Identification of *CDH11* as an ASD Risk Gene by Matched-gene Co-expression Analysis and Mouse Behavioral Studies

Nan Wu
East China Normal University
Yue Wang
Hussman Institute for Autism
Jing-Yan Jia
East China Normal University
Yi-Hsuan Pan
East China Normal University https://orcid.org/0000-0001-5156-9887
Xiao-Bing Yuan (xbyuan@brain.ecnu.edu.cn)
East China Normal University and University of Maryland School of Medicine https://orcid.org/0000-0002-1632-8460

Research

Keywords: CDH11, ASD, Gene co-expression analysis, matched-gene co-expression analysis

Posted Date: October 14th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-90298/v1

Abstract

Background: Gene co-expression analysis (GCA) has emerged as an important tool to identify convergent molecular pathways of ASD risk genes. The aim of this study is to identify ASD-relevant genes at the whole-genome level using GCA with consideration of the effect of confounding factors on GCA, including the size, expression level, and guanine-cytosine content of genes.

Methods: Pearson's correlation coefficient was computed to indicate the co-expression of a gene pair based on the BrainSpan human brain transcriptome dataset. Whether a gene is significantly co-expressed with a group of high-confidence ASD risk genes (hcASDs) was determined by statistically comparing the co-expression of this gene with the hcASD gene set to that of this gene with permuted gene sets of matched gene features. This method is referred to as "matched-gene co-expression analysis" (MGCA). Gene ontology (GO) analysis and construction of integrated GO enrichment networks were performed to reveal convergent pathways of co-expressed genes. Behavioral tests were carried out in gene knockout mice.

Results: Gene size, mRNA length, mRNA abundance, and guanine-cytosine content were found to affect co-expression profiles of ASD genes. Using the MGCA method, we confirmed the convergence in the developmental expression profiles of hcASDs. MGCA also effectively revealed convergent molecular pathways of ASD risk genes and determined that *CDH11*, but not *CDH9*, is associated with ASD. Mouse behavioral studies showed that *Cdh11*-null mice, but not *Cdh9*-null mice, have multiple autistic-like behavioral alterations.

Limitations: The use of tissue-derived transcriptomes instead of single-cell transcriptomes may have detected coincident expression of some functionally irrelevant genes in different cell types. Some ASD risk genes may have been missed due to the highly stringent statistical standard of MGCA. Another limitation is the relatively small number of animals analyzed in behavioral tests.

Conclusions: Results of this study revealed the importance of considering matched gene features in GCA. CDH11 was confirmed to be an important ASD risk gene and *Cdh11*-null mice were found to be a very useful animal model for investigation of ASD.

Background

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 to determine 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 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 crucial 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 effective in identifying functionally convergent molecular pathways of ASD risk genes. MGCA also revealed “homophilic cell adhesion” as one of the most significantly converged molecular pathways of

risk genes. Further analysis of CDH11 and CDH9, two ASD candidate genes belonging to the cadherin family of adhesion molecules, showed that *CDH11*, but not *CDH9*, is associated with ASD. This finding was corroborated by mouse behavioral studies using *Cdh11* and *Cdh9* knockout mice.

Materials And Methods

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 [25], 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, **Additional file 2: Table S1**). Genes with an average expression level lower than the lowest expression level of hcASDs were filtered out (**Additional file 2: 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 (**Additional file 2: Table S1**). 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

Fig. 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 (**Additional file 1: Fig. S6a**), 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 (**Fig. 3d; Additional file1: Fig. S6b**), 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 [26] were provided by Dr. Joshua R. Sanes at Harvard University. *Cdh11* KO mice [27] 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 [26]. 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^{o}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^{o}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 the R software package. For the permutation test, 100,000 mRand or Rand gene sets were used; hcASDs were not excluded from the permuted gene sets. 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. 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).

Results

Effects of gene features on gene co-expression profiles

The potential effect of the four gene features, including mRNA abundance, mRNA size, gDNA size, and GC content of the coding region, on gene co-expression profiles was 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 (**Additional file 2: Table S1**). 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 (**Fig. 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 (**Fig. 1a**). This result suggests that all four gene features 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 (**Additional file1: Fig. S1**). 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 **Fig. 1b; Additional file 2: Table S1**). In each of the 4 panels (**Fig. 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 (**Fig. 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 (**Fig. 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 (**Fig. 1b**, right panel; **Additional file 2: Table S1**).

With cubic regression, each noise-reduced CEC distribution curve was found to have an R^2 value > 0.9 (**Additional file 1: Fig. S2;Additional file 3: Table S2**), 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 (**Fig. 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 (**Additional file 1: Fig. S3a**), both gender, and different brain regions (**Additional file 1: Fig. S3b, 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 mRNA abundance and GC content or largest, medium, and smallest mRNA and gDNA sizes (**Additional file 4: Table S3; Fig. 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 **Fig. 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 **Fig. 1d**) were positively correlated with mRNA abundance, whereas those of the bottom mRNA abundance gene set (Bottom in **Fig. 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 **Fig. 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 (**Fig. 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 (**Fig. 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 **Fig. 2b**), suggesting that genes of similar features tend to co-express with each other. The CEC of hcASD-hcASD (dashed line in **Fig. 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 this tendency (*** in **Fig. 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 **Fig. 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 (**Additional file 1: Fig. S4a, b**), both gender (**Additional file 1: Fig. S4c, d**), and different brain regions (**Additional file 1: Fig. S5a-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 (**Additional file 1: Fig. S5e**). 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 [28-34]. 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 (**Fig. 3a; Additional file 5: Table S4**), with a false discovery rate below 5% (FDR < 0.05).

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 (**Fig. 3a; Additional file 5: Table S4**). 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, **Additional file 5: Table S4**). 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 (**Fig. 3b; Additional file 5: Table S4**). 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 (**Fig. 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 (**Fig. 3c; Additional file 6: Table S5**).

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 [35]. 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 (**Fig. 3d;Additional file 7: Table S6**). 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 (**Fig. 3d; Additional file 7: Table S6**), 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, 25, 36-41] (**Additional file 8: Table S7**). 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 (**Additional file 1: Fig. S6a**). 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. (**Additional file 1: Fig. S6a; Additional file 9: Table S8**). More co-localized nodes of cASD and TetreM-only genes than those of cASD and Top-only genes were seen in the network graph (**Additional file 1: Fig. S6b;Additional file 9: Table S8**). 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 (**Fig. 3c, d; Additional file 6: Table S5; Additional file 7: Table S6**). Some cadherin family members, such as CDH2, in the TetraM-2515 gene set are known to be high risk genes of ASD (**Additional file 10: Table S9**) 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; **Additional file 10: Table S9**). Several recent genetic studies have implicated two type II cadherins, *CDH11* and *CDH9*, in ASD and other psychiatric diseases [42-46]. 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 (**Fig. 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 (**Fig. 4b, c**). Both male and female *Cdh9*-null mice were largely normal in the OFT (**Fig. 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 (**Fig. 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 [47]. 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 (**Fig. 4j, k**).

