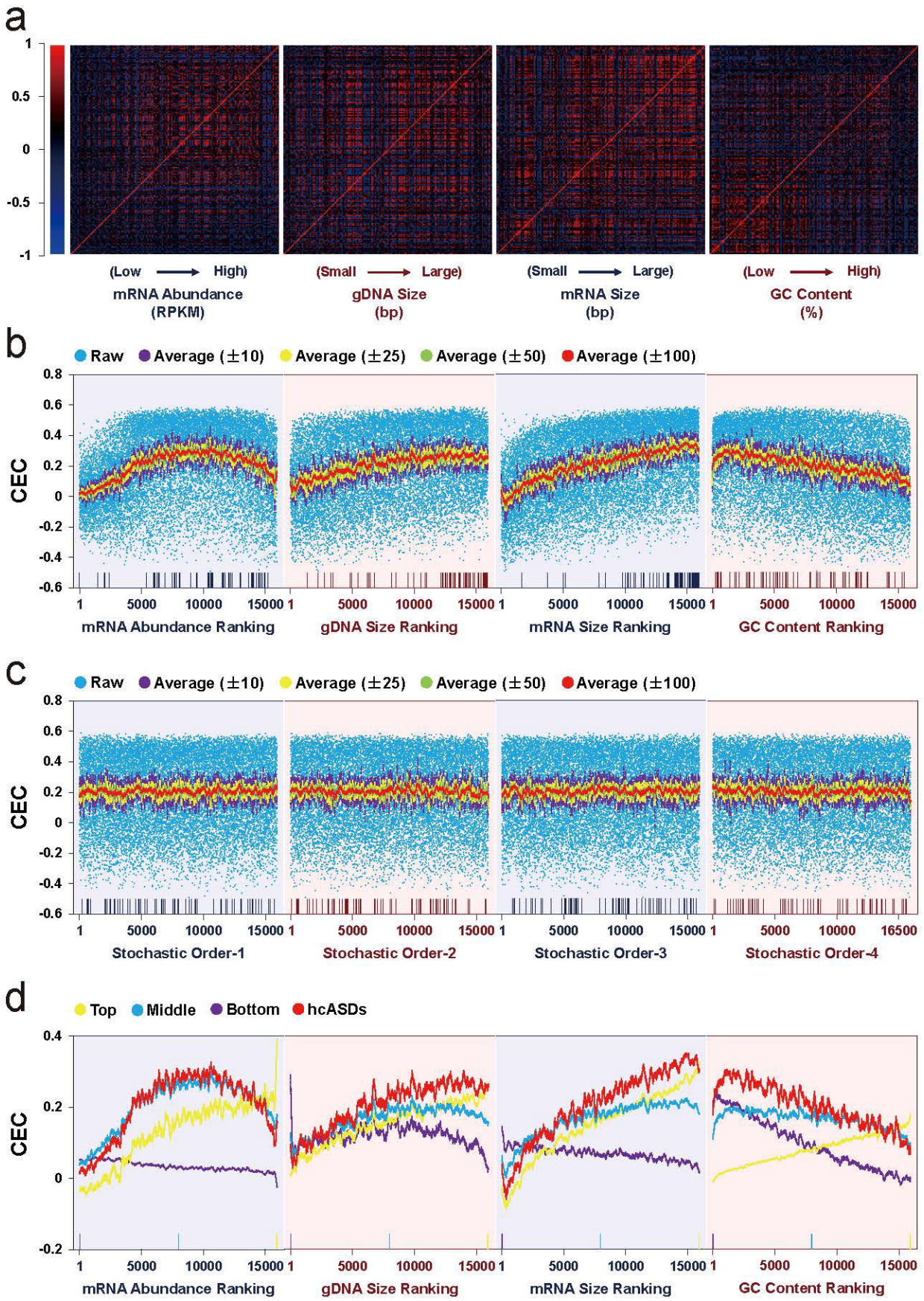
a
1
0.5
0
-0.5
-1
(Low → High)
mRNA Abundance
(RPKM)
(Small → Large)
gDNA Size
(bp)
(Small → Large)
mRNA Size
(bp)
(Low → High)
GC Content
(%)
b
Raw
Average (±10)
Average (±25)
Average (±50)
Average (±100)
CEC
mRNA Abundance Ranking
gDNA Size Ranking
mRNA Size Ranking
GC Content Ranking
c
Raw
Average (±10)
Average (±25)
Average (±50)
Average (±100)
CEC
Stochastic Order-1
Stochastic Order-2
Stochastic Order-3
Stochastic Order-4
d
Top
Middle
Bottom
hcASDs
CEC
mRNA Abundance Ranking
gDNA Size Ranking
mRNA Size Ranking
GC Content Ranking

Fig 1 Wu et al.

