

1 **A brain structural connectivity biomarker for diagnosis of autism** 2 **spectrum disorder in early childhood**

3

4 **Authors**

5 Xi Jiang, PhD^{1*}, Xiao-Jing Shou, PhD^{2,3*}, Zhongbo Zhao, MSc¹, Fanchao Meng, MD², Jiao
6 Le, PhD^{1,5}, Tianjia Song, PhD², Xinjie Xu, PhD², Xiaoyan Ke, MD⁴, Yuzhong Chen, BS¹,
7 Xiaoe Cai, MPhil², Weihua Zhao, PhD¹, Juan Kou, PhD¹, Ran Huo MD², Ying Liu MD⁶,
8 Huishu Yuan MD⁶, Yan Xing MD⁷, Jisheng Han, BS², Songping Han PhD⁸, Yun Li, MD⁴,
9 Hua Lai, BS⁵, Lan Zhang, MM⁵, Meixiang Jia, BS⁹, Jing Liu, MD⁹, Keith M. Kendrick,
10 PhD^{1#}, Rong Zhang, MD^{2,10#}

11

12 ¹The Clinical Hospital of Chengdu Brain Science Institute, MOE Key Lab for
13 Neuroinformatics, University of Electronic Science and Technology of China, Chengdu,
14 611731, China.

15 ²Neuroscience Research Institute; Key Laboratory for Neuroscience, Ministry of Education
16 of China; Key Laboratory for Neuroscience, National Committee of Health and Family
17 Planning of China; and Department of Neurobiology, School of Basic Medical Sciences,
18 Peking University, Beijing, 100191, China.

19 ³State Key Laboratory of Cognitive Neuroscience and Learning; Beijing Key Laboratory of
20 Brain Imaging and Connectomics; and IDG/McGovern Institute for Brain Research, Beijing
21 Normal University, Beijing, 100875, China.

22 ⁴Child Mental Health Research Center, Nanjing Brain Hospital Affiliated of Nanjing Medical
23 University, Nanjing, 210029, China.

24 ⁵Chengdu Women's and Children's Central Hospital, School of Medicine, University of
25 Electronic Science and Technology of China, Chengdu, 611731, China.

26 ⁶Radiology Department, Peking University Third Hospital, Beijing, 100191, China.

27 ⁷Department of Pediatrics, Peking University Third Hospital, Beijing, 100191, China.

28 ⁸Wuxi Shengpingxintai Medical Technology Co., Ltd., Wuxi, 214091, China.

29 ⁹Mental Health Institute, Peking University, Key Laboratory of Ministry of Health, The
30 Ministry of Public Health, Beijing, 100191, China.

31 ¹⁰Autism Research Center of Peking University Health Science Center, Beijing, 100191,
32 China

33

34 ***Co-first authors:** Xi Jiang and Xiao-Jing Shou

35 **#Joint corresponding authors:**

36 Professor Keith M. Kendrick,

37 The Clinical Hospital of Chengdu Brain Science Institute, MOE Key Laboratory for

38 Neuroinformation,

39 University of Electronic Science and Technology of China,

40 No. 2006, Xiyuan Ave, West High-Tech Zone, Chengdu, 611731, China

41 Tel/Fax: 86-28-61830811

42 E-mail: kkendrick@uestc.edu.cn

43

44 Dr. Rong Zhang,

45 Neuroscience Research Institute; Key Laboratory for Neuroscience, Ministry of Education of

46 China; Key Laboratory for Neuroscience, National Committee of Health and Family

47 Planning of China; Department of neurobiology, School of Basic Medical Sciences, Peking

48 University,

49 No. 38, Xueyuan Road, Haidian District, Beijing, 100191, China

50 Tel: 86-10-82801152

51 E-mail: zhangrong@bjmu.edu.cn

52

53 **Abstract**

54 **Objective:** Autism spectrum disorder (ASD) is associated with altered brain development,
55 but it is unclear which specific structural changes may serve as potential diagnostic markers.
56 This study aimed to identify and model brain-wide differences in structural connectivity
57 using MRI diffusion tensor imaging (DTI) in young ASD and typically developing (TD)
58 children (3-5-6 years old).