Individuals with ASD often have a weaker grip strength than age-matched controls [48]. 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 (**Fig. 5a, b**), indicating reduced gripping strength and/or impaired motor coordination. The gripping strength of *Cdh9*-null mice was normal (**Fig. 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 (**Fig. 5d-g**). In subsequent trials, *Cdh11*-null mice did not display significant improvement in performance (**Fig. 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 [49].

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 **Fig. 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 (**Fig. 5j, k**), indicating elevated repetitive behavior even in a relatively familiar environment. No such behavioral alteration was observed in *Cdh9*-null mice (**Fig. 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 (**Fig. 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 (**Fig. 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 (**Fig. 6b, c**), indicating reduced sociability. In contrast, *Cdh9*-null mice did not show any abnormality in this test (**Fig. 6d, e**).

Discussion

Gene co-expression analysis (GCA) is a powerful tool to find functionally convergent genes. Several previous GCA studies had considered the potential effect of gene size and GC content, but not mRNA abundance on the co-expression of ASD risk genes [6]. 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 effect 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 (Fig. 3c, d; **Additional file 1: Fig. S6a, b**). Many of these genes are either upstream genes that regulate the transcription of hcASDs or downstream target genes whose expression is regulated by hcASDs, and thus display a 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 not realized by previous GCA studies due to a relatively low CC with other genes (**Additional file 10: Table S9**) and was excluded from the gene co-expression network. 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 [50–53]. Several cadherin family members have been implicated in ASD [16, 54–62]. For example, some protocadherins in the *FAT* cadherin subfamily were found by whole-exome sequencing to be associated with ASD [16, 54]. 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 [55]. Mutations in the *PCDH19* gene have been shown to cause early-onset epilepsy, and many individuals with these mutations also display autistic features [56–58]. 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 [63]. 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 profile 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 (**Additional file 10: Table S9**). The functions of these *PCDHB*s in the brain remain to be determined. One of such cadherins identified by MGCA is CDH11. In this study, we found that *Cdh11*-null mice had significantly increased repetitive activities. The brain regions including neocortex, CB, and STR are known to be involved in the control of repetitive behaviors [64]. 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 [65].

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) [42]. A case-control association study revealed a high prevalence of a homozygous single nucleotide variant rs7187376C/C of *CDH11* in patients with ASD [42]. Several other coding variants of *CDH11* were also discovered in ASD individuals [42]. Behavioral changes that we have observed in *Cdh11*-null mice, including reduced anxiety, increased repetitive behavior, and reduced sociability, are highly consistent with the non-syndromic ASD case with partial deletion of *CDH11* [42]. 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 [66–69]. 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, but not at a statistically significant level (Fig. 5j, k; Fig. 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 [26, 70]. Its association with ASD has been suggested by several genome-wide association studies (GWAS) [71, 72]. 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 [71]. However, this rs4307059 genotype was not correlated with the expression of either *CDH9* or *CDH10* in adult brains [71, 73], 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 contribute to ASD [73–75]. 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.

Limitations

The current study was based on the BrainSpan human brain transcriptome dataset. The advantage of using this dataset for MGCA is that it contains data of a wide variety of brain tissue. Multiple brain regions that are known to be closely relevant to ASD etiology were included and a broad range of developmental stages of each brain region were sampled in this dataset. It is a valuable data resource for investigating gene co-expression related to ASD. A major limitation of using this dataset is that the transcriptome data were from tissues of mixed cell types. Some co-expressed genes detected using this dataset may be attributed to coincident expression of functionally irrelevant genes in different cell types instead of real co-expression of genes that play coordinated functions within the same type of cells. In recent years, single-cell transcriptome analysis has emerged as a powerful technology for high-throughput analysis of various developmental and pathological processes. We expect that future GCA using single-cell transcriptomes will be an important complement to the MGCA method in the identification of ASD-relevant genes and molecular pathways.

In the present study, four different gene features were taken into consideration in MGCA, and only co-expression relationships with a permutation *p*-value less than 0.001 under all four matched conditions (tetra-matched) were considered as significant. Although such a highly stringent condition could greatly increase the accuracy of co-expression detection, some meaningful co-expression relationships may be filtered out. Since 34 of the 64 hcASDs were found in the TetraM gene set ($p < 0.001$), we estimate that MGCA detected 53% ASD-relevant genes in the whole genome.

Another limitation of the present study is the relatively small number of animals analyzed in some behavioral tests, such as the modified 3-chamble sociability test. Statistical analysis based on test results of this small animal cohort indicated impaired social interaction of *Cdh11*-null mice. It is conceivable that analysis of more animals could strengthen the conclusion of the present study.

Conclusions

Results of this study revealed the importance of considering matched gene features in the analysis of gene co-expression and demonstrated the effectiveness of MGCA in identification of ASD susceptibility genes and their convergent signaling pathways. Application of MGCA led to the identification of *CDH11* as an important ASD risk gene. Our results also showed that *Cdh11*-null mice can be used to study circuit mechanisms of a subgroup of ASD and explore therapeutic strategies for ASD.

Declarations

Ethics approval and consent to participate

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).

Consent for publication

Not applicable.

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.

Funding

This work was funded by the National Science Foundation of China (31771116, 31871501) and Hussman Foundation (HIAS15006). X.Y. was supported by the Simons Foundation (296143).

Authors' contributions

N.W., Y.W., and Y.P. analyzed data and generated figures. J.J and X.Y. conducted behavioral tests. Y.P. and X.Y. designed the study and wrote the manuscript.

Acknowledgments

We thank Dr. Chao-Hung Lee and Elizabeth Benevides for editing the manuscript.

References

1. Berg JM, Geschwind DH: **Autism genetics: searching for specificity and convergence.** *Genome Biol* 2012, **13:**247.
2. Jeste SS, Geschwind DH: **Disentangling the heterogeneity of autism spectrum disorder through genetic findings.** *Nature Reviews Neurology* 2014, **10:**74-81.
3. de la Torre-Ubieta L, Won H, Stein JL, Geschwind DH: **Advancing the understanding of autism disease mechanisms through genetics.** *Nat Med* 2016, **22:**345-361.