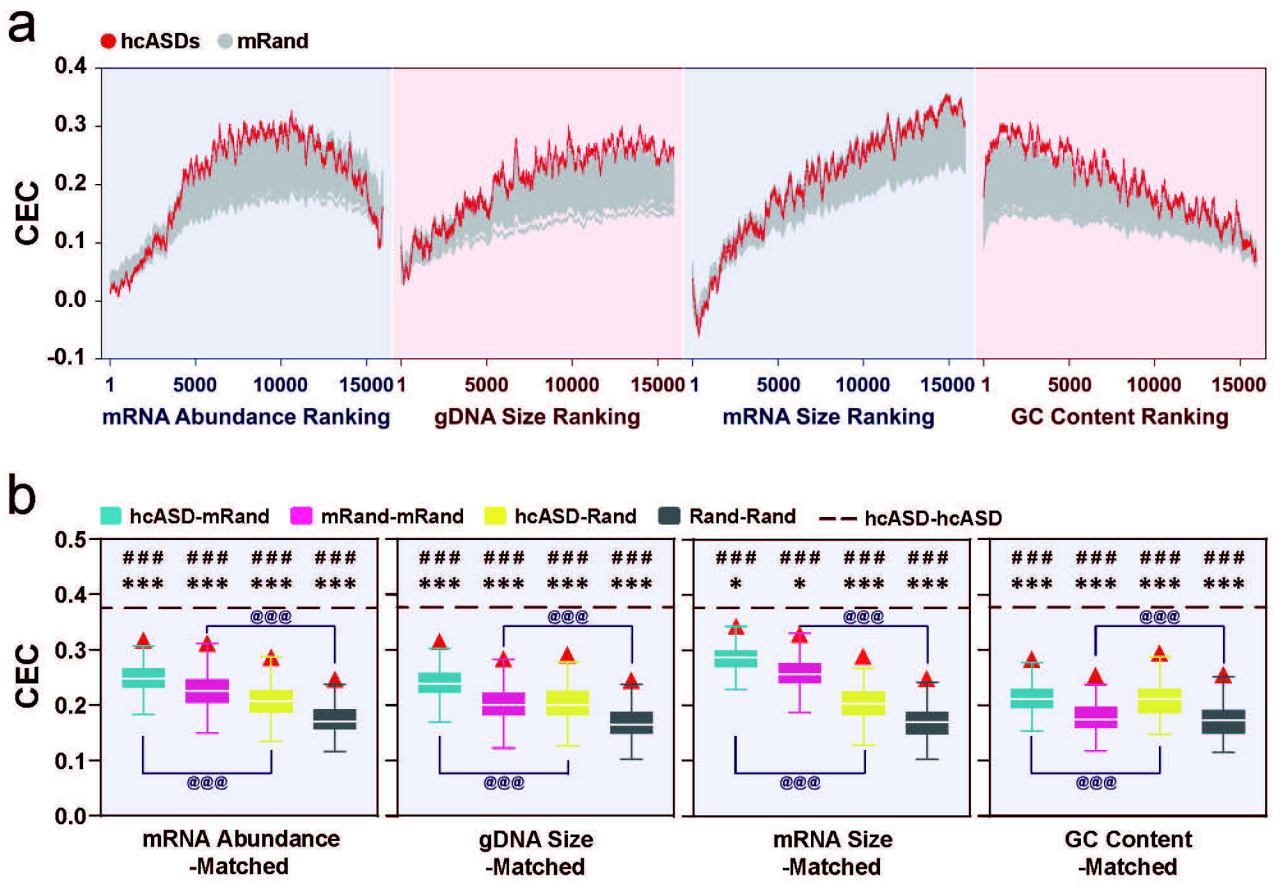
a
hcASDs
mRand
CEC
0.4
0.3
0.2
0.1
0.0
-0.1
1
5000
10000
15000
mRNA Abundance Ranking
gDNA Size Ranking
mRNA Size Ranking
GC Content Ranking
b
hcASD-mRand
mRand-mRand
hcASD-Rand
Rand-Rand
hcASD-hcASD
0.5
0.4
0.3
0.2
0.1
0.0
###
***
*
@@@
mRNA Abundance -Matched
gDNA Size -Matched
mRNA Size -Matched
GC Content -Matched

Fig 2 Wu et al.