59 **Methods:** Ninety-three ASD and 26 TD children were included in a discovery dataset and 12
60 ASD and 9 TD children from different sites included as independent validation datasets.
61 Brain-wide (294 regions) structural connectivity was measured using DTI (fractional
62 anisotropy, FA) under sedation together with symptom severity and behavioral and cognitive
63 development. A connection matrix was constructed for each child for comparisons between
64 ASD and TD groups. Pattern classification was performed and the resulting model tested on
65 two independent datasets.

66 **Results:** Thirty-three structural connections showed increased FA in ASD compared to TD
67 children and associated with both symptom severity and general cognitive development. The
68 majority (29/33) involved the frontal lobe and comprised five different networks with
69 functional relevance to default mode, motor control, social recognition, language and reward.
70 Overall, classification accuracy is very high in the discovery dataset 96.77%, and 91.67% and
71 88.89% in the two independent validation datasets.

72 **Conclusions:** Identified structural connectivity differences primarily involving the frontal
73 cortex can very accurately distinguish individual ASD from TD children and may therefore
74 represent a robust early brain biomarker.

75

76 **Keywords**

77 autism; neuroimaging; diffusion tensor imaging; fractional anisotropy; brain structural
78 connectivity; diagnosis; early childhood

79 **Introduction**

80 There is an increasing consensus that children with autism spectrum disorder (ASD) have an
81 aberrant pattern of brain development.¹⁻² A number of structural magnetic resonance imaging
82 studies using diffusion tensor imaging (DTI) have identified altered fractional anisotropy
83 (FA), which is a widely used index in DTI to reflect axonal density and myelination, in
84 individuals with ASD, particularly in the frontal,³⁻⁵ occipital lobes,⁶ and corpus callosum.^{4,7}
85 Developmental changes in FA occur with increases in fiber tracts during the first few years,
86 followed by decreases in later childhood and adolescence, through into adulthood.^{3,5}
87 However, the discrimination accuracies between ASD and typically developing (TD)
88 individuals reported by a small number of studies to date following machine learning
89 classification approaches only found modest effects and have not been validated using
90 independent validation datasets (75-80%⁶⁻⁷). Analysis of large fiber tracts connecting many
91 different brain regions may also obscure changes in altered structural connectivity between
92 specific brain regions.

93

94 In the current study we have therefore used DTI to identify differences in inter-regional
95 structural connectivity at the whole-brain level in ASD compared to TD children in 294
96 different brain regions. We chose to restrict the age range of children to 3-5-6 years old since
97 this corresponds to the period when ASD symptoms have become robustly established.⁸
98 Previous tractography-based research also suggests that at this age overall structural
99 connectivity differences between ASD and TD children may be less pronounced.⁴

100

101 Based on previous studies, we firstly hypothesized that ASD children would exhibit
102 significantly greater FA in structural connections at the whole brain level and particularly
103 involving frontal regions compared to TD children. Secondly, we hypothesized that altered
104 structural connections would be in networks associated with ASD symptoms and cognitive
105 and behavioral development. Finally, we hypothesized that the structural connectivity
106 changes identified would accurately predict ASD diagnosis at the individual level not only
107 within the original discovery dataset but also in two independent validation datasets.

108

109 **Methods**

110 **Participants**

111 The present study included three independent datasets: a discovery and two validation
112 datasets.

113

114 **Discovery dataset (Beijing).** The experiment was approved by the ethics committee of the
115 Peking University Institutional Review Board (approval no. IRB00001052-13079). A total of
116 119 pre-school children either diagnosed with ASD (n = 93) or TD children (n = 26) were
117 recruited. The age range of participants was 3-5 to 6 years, which is regarded as the time of
118 the most severe emerging symptoms of autism.⁸ Children with ASD were recruited through
119 pediatric psychiatric clinics and autism rehabilitation training centers in Beijing. Age and
120 gender matched TD children were also recruited through online social platforms or day care
121 centers in Beijing.