4. Sullivan PF, Geschwind DH: **Defining the Genetic, Genomic, Cellular, and Diagnostic Architectures of Psychiatric Disorders.** *Cell* 2019, **177:**162-183.
5. Courchesne E, Pramparo T, Gazestani VH, Lombardo MV, Pierce K, Lewis NE: **The ASD Living Biology: from cell proliferation to clinical phenotype.** *Mol Psychiatry* 2019, **24:**88-107.
6. Willsey AJ, Sanders SJ, Li M, Dong S, Tebbenkamp AT, Muhle RA, Reilly SK, Lin L, Fertuzinhos S, Miller JA, et al: **Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism.** *Cell* 2013, **155:**997-1007.
7. Parikshak NN, Luo R, Zhang A, Won H, Lowe JK, Chandran V, Horvath S, Geschwind DH: **Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism.** *Cell* 2013, **155:**1008-1021.
8. Stuart JM, Segal E, Koller D, Kim SK: **A gene-coexpression network for global discovery of conserved genetic modules.** *Science* 2003, **302:**249-255.
9. Hormozdiari F, Penn O, Borenstein E, Eichler EE: **The discovery of integrated gene networks for autism and related disorders.** *Genome Res* 2015, **25:**142-154.
10. Mahfouz A, Ziats MN, Rennert OM, Lelieveldt BP, Reinders MJ: **Shared Pathways Among Autism Candidate Genes Determined by Co-expression Network Analysis of the Developing Human Brain Transcriptome.** *J Mol Neurosci* 2015, **57:**580-594.
11. Kopp N, Climer S, Dougherty JD: **Moving from capstones toward cornerstones: successes and challenges in applying systems biology to identify mechanisms of autism spectrum disorders.** *Front Genet* 2015, **6:**301.
12. Gabel HW, Kinde B, Stroud H, Gilbert CS, Harmin DA, Kastan NR, Hemberg M, Ebert DH, Greenberg ME: **Disruption of DNA-methylation-dependent long gene repression in Rett syndrome.** *Nature* 2015, **522:**89-93.
13. Krueger DD, Tuffy LP, Papadopoulos T, Brose N: **The role of neurexins and neuroligins in the formation, maturation, and function of vertebrate synapses.** *Curr Opin Neurobiol* 2012, **22:**412-422.
14. Sudhof TC: **Neuroligins and neurexins link synaptic function to cognitive disease.** *Nature* 2008, **455:**903-911.
15. Hussman JP, Chung RH, Griswold AJ, Jaworski JM, Salyakina D, Ma D, Konidari I, Whitehead PL, Vance JM, Martin ER, et al: **A noise-reduction GWAS analysis implicates altered regulation of neurite outgrowth and guidance in autism.** *Mol Autism* 2011, **2:**1.
16. Cukier HN, Dueker ND, Slifer SH, Lee JM, Whitehead PL, Lalanne E, Leyva N, Konidari I, Gentry RC, Hulme WF, et al: **Exome sequencing of extended families with autism reveals genes shared across neurodevelopmental and neuropsychiatric disorders.** *Mol Autism* 2014, **5:**1.
17. Iossifov I, O'Roak BJ, Sanders SJ, Ronemus M, Krumm N, Levy D, Stessman HA, Witherspoon KT, Vives L, Patterson KE, et al: **The contribution of de novo coding mutations to autism spectrum disorder.** *Nature* 2014, **515:**216-221.
18. Sanders SJ, Murtha MT, Gupta AR, Murdoch JD, Raubeson MJ, Willsey AJ, Ercan-Sencicek AG, DiLullo NM, Parikshak NN, Stein JL, et al: **De novo mutations revealed by whole-exome sequencing**

are strongly associated with autism. *Nature* 2012, **485:**237-241.

19. Neale BM, Kou Y, Liu L, Ma'ayan A, Samocha KE, Sabo A, Lin CF, Stevens C, Wang LS, Makarov V, et al: **Patterns and rates of exonic de novo mutations in autism spectrum disorders.** *Nature* 2012, **485:**242-245.
20. Seong E, Yuan L, Arikkath J: **Cadherins and catenins in dendrite and synapse morphogenesis.** *Cell Adh Migr* 2015, **9:**202-213.
21. Stocker AM, Chenn A: **The role of adherens junctions in the developing neocortex.** *Cell Adh Migr* 2015, **9:**167-174.
22. Gartner A, Fornasiero EF, Dotti CG: **Cadherins as regulators of neuronal polarity.** *Cell Adh Migr* 2015, **9:**175-182.
23. Luccardini C, Hennekinne L, Viou L, Yanagida M, Murakami F, Kessaris N, Ma X, Adelstein RS, Mege RM, Metin C: **N-cadherin sustains motility and polarity of future cortical interneurons during tangential migration.** *J Neurosci* 2013, **33:**18149-18160.
24. Takeichi M: **The cadherin superfamily in neuronal connections and interactions.** *Nat Rev Neurosci* 2007, **8:**11-20.
25. Sanders SJ, He X, Willsey AJ, Ercan-Sencicek AG, Samocha KE, Cicek AE, Murtha MT, Bal VH, Bishop SL, Dong S, et al: **Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci.** *Neuron* 2015, **87:**1215-1233.
26. Duan X, Krishnaswamy A, De la Huerta I, Sanes JR: **Type II cadherins guide assembly of a direction-selective retinal circuit.** *Cell* 2014, **158:**793-807.
27. Horikawa K, Radice G, Takeichi M, Chisaka O: **Adhesive subdivisions intrinsic to the epithelial somites.** *Dev Biol* 1999, **215:**182-189.
28. Tsai PT, Hull C, Chu Y, Greene-Colozzi E, Sadowski AR, Leech JM, Steinberg J, Crawley JN, Regehr WG, Sahin M: **Autistic-like behaviour and cerebellar dysfunction in Purkinje cell Tsc1 mutant mice.** *Nature* 2012, **488:**647-651.
29. Khan AJ, Nair A, Keown CL, Datko MC, Lincoln AJ, Muller RA: **Cerebro-cerebellar Resting-State Functional Connectivity in Children and Adolescents with Autism Spectrum Disorder.** *Biol Psychiatry* 2015, **78:**625-634.
30. Buxhoeveden DP, Semendeferi K, Buckwalter J, Schenker N, Switzer R, Courchesne E: **Reduced minicolumns in the frontal cortex of patients with autism.** *Neuropathol Appl Neurobiol* 2006, **32:**483-491.
31. Green SA, Hernandez L, Tottenham N, Krasileva K, Bookheimer SY, Dapretto M: **Neurobiology of Sensory Overresponsivity in Youth With Autism Spectrum Disorders.** *JAMA Psychiatry* 2015, **72:**778-786.
32. Green SA, Rudie JD, Colich NL, Wood JJ, Shirinyan D, Hernandez L, Tottenham N, Dapretto M, Bookheimer SY: **Overreactive brain responses to sensory stimuli in youth with autism spectrum disorders.** *J Am Acad Child Adolesc Psychiatry* 2013, **52:**1158-1172.