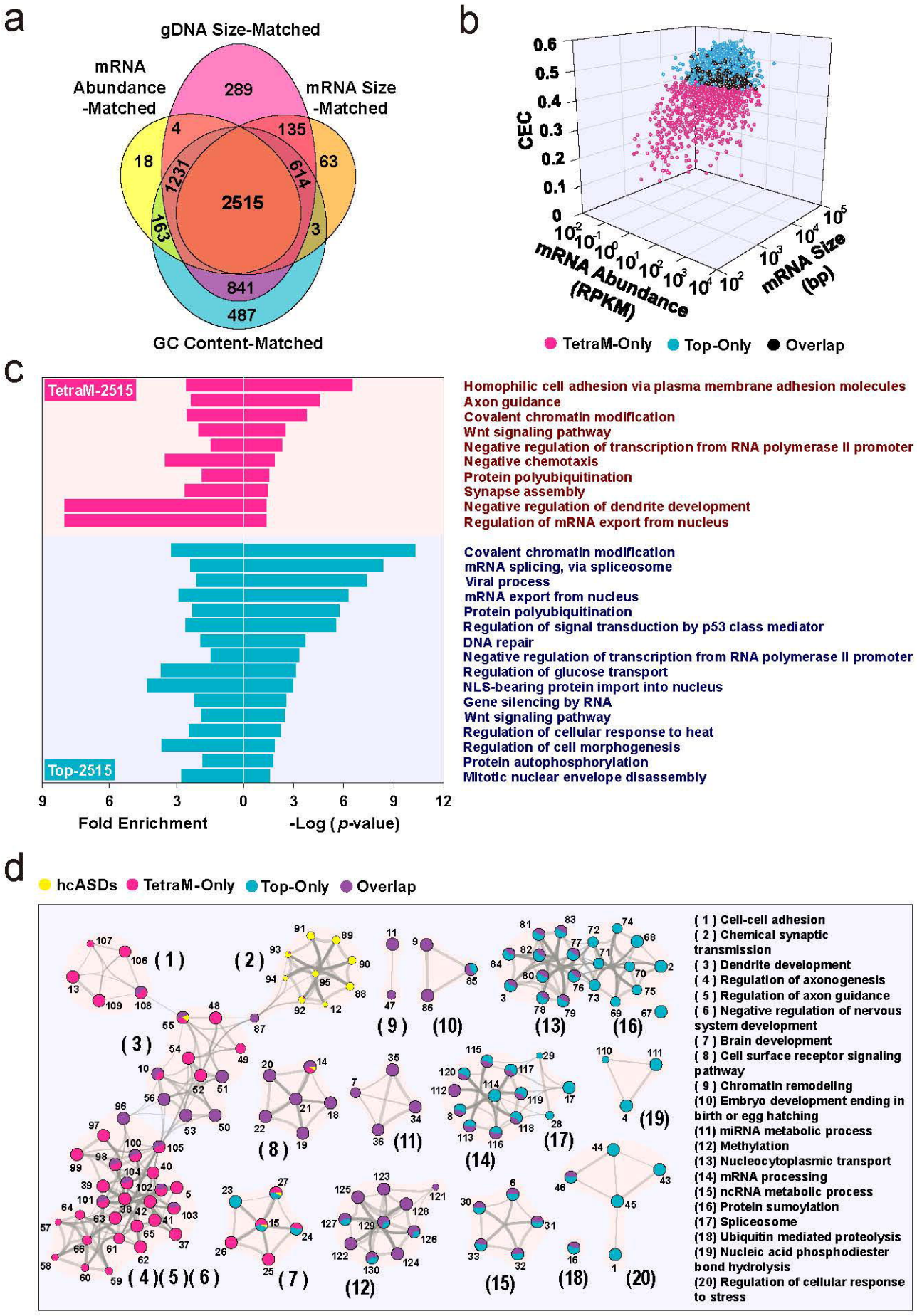
a
gDNA Size-Matched
mRNA Abundance -Matched
mRNA Size -Matched
GC Content-Matched
289
4
135
18
63
1231
614
2515
163
3
841
487
b
CEC
mRNA Abundance (RPKM)
mRNA Size (bp)
TetraM-Only
Top-Only
Overlap
c
TetraM-2515
Top-2515
Homophilic cell adhesion via plasma membrane adhesion molecules
Axon guidance
Covalent chromatin modification
Wnt signaling pathway
Negative regulation of transcription from RNA polymerase II promoter
Negative chemotaxis
Protein polyubiquitination
Synapse assembly
Negative regulation of dendrite development
Regulation of mRNA export from nucleus
Covalent chromatin modification
mRNA splicing, via spliceosome
Viral process
mRNA export from nucleus
Protein polyubiquitination