122

123 **ASD validation dataset (Chengdu).** The experiment was approved by the ethics committee
124 of the University of Electronic Science and Technology of China (approval no. 1420190601).
125 A total of 12 ASD children were recruited aged 3 to 8 years through the child healthcare
126 department of Chengdu Women's and Children's Central Hospital.

127

128 **TD validation dataset (Nanjing).** The experiment was approved by the medical ethics
129 committee of the Brain Hospital affiliated to Nanjing Medical University (approval no.
130 KY043). A total of 9 TD children were recruited aged 4 to 6 years either through the Nanjing
131 child mental health research center, online social platforms or day care centers.

132

133 All the participants' parents in the three datasets were informed in detail of the research
134 objectives and procedures, and provided written informed consents. Exclusion criteria were:
135 (1) neurological complications, such as epilepsy, cerebral palsy, Fragile X syndrome etc.; (2)
136 medical intervention, such as antipsychotic drugs, transcranial magnetic stimulation,
137 acupuncture etc.; (3) diagnostic imaging anomalies or craniocerebral trauma; (4) other
138 contraindications to MRI; (5) TD children had no family histories of any mental disorders
139 and exhibited no evidence of developmental delay.

140

141 **Clinical Diagnosis**

142 Participants in ASD groups were diagnosed at either Peking University Sixth Hospital or
143 Beijing Children's Hospital, Beijing, China for Beijing dataset, or at Chengdu Women's and
144 Children's Central Hospital, Chengdu, China for Chengdu dataset. All children in the ASD

145 group met the diagnostic criteria of Diagnostic and Statistical Manual of Mental Disorders
146 IV-Text Revision (DSM-IV-TR)⁹ or Fifth Edition (DSM-5)¹⁰ and International Statistical
147 Classification of Diseases and Related Health Problems 10th revision (ICD-10)¹¹. In addition,
148 ASD diagnosis was confirmed in the Beijing dataset using the Autism Diagnostic
149 Observation Schedule (ADOS)¹² Traditional Mandarin version, module 1 or module 2 based
150 on the child's language ability. For children in the Chengdu dataset, diagnosis was confirmed
151 using ADOS-2.¹³ Moreover, in the Beijing ASD and TD cohorts, cognitive ability was also
152 assessed using the Gesell Developmental Scale (GDS)¹⁴ administered by an experienced
153 pediatrician. This is a measure of cognitive and behavioral development and adaptability
154 including five components (gross motor, fine motor, adaptive, language and personal social
155 behaviors).

156

157 **MRI Acquisition and Preprocessing**

158 Children in both ASD and TD groups of the three datasets were sedated by oral
159 administration of chloral hydrate at the 50 mg/kg body weight (1 g maximum dose),
160 commonly used for pediatric clinical imaging. During the MRI scan, children wore earplugs
161 and de-noising headsets to reduce the noise, and parents were encouraged to remain in the
162 scanning room to ensure safety in case the child awoke. In all sites, brain images were
163 reviewed by neuroradiologists to confirm absence of neurological abnormalities.

164

165 For the discovery dataset MRI images were acquired on a GE 3T MR750 scanner with a 12-
166 channel head coil at the Peking University Third Hospital. DTI data were obtained with an
167 echo-planar imaging sequence: TR = 9,000 ms, TE = 89.4 ms, FOV = 256 mm, matrix size =
168 128 × 128, voxel size = 2 mm isotropic, 75 slices covering the whole brain with no gap, 32
169 diffusion directions, b-value = 1000 s/mm². For independent ASD and TD validation datasets
170 see supplementary methods. Pre-processing methods for DTI raw data are described in the
171 supplementary methods.