33. Di Martino A, Kelly C, Grzadzinski R, Zuo XN, Mennes M, Mairena MA, Lord C, Castellanos FX, Milham MP: **Aberrant striatal functional connectivity in children with autism.** *Biol Psychiatry* 2011, **69:**847-856.
34. Wang X, Bey AL, Katz BM, Badea A, Kim N, David LK, Duffney LJ, Kumar S, Mague SD, Hulbert SW, et al: **Altered mGluR5-Homer scaffolds and corticostriatal connectivity in a Shank3 complete knockout model of autism.** *Nat Commun* 2016, **7:**11459.
35. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, Benner C, Chanda SK: **Metascape provides a biologist-oriented resource for the analysis of systems-level datasets.** *Nat Commun* 2019, **10:**1523.
36. Werling DM, Parikshak NN, Geschwind DH: **Gene expression in human brain implicates sexually dimorphic pathways in autism spectrum disorders.** *Nat Commun* 2016, **7:**10717.
37. Basu SN, Kollu R, Banerjee-Basu S: **AutDB: a gene reference resource for autism research.** *Nucleic Acids Res* 2009, **37:**D832-836.
38. Abrahams BS, Arking DE, Campbell DB, Mefford HC, Morrow EM, Weiss LA, Menashe I, Wadkins T, Banerjee-Basu S, Packer A: **SFARI Gene 2.0: a community-driven knowledgebase for the autism spectrum disorders (ASDs).** *Mol Autism* 2013, **4:**36.
39. Krishnan A, Zhang R, Yao V, Theesfeld CL, Wong AK, Tadych A, Volfovsky N, Packer A, Lash A, Troyanskaya OG: **Genome-wide prediction and functional characterization of the genetic basis of autism spectrum disorder.** *Nat Neurosci* 2016, **19:**1454-1462.
40. RK CY, Merico D, Bookman M, J LH, Thiruvahindrapuram B, Patel RV, Whitney J, Deflaux N, Bingham J, Wang Z, et al: **Whole genome sequencing resource identifies 18 new candidate genes for autism spectrum disorder.** *Nat Neurosci* 2017, **20:**602-611.
41. pfeliciano@simonsfoundation.org SCEa, Consortium S: **SPARK: A US Cohort of 50,000 Families to Accelerate Autism Research.** *Neuron* 2018, **97:**488-493.
42. Crepel A, De Wolf V, Brison N, Ceulemans B, Walleghem D, Peuteman G, Lambrechts D, Steyaert J, Noens I, Devriendt K, Peeters H: **Association of CDH11 with non-syndromic ASD.** *Am J Med Genet B Neuropsychiatr Genet* 2014, **165B:**391-398.
43. Gasso P, Ortiz AE, Mas S, Morer A, Calvo A, Bargallo N, Lafuente A, Lazaro L: **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* 2015, **186:**284-292.
44. Sokolowski M, Wasserman J, Wasserman D: **Polygenic associations of neurodevelopmental genes in suicide attempt.** *Mol Psychiatry* 2016, **21:**1381-1390.
45. Wang K, Zhang H, Bloss CS, Duvvuri V, Kaye W, Schork NJ, Berrettini W, Hakonarson H, Price Foundation Collaborative G: **A genome-wide association study on common SNPs and rare CNVs in anorexia nervosa.** *Mol Psychiatry* 2011, **16:**949-959.
46. Anazi S, Maddirevula S, Faqeih E, Alsedairy H, Alzahrani F, Shamseldin HE, Patel N, Hashem M, Ibrahim N, Abdulwahab F, et al: **Clinical genomics expands the morbid genome of intellectual**

disability and offers a high diagnostic yield. *Mol Psychiatry* 2017, **22:**615-624.

47. Manabe T, Togashi H, Uchida N, Suzuki SC, Hayakawa Y, Yamamoto M, Yoda H, Miyakawa T, Takeichi M, Chisaka O: **Loss of cadherin-11 adhesion receptor enhances plastic changes in hippocampal synapses and modifies behavioral responses.** *Mol Cell Neurosci* 2000, **15:**534-546.
48. Travers BG, Bigler ED, Tromp do PM, Adluru N, Destiche D, Samsin D, Froehlich A, Prigge MD, Duffield TC, Lange N, et al: **Brainstem White Matter Predicts Individual Differences in Manual Motor Difficulties and Symptom Severity in Autism.** *J Autism Dev Disord* 2015, **45:**3030-3040.
49. Rothwell PE, Fuccillo MV, Maxeiner S, Hayton SJ, Gokce O, Lim BK, Fowler SC, Malenka RC, Sudhof TC: **Autism-associated neuroligin-3 mutations commonly impair striatal circuits to boost repetitive behaviors.** *Cell* 2014, **158:**198-212.
50. Bekirov IH, Nagy V, Svoronos A, Huntley GW, Benson DL: **Cadherin-8 and N-cadherin differentially regulate pre- and postsynaptic development of the hippocampal mossy fiber pathway.** *Hippocampus* 2008, **18:**349-363.
51. Friedman LG, Riemslagh FW, Sullivan JM, Mesias R, Williams FM, Huntley GW, Benson DL: **Cadherin-8 expression, synaptic localization, and molecular control of neuronal form in prefrontal corticostriatal circuits.** *J Comp Neurol* 2015, **523:**75-92.
52. Bian WJ, Miao WY, He SJ, Qiu Z, Yu X: **Coordinated Spine Pruning and Maturation Mediated by Inter-Spine Competition for Cadherin/Catenin Complexes.** *Cell* 2015, **162:**808-822.
53. Tsai NP, Wilkerson JR, Guo W, Maksimova MA, DeMartino GN, Cowan CW, Huber KM: **Multiple autism-linked genes mediate synapse elimination via proteasomal degradation of a synaptic scaffold PSD-95.** *Cell* 2012, **151:**1581-1594.
54. Butler MG, Rafi SK, Hossain W, Stephan DA, Manzardo AM: **Whole exome sequencing in females with autism implicates novel and candidate genes.** *Int J Mol Sci* 2015, **16:**1312-1335.
55. Anitha A, Thanseem I, Nakamura K, Yamada K, Iwayama Y, Toyota T, Iwata Y, Suzuki K, Sugiyama T, Tsujii M, et al: **Protocadherin alpha (PCDHA) as a novel susceptibility gene for autism.** *J Psychiatry Neurosci* 2013, **38:**192-198.
56. Cappelletti S, Specchio N, Moavero R, Terracciano A, Trivisano M, Pontrelli G, Gentile S, Vigevano F, Cusmai R: **Cognitive development in females with PCDH19 gene-related epilepsy.** *Epilepsy Behav* 2015, **42:**36-40.
57. van Harssel JJ, Weckhuysen S, van Kempen MJ, Hardies K, Verbeek NE, de Kovel CG, Gunning WB, van Daalen E, de Jonge MV, Jansen AC, 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* 2013, **14:**23-34.
58. Camacho A, Simon R, Sanz R, Vinuela A, Martinez-Salio A, Mateos F: **Cognitive and behavioral profile in females with epilepsy with PDCH19 mutation: two novel mutations and review of the literature.** *Epilepsy Behav* 2012, **24:**134-137.
59. Sanders SJ, Ercan-Sencicek AG, Hus V, Luo R, Murtha MT, Moreno-De-Luca D, Chu SH, Moreau MP, Gupta AR, Thomson SA, et al: **Multiple recurrent de novo CNVs, including duplications of the 7q11.23**

Williams syndrome region, are strongly associated with autism. *Neuron* 2011, **70:**863-885.