Regulation of signal transduction by p53 class mediator
DNA repair
Negative regulation of transcription from RNA polymerase II promoter
Regulation of glucose transport
NLS-bearing protein import into nucleus
Gene silencing by RNA
Wnt signaling pathway
Regulation of cellular response to heat
Regulation of cell morphogenesis
Protein autophosphorylation
Mitotic nuclear envelope disassembly
Fold Enrichment
-Log (p-value)
d
hcASDs
TetraM-Only
Top-Only
Overlap
(1) Cell-cell adhesion
(2) Chemical synaptic transmission
(3) Dendrite development
(4) Regulation of axonogenesis
(5) Regulation of axon guidance
(6) Negative regulation of nervous system development
(7) Brain development
(8) Cell surface receptor signaling pathway
(9) Chromatin remodeling
(10) Embryo development ending in birth or egg hatching
(11) miRNA metabolic process
(12) Methylation
(13) Nucleocytoplasmic transport
(14) mRNA processing
(15) ncRNA metabolic process
(16) Protein sumoylation
(17) Spliceosome
(18) Ubiquitin mediated proteolysis
(19) Nucleic acid phosphodiester bond hydrolysis
(20) Regulation of cellular response to stress

Fig 3 Wu et al.

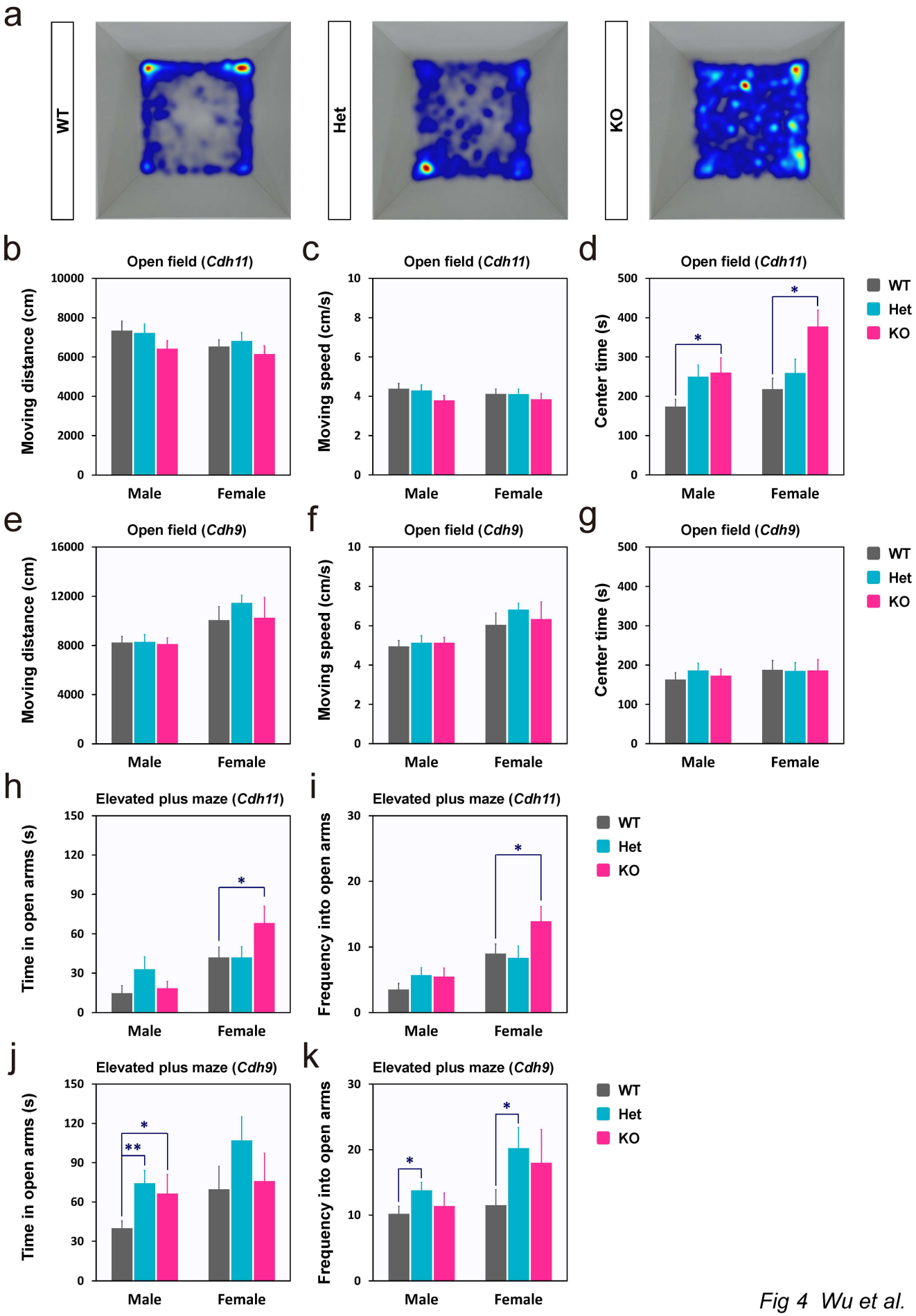
a
WT
Het
KO
b
Open field (Cdh11)
Moving distance (cm)
Male
Female
c
Open field (Cdh11)
Moving speed (cm/s)
d
Open field (Cdh11)
Center time (s)
WT
Het
KO
e
Open field (Cdh9)
f
Open field (Cdh9)
g
Open field (Cdh9)
h
Elevated plus maze (Cdh11)
Time in open arms (s)
i
Elevated plus maze (Cdh11)
Frequency into open arms
j
Elevated plus maze (Cdh9)
k
Elevated plus maze (Cdh9)

Fig 4 Wu et al.