172

173 **Overview of analysis**

174 The flow chart of ASD identification framework is illustrated in Fig. 1. Based on the pre-
175 processed DTI FA map (Fig. 1(A)), the whole-brain streamline fibers were reconstructed (Fig.
176 1(B)) and adopted to the brain atlas (Fig. 1(C)). To construct the structural connection matrix
177 for each subject (Fig. 1(D)), we calculated the pair-wise FA value between every two brain
178 regions. The constructed structural connection matrix was further projected into the brain for

179 each participant (Fig. 1(E)). Group comparisons were then performed based on all structural
180 connection networks between ASD and TD groups (Fig. 1(F)) to obtain the between-group
181 differences (Fig. 1(G)). The resulting connections were finally adopted as features to perform
182 pattern classification between ASD and TD groups (Fig. 1(H)). The details of each step are
183 demonstrated in the following sections.

184

185 **Construction of Structural Connection Network**

186 The brain atlas used in this study included 294 non-overlapping brain regions consisting of
187 246 cortical and subcortical regions, 34 cerebellar regions, and 14 brainstem regions (Fig.
188 1(C)). The cortical and subcortical regions were from the Brainnetome Atlas¹⁵ with 123
189 homotopic regions in each hemisphere; cerebellar regions were from the Human Cerebellar
190 Probabilistic Magnetic Resonance Atlas¹⁶ and brainstem regions were from the Human
191 Brainstem Standard Neuroanatomy Atlas¹⁷. Since the three atlases were originally defined in
192 the adult MNI152 standard space, we first performed linear registration to warp the T1-
193 weighted adult MNI152 image to the space of T1-weighted template image of children aged
194 4-5-8-5 years ([https://www.mcgill.ca/bic/software/tools-data-analysis/anatomical-](https://www.mcgill.ca/bic/software/tools-data-analysis/anatomical-mri/atlases/nihpd)
195 [mri/atlases/nihpd](https://www.mcgill.ca/bic/software/tools-data-analysis/anatomical-mri/atlases/nihpd)),¹⁸ and then applied the linear transformation to the brain atlas to warp it to
196 the children's template image space as well in order to obtain the young children's brain atlas.

197

198 The reconstructed whole-brain streamline fibers in each child (Fig. 1(B)) were then aligned to
199 the young children's brain atlas space via DSI Studio¹⁹. The structural connection matrix for
200 each child (Fig. 1(D)) was obtained by calculating the pair-wise mean FA value between
201 every two brain regions via DSI Studio¹⁹, and further represented as a structural connection
202 network (Fig. 1(E)) in which the nodes were the brain regions and edges were the mean FA
203 values between two nodes.

204

205 **Identification of Structural Connections Showing Differences between ASD and TD**

206 Based on the structural connection networks of all children in the ASD and TD groups (Fig.
207 1(F)), we adopted the widely used Network-based Statistic (NBS) approach²⁰ to identify the
208 structural connections and associated networks showing between-group differences (Fig.
209 1(G)). As a non-parametric statistical method, NBS first performs a large number of
210 univariate hypothesis tests on all edges in the network, then clustering-based statistics, and
211 finally permutation tests to calculate the family-wise error rate (FWER) corrected *p*-values
212 for each sub-network consisting of edges with group differences. We adopted the NBS

213 Connectome toolbox implemented in Matlab²⁰ to perform the analysis. The structural
214 connection networks of all participants in ASD and TD groups were the inputs with gender
215 and age as covariates. Next we performed the NBS analysis to identify both increased and
216 decreased FA values of structural connections and associated networks in ASD compared to
217 TD.