60. Pagnamenta AT, Khan H, Walker S, Gerrelli D, Wing K, Bonaglia MC, Giorda R, Berney T, Mani E, Molteni M, et al: **Rare familial 16q21 microdeletions under a linkage peak implicate cadherin 8 (CDH8) in susceptibility to autism and learning disability.** *J Med Genet* 2011, **48:**48-54.
61. Moya PR, Dodman NH, Timpano KR, Rubenstein LM, Rana Z, Fried RL, Reichardt LF, Heiman GA, Tischfield JA, King RA, et al: **Rare missense neuronal cadherin gene (CDH2) variants in specific obsessive-compulsive disorder and Tourette disorder phenotypes.** *Eur J Hum Genet* 2013, **21:**850-854.
62. Redies C, Hertel N, Hubner CA: **Cadherins and neuropsychiatric disorders.** *Brain Res* 2012, **1470:**130-144.
63. Holroyd S, Reiss AL, Bryan RN: **Autistic features in Joubert syndrome: a genetic disorder with agenesis of the cerebellar vermis.** *Biol Psychiatry* 1991, **29:**287-294.
64. Kim H, Lim CS, Kaang BK: **Neuronal mechanisms and circuits underlying repetitive behaviors in mouse models of autism spectrum disorder.** *Behav Brain Funct* 2016, **12:**3.
65. Wang C, Pan YH, Wang Y, Blatt G, Yuan XB: **Segregated expressions of autism risk genes Cdh11 and Cdh9 in autism-relevant regions of developing cerebellum.** *Mol Brain* 2019, **12:**40.
66. Gadow KD, Devincent CJ, Pomeroy J, Azizian A: **Comparison of DSM-IV symptoms in elementary school-age children with PDD versus clinic and community samples.** *Autism* 2005, **9:**392-415.
67. Simonoff E, Pickles A, Charman T, Chandler S, Loucas T, Baird G: **Psychiatric disorders in children with autism spectrum disorders: prevalence, comorbidity, and associated factors in a population-derived sample.** *J Am Acad Child Adolesc Psychiatry* 2008, **47:**921-929.
68. Greene RW, Biederman J, Faraone SV, Ouellette CA, Penn C, Griffin SM: **Toward a new psychometric definition of social disability in children with attention-deficit hyperactivity disorder.** *J Am Acad Child Adolesc Psychiatry* 1996, **35:**571-578.
69. 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* 2016, **7:**7.
70. Williams ME, Wilke SA, Daggett A, Davis E, Otto S, Ravi D, Ripley B, Bushong EA, Ellisman MH, Klein G, Ghosh A: **Cadherin-9 regulates synapse-specific differentiation in the developing hippocampus.** *Neuron* 2011, **71:**640-655.
71. Wang K, Zhang H, Ma D, Bucan M, Glessner JT, Abrahams BS, Salyakina D, Imielinski M, Bradfield JP, Sleiman PM, et al: **Common genetic variants on 5p14.1 associate with autism spectrum disorders.** *Nature* 2009, **459:**528-533.
72. Ma DQ, Salyakina D, Jaworski JM, Konidari I, Whitehead PL, Andersen AN, Hoffman JD, Slifer SH, Hedges DJ, Cukier HN, et al: **A Genome-wide Association Study of Autism Reveals a Common Novel Risk Locus at 5p14.1.** *Annals of Human Genetics* 2009, **73:**263-273.
73. Kerin T, Ramanathan A, Rivas K, Grepo N, Coetzee GA, Campbell DB: **A Noncoding RNA Antisense to Moesin at 5p14.1 in Autism.** *Science Translational Medicine* 2012, **4**.

74. DeWitt JJ, Grepo N, Wilkinson B, Evgrafov OV, Knowles JA, Campbell DB: **Impact of the Autism-Associated Long Noncoding RNA MSNP1AS on Neuronal Architecture and Gene Expression in Human Neural Progenitor Cells.** *Genes (Basel)* 2016, **7**.
75. DeWitt JJ, Hecht PM, Grepo N, Wilkinson B, Evgrafov OV, Morris KV, Knowles JA, Campbell DB: **Transcriptional Gene Silencing of the Autism-Associated Long Noncoding RNA MSNP1AS in Human Neural Progenitor Cells.** *Dev Neurosci* 2016, **38:**375-383.