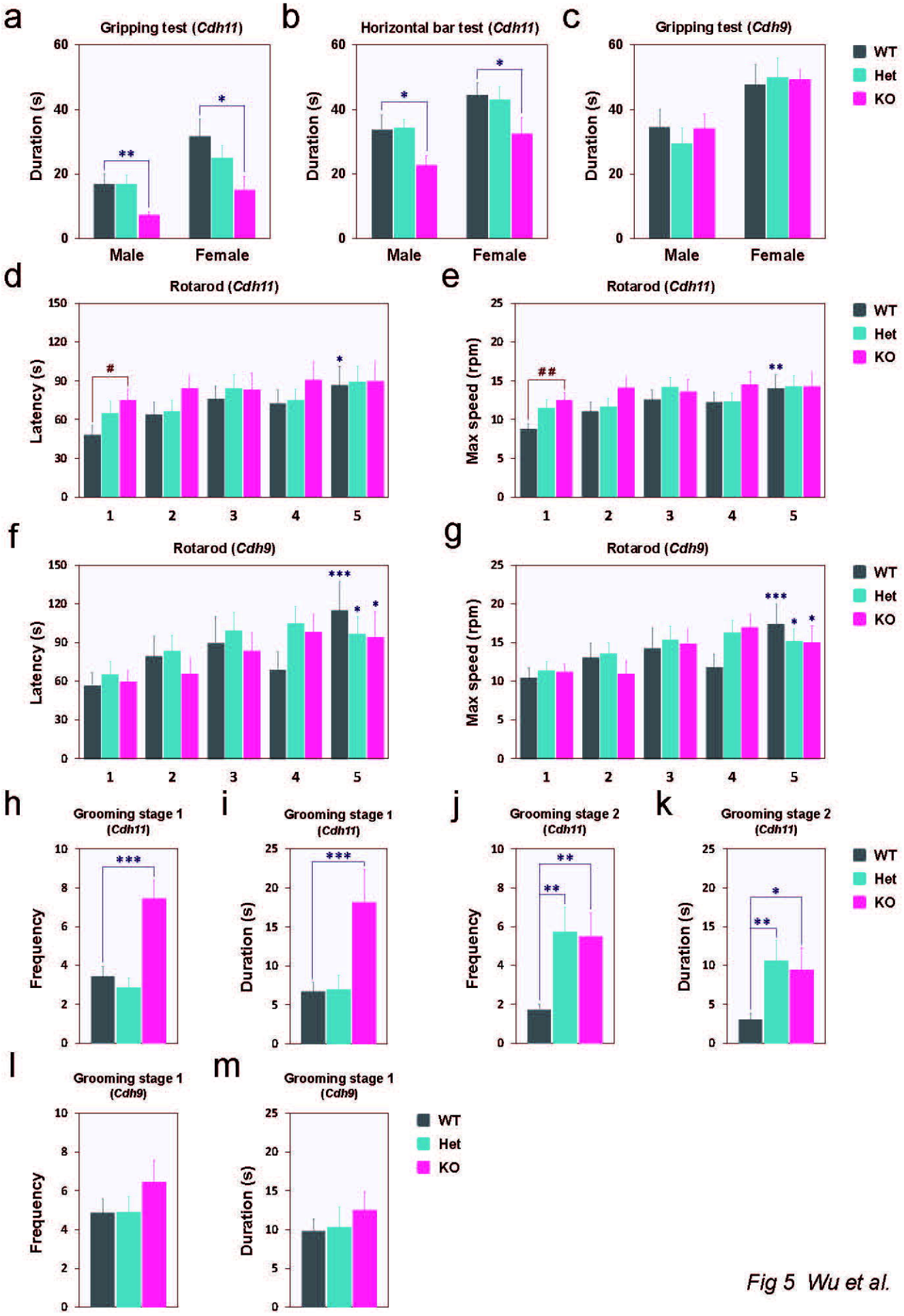
a
Gripping test (Cdh11)
Duration (s)
Male
Female
b
Horizontal bar test (Cdh11)
c
Gripping test (Cdh9)
WT
Het
KO
d
Rotarod (Cdh11)
Latency (s)
e
Max speed (rpm)
f
Rotarod (Cdh9)
g
h
Grooming stage 1 (Cdh11)
Frequency
i
Grooming stage 1 (Cdh11)
j
Grooming stage 2 (Cdh11)
k
Grooming stage 2 (Cdh11)
l
Grooming stage 1 (Cdh9)
m
Grooming stage 1 (Cdh9)

Fig 5 Wu et al.