218

219 **Pattern Classification between ASD and TD Children**

220 Based on the identified structural connections showing between-group differences (Fig. 1(G)),
221 we further adopted FA values of those connections as features to perform pattern
222 classification between the ASD and TD groups. We used the discovery dataset as the training
223 dataset to establish the classification model. For training and leave-one-out cross-validation,
224 we employed the widely used support vector machine (SVM) approach. The training model
225 was then applied to the two independent validation datasets to validate its generalizability.
226 Specifically, we adopted the widely used cost-support vector classification (C-SVC), and
227 the radial basis function (RBF) as the kernel function in SVM. The optimal values of
228 parameters c (i.e., cost, c ranged from 15 to 16) and g (i.e., gamma, g ranged from 0.08
229 to 1) in RBF were obtained by a hyperparameter optimization framework of optuna.²¹

230

231 **Potential Effect of Imbalanced Sample Size between ASD and TD**

232 In view of the imbalanced group sample sizes in the discovery dataset (93 ASD and 26 TD),
233 we adopted both up-sampling and down-sampling approaches to alleviate the potential model
234 overfitting as well as classification bias problem (see supplementary methods). Both
235 approaches confirmed that our classification model was not influenced by the imbalanced
236 sample sizes.

237

238 **Statistical Analysis**

239 Independent two-sample t -tests were utilized to identify the different structural connections
240 between ASD and TD ($n=10,000$ permutation times, significance level $p<0.05$, FWE
241 corrected) in NBS. Independent sample t -tests were used for continuous variables including
242 age, BMI, head circumference, GDS total score, and FA between ASD and TD. Chi-square
243 tests were used for categorical variables including gender and handedness between ASD and
244 TD. Pearson's linear correlation coefficients were computed between the averaged FA value
245 and ADOS and GDS scores (one-sample t -tests, significance level at $p<0.05$, FDR corrected).
246 A mediation model was conducted (PROCESS)²² using bootstrap analysis to investigate the

247 relationship between the averaged FA value, ADOS total and GDS total scores (bootstrap =
248 1000).

249

250 **Results**

251 **Subject Demographics and Behavioral Measures**

252 Table 1 summarizes demographic and other information for ASD and TD groups in the
253 different datasets and ADOS scores for ASD children. There were no group differences in
254 age, gender, BMI, handedness, and head circumference in the discovery dataset, although as
255 expected the total GDS score was significantly less in the ASD group indicating impaired
256 cognitive and behavioral development.

257

258 **Increased FA Connections and Networks in ASD Compared to TD Children**

259 We identified 33 increased but no decreased FA values of structural connections in ASD
260 compared to TD children (Figs. 2(A) and 2(C), Table 2). Notably, 29 out of 33 connections
261 were associated with the frontal lobe. Moreover, 30 out of 33 connections were intra-
262 hemispheric. Fig. 2(B) illustrates the locations of 33 connections on the cortical surface. In
263 the ASD group averaged FA values were significantly negatively correlated (FDR corrected)
264 with ADOS total and ‘social interaction’ sub-scale scores (Fig. 2(D)) and positively
265 correlated with the GDS total score (Fig. 2(E)). In the TD group there was a slight but not
266 significant negative correlation between averaged FA values and GDS score (Fig. 2(E)). A
267 mediation analysis indicated that within the ASD group the ADOS total score was the main
268 mediator of the effects on the averaged FA value and GDS total score (Fig. 2(F)).

269

270 Connections with increased FA were further categorized into 5 structural networks via NBS
271 (Fig. 2(G) and Table 2) together with a functional characterization using the Neurosynth
272 platform, and visualized as a word cloud. Network 1: default mode function and memory
273 retrieval, including: left superior temporal gyrus, precuneus, inferior parietal lobule, superior
274 frontal gyrus, and medioventral occipital cortex; Network 2: motor function, including: right
275 superior and inferior frontal gyri, right cerebellum lobe IX and left postcentral gyrus;
276 Network 3 visual and facial recognition function, including: right inferior and superior frontal
277 gyri, lateral occipital cortex, pre-motor thalamus, middle frontal gyrus, medioventral occipital
278 cortex, basal ganglia, superior temporal gyrus and insula; Network 4 language and cognitive
279 function, including: left middle and inferior frontal gyri, thalamus, cingulate gyrus, basal

280 ganglia and cerebellum lobe IX; Network 5: social and general reward functions, including:
281 bilateral orbitofrontal and cingulate gyri and basal ganglia.

282

283 **Classification Accuracy between ASD and TD Children**

284 We first up-sampled the discovery dataset to 186 subjects with 93 ASD and 93 TD and
285 trained the classification model in a 33-dimensional feature space based on the 33
286 connections using a leave-one-out cross-validation strategy. Fig. 3(A) shows the
287 classification model in a 3-dimensional feature space after performing dimensionality
288 reduction using the t-distributed stochastic neighbor embedding (t-SNE) algorithm.²³ The
289 Receiver Operating Characteristic (ROC) curve of the training model is shown in Fig. 3(B).
290 The area under the ROC curve (AUC) was 0.981, indicating the robustness of the training
291 model. The confusion matrix of the training model is shown in Fig. 3(C). Accuracy,
292 sensitivity, specificity, precision, and F measure in both discovery and validation datasets are
293 reported in Fig. 3(D) with the proposed model achieving high classification accuracy in both
294 discovery (96.77%) and independent validation datasets (91.67% and 88.89%). The
295 alternative down-sampling strategy of the discovery dataset by 1,000 times also achieved
296 high classification accuracy in both discovery (94.85±1.30%) and independent validation
297 datasets (91.63±5.55% and 80.04±5.52%). Thus overall, the classification model showed
298 both high accuracy and generalizability for ASD identification across different datasets
299 without being influenced by the imbalanced sample sizes of the discovery dataset.

300

301 **Discussion**

302 Our findings have revealed the presence of a small number of inter-regional structural
303 connections within the brains of young children with ASD which exhibit increased FA
304 compared to TD and negatively associated with symptom severity. The majority of affected
305 regional connections involve the frontal cortex and overall they achieved a classification
306 accuracy of 96.77% for discriminating between ASD and TD individuals in the discovery
307 dataset and 91.67% and 88.89% in two small independent datasets. The 33 inter-regional
308 structural connections could be clustered into 5 independent networks with relevance to a
309 range of behavioral functions influenced in ASD.

310

311 In support of our original hypothesis, the majority of the 33 structural connections showing
312 increased FA in children with ASD in the current study involved the frontal lobe including

313 intrinsic short range frontal-frontal connections and longer range frontal-occipital, frontal-
314 thalamic and frontal-limbic ones. This is in agreement with findings from other studies^{3-4,24-26}
315 and supports the conclusion that alterations in both intrinsic and extrinsic frontal lobe
316 structural connectivity contribute fundamentally to ASD.

317

318 The altered structural connections in children with ASD could be clustered into 5 individual
319 networks encompassing default mode, motor, visual and facial recognition, language and
320 memory and reward functions. The largest single frontal lobe cluster involved orbitofrontal
321 regions and their connections with the basal ganglia. These intrinsic frontal connections are
322 strongly associated with social and other types of reward processing as well as decision
323 making²⁷ and these functions are known to be impaired in ASD.²⁸⁻²⁹ Three other clusters
324 involved inferior, medial and superior frontal gyri connections with thalamus, basal ganglia,
325 cingulate, insula, occipital cortex, post-central gyrus and cerebellum associated with social
326 cognition, language comprehension, sensory and motor processing functions,³⁰⁻³² all of which
327 are impaired in ASD.³³⁻³⁶ The remaining cluster primarily involved connections between the
328 superior temporal gyrus and inferior parietal lobule with the precuneus in the default mode
329 network, associated with self-processing, experience of agency, autobiographic and episodic
330 memory retrieval and visuospatial imagery. Default mode dysfunction has been consistently
331 reported in ASD³⁷ as well as impaired self-processing, sense of agency, autobiographical and
332 episodic memory.³⁸⁻³⁹

333

334 A previous study using DTI measures and classification techniques to identify ASD
335 compared to TD children employed shape representations of white matter fiber tracts as
336 features, and achieved 75.34% discrimination accuracy using a leave-one-out cross-validation
337 approach.⁷ Another study adopted the anisotropy scores of regions of interest as features, and
338 achieved 80% accuracy using leave-one-out cross-validation.⁶ In our current study, we
339 adopted the DTI-derived FA values of structural connections as features, and achieved a
340 much higher classification accuracy in both the discovery dataset (96.77%) and, importantly,
341 in independent validation datasets (91.67% and 88.89%), demonstrating satisfying
342 classification and generalization ability of our model across different datasets.

343

344 Unexpectedly, we found a significant negative correlation between the averaged FA values of
345 the 33 altered connections in children with ASD and ADOS total and social sub-scale scores,
346 indicating that symptom severity was actually lower in children with greater FA. Scores on

347 GDS were also positively correlated with FA values in the ASD group but slightly negatively
348 correlated in the TD group. A mediation analysis identified that ADOS scores were primarily
349 mediating both FA values and GDS scores in the ASD group. This may indicate an
350 experience-dependent compensatory effect is occurring in children with ASD whereby
351 increased FA contributes to reduced symptom severity and enhanced cognitive and
352 behavioral development. A social experience compensation effect has previously been
353 described in behavioral studies of autism.⁴⁰ Interestingly, a tractography study has reported a
354 positive association between increased frontal lobe FA and symptom severity in very young
355 children but a negative one in older children in the age-range of the current study.⁴ Thus,
356 children who experience more severe symptoms at the age of 3.5-6 years may have reduced
357 FA in these neural circuits compared to when they were younger, whereas those with milder
358 symptoms may instead have maintained or even increased their FA. A longitudinal study
359 would clearly be required to confirm this possibility.

360

361 **Limitations**

362 A limitation of the current study is its cross-sectional nature and restricted age range (3-5-6
363 years old). Patterns of structural differences may differ in both younger and older individuals
364 and only a longitudinal design study can address this. A second limitation is we did not
365 determine whether observed changes are specific to ASD or might also occur in children with
366 developmental delay, for example. A final limitation is that we could only obtain two small
367 datasets for independent analysis of discrimination accuracy although the findings were very
368 encouraging.

369

370 **Conclusion**

371 By employing a fine-grained, brain-wide analysis of structural differences for regional
372 connections in the brains of young (3-5-6 years old) autistic compared with typically
373 developing children we have identified a number of structural connections mainly involving
374 the frontal lobe exhibiting increased FA but negatively associated with symptom severity.
375 Differences in these structural connections show high accuracy (>96%) in discriminating
376 autistic children from TD children which generalizes to independent novel datasets. These
377 new findings suggest that differences in structural connections primarily involving the frontal
378 cortex of young autistic children are a potential reliable and generalizable biomarker for ASD
379 diagnosis and for assessing the efficacy of therapeutic interventions.

380

381 **Contributors**

382 XJ: conceptualization, methodology, software, validation, formal analysis, resources, data
383 curation, writing - original draft, writing - review & editing, visualization, supervision,
384 project administration, funding acquisition. XJS: methodology, validation, investigation,
385 resources, data curation, writing – review & editing, project administration, funding
386 acquisition. ZBZ: methodology, software, validation, formal analysis, visualization. FCM:
387 investigation. JL: investigation. TJS: investigation. XJX: investigation. XYK: resources.
388 YZC: formal analysis, visualization. XEC: investigation. WHZ: methodology, formal
389 analysis. JK: investigation. RH: investigation. YL: investigation. HSY: resources. YX:
390 investigation. JSH: conceptualization, resources. SPH: investigation. YL: investigation. HL:
391 conceptualization. LZ: conceptualization. MXJ: investigation. JL