Figures

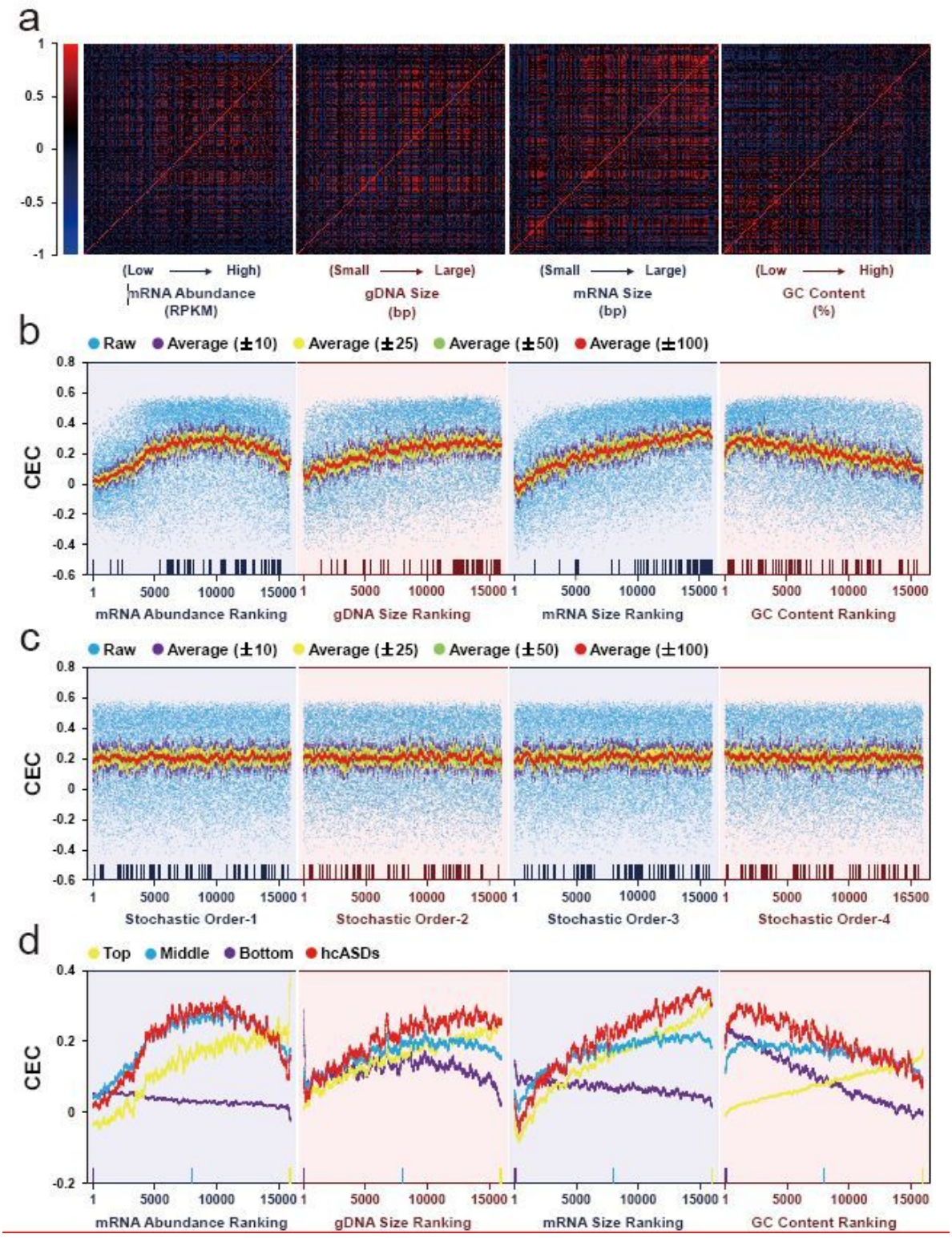

Figure 1

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 Additional file 2: Table S1 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 mRNA abundance and GC content or longest, medium, and shortest 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.

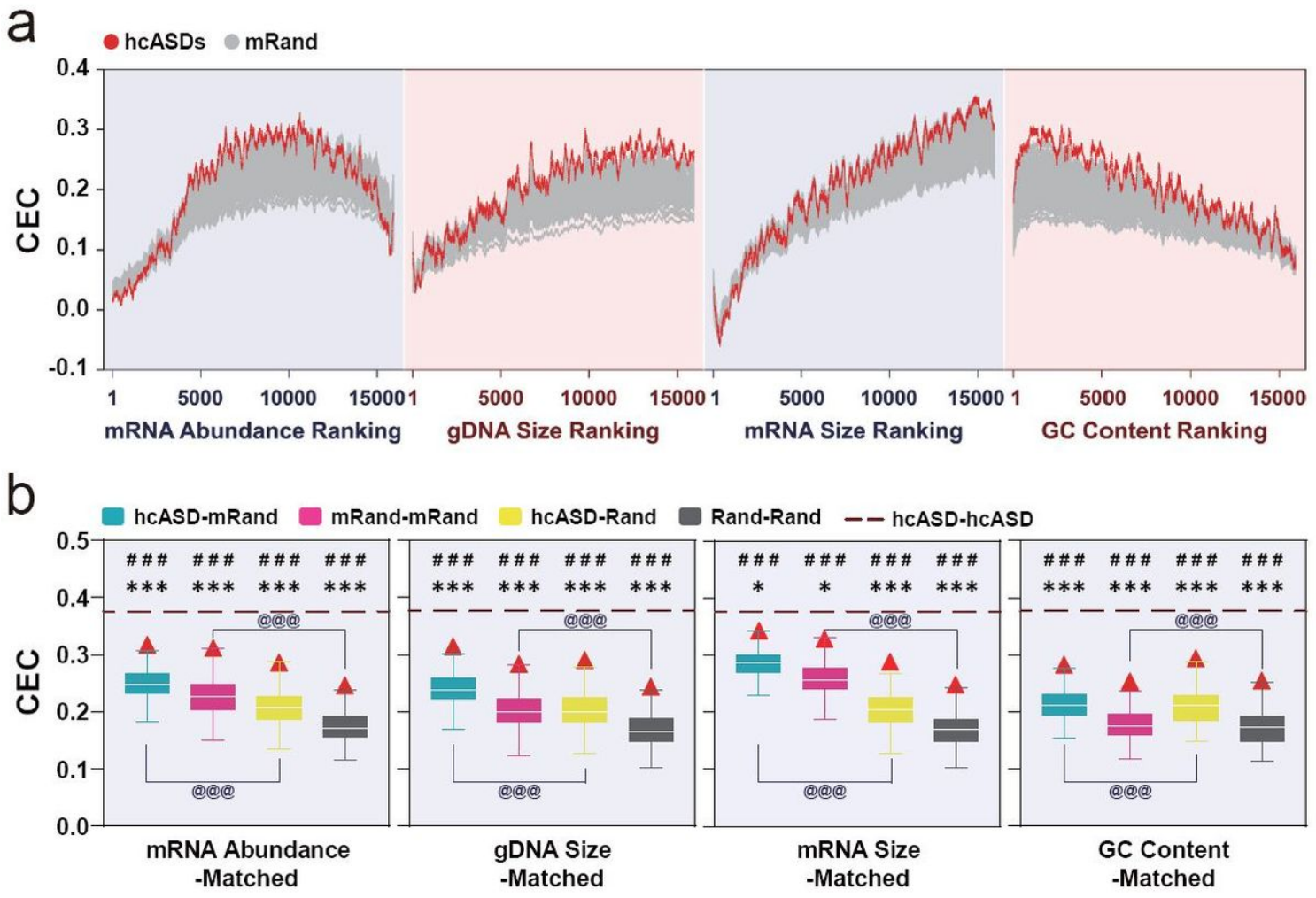

Figure 2

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. @@@ $P < 0.001$.

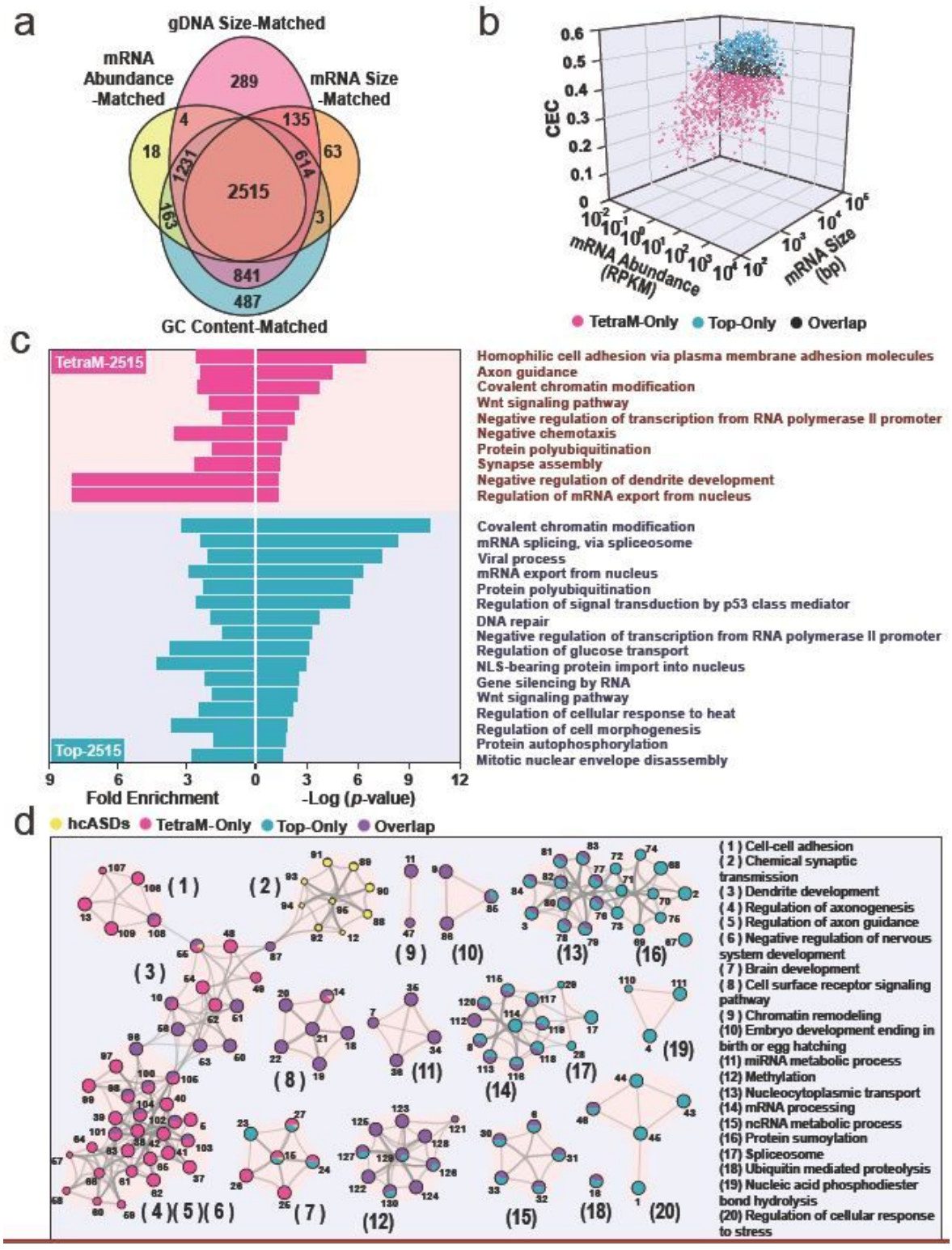

Figure 3

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.

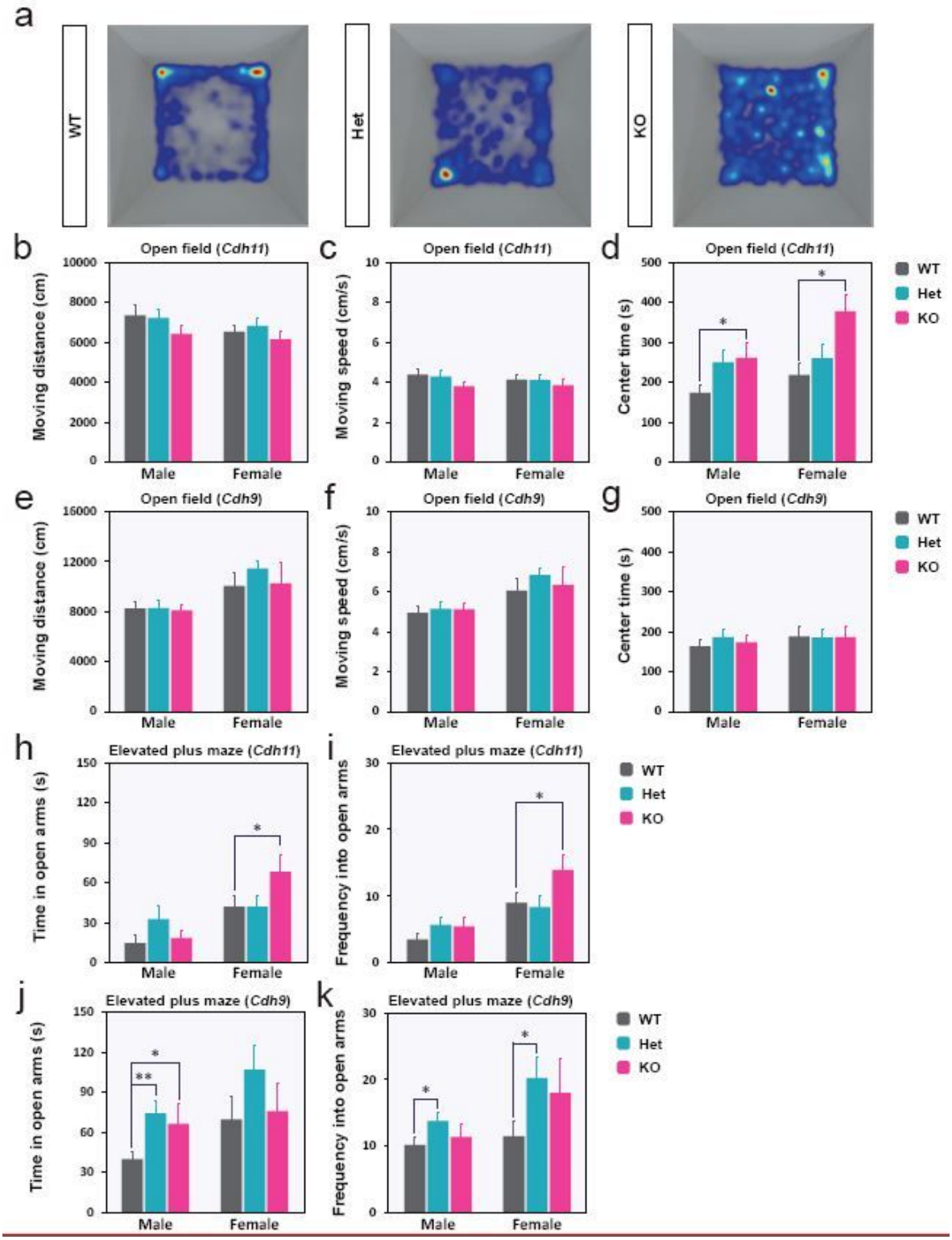

Figure 4

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. * $P < 0.05$, ** $P < 0.01$.

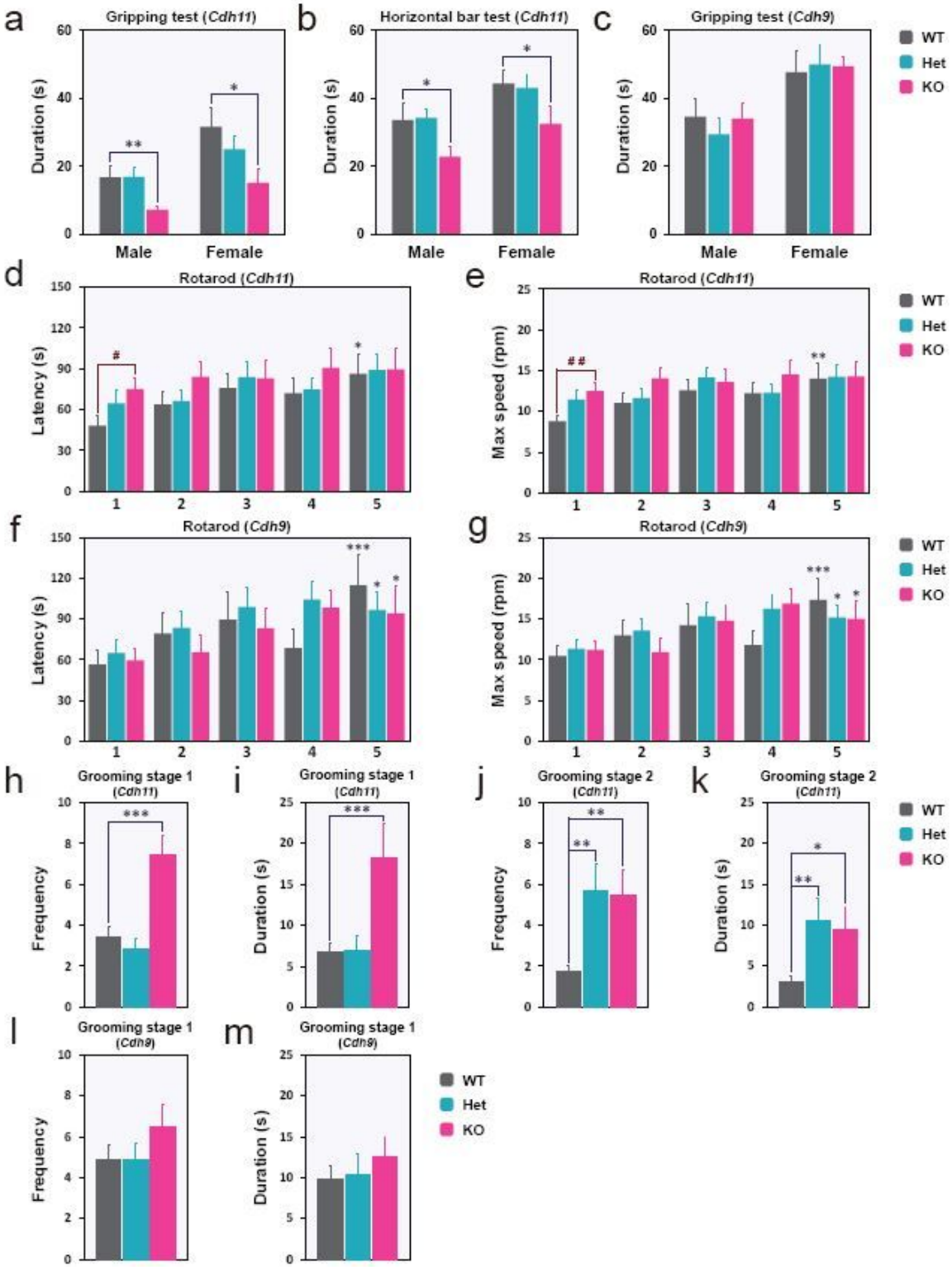

Figure 5

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.

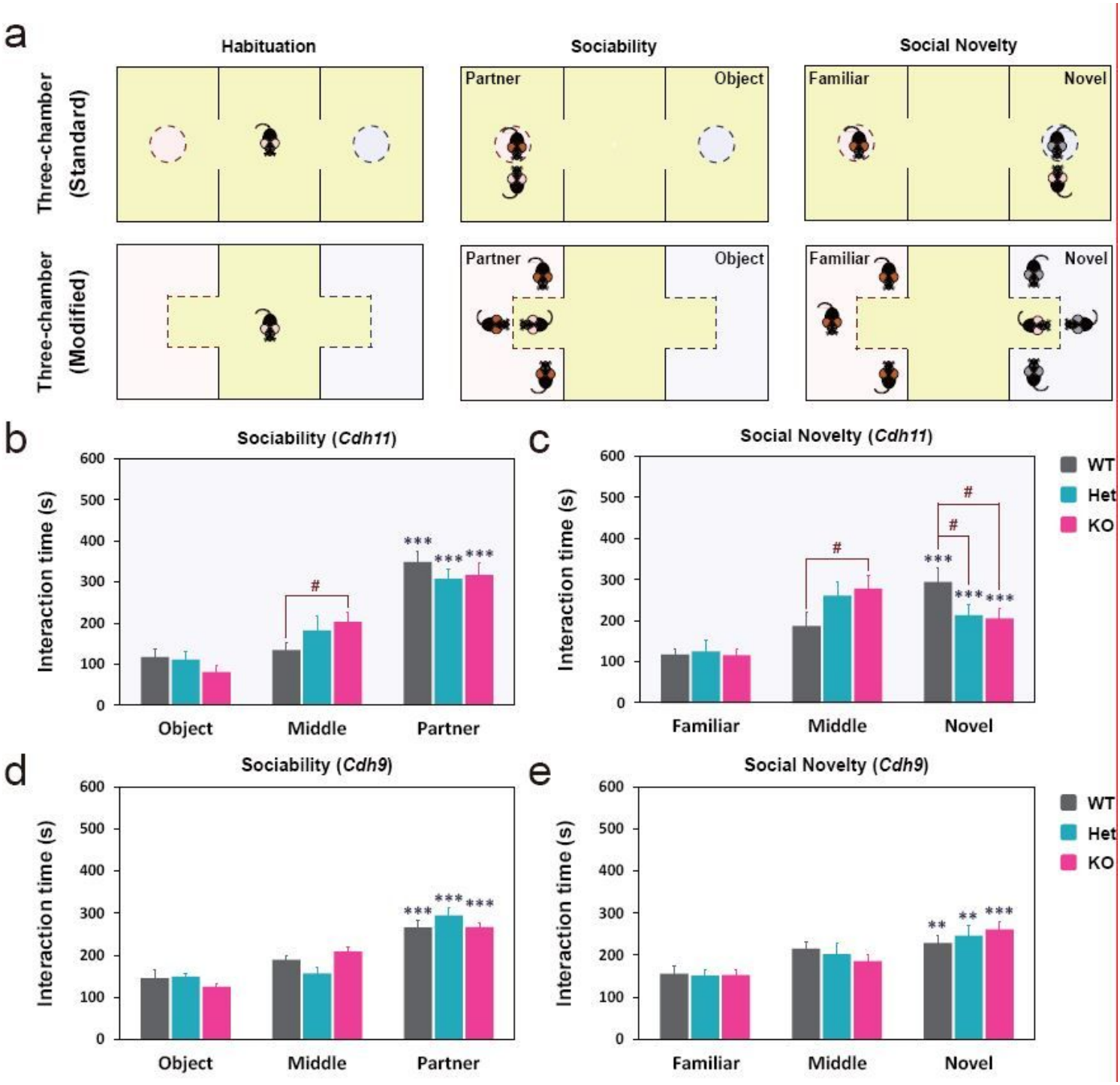

Figure 6

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.

Supplementary Files

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

- Additionalfile1.pdf
- Additionalfile2.xlsx
- Additionalfile3.xlsx
- Additionalfile4.xlsx
- Additionalfile5.xlsx
- Additionalfile6.xlsx
- Additionalfile7.xlsx
- Additionalfile8.xlsx
- Additionalfile9.xlsx
- Additionalfile10.xlsx