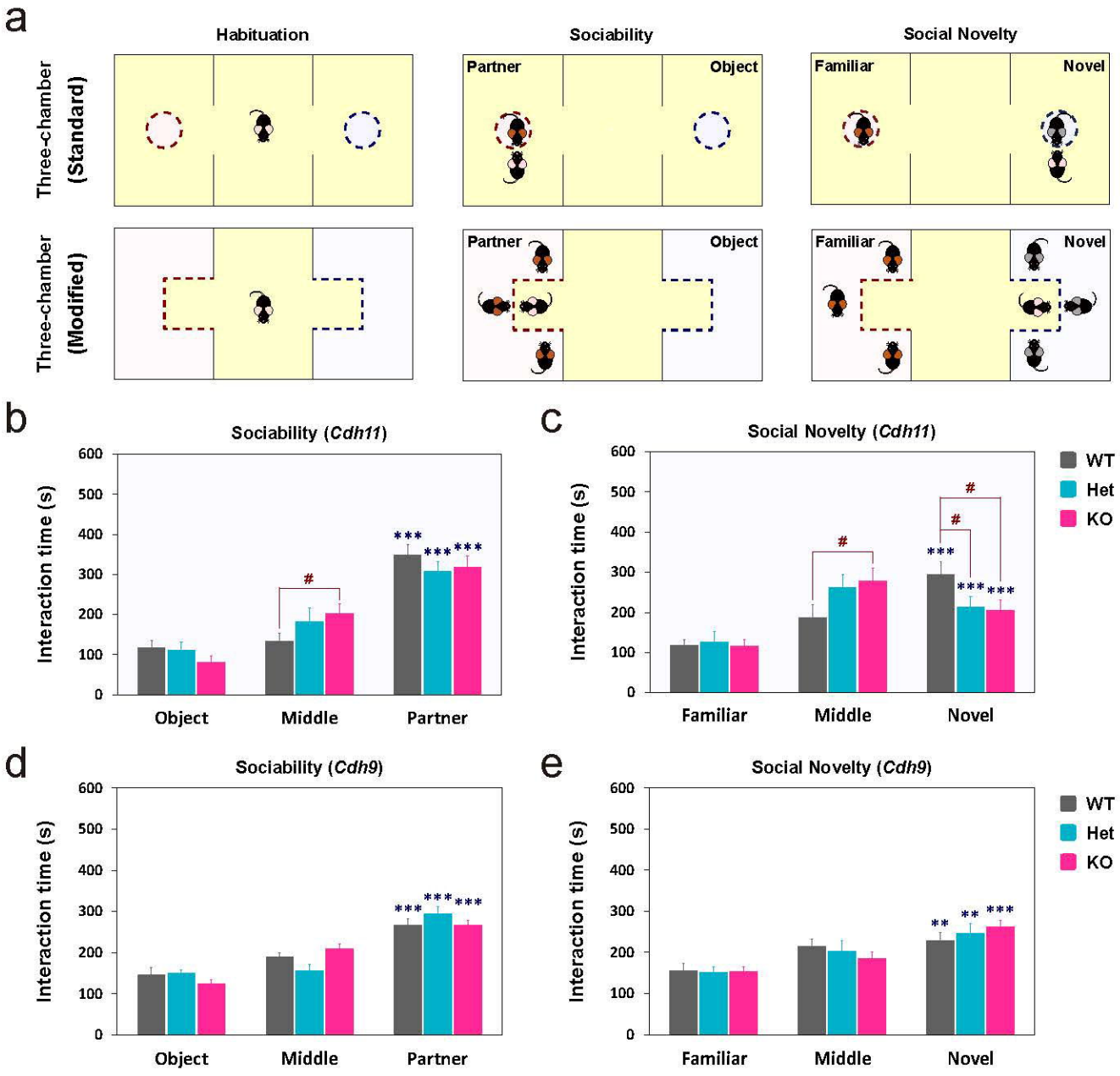
a
Habituation
Sociability
Social Novelty
Three-chamber (Standard)
Three-chamber (Modified)
Partner
Object
Familiar
Novel
b
Sociability (Cdh11)
Interaction time (s)
Object
Middle
Partner
c
Social Novelty (Cdh11)
Familiar
Middle
Novel
WT
Het
KO
d
Sociability (Cdh9)
e
Social Novelty (Cdh9)

Fig 6 Wu et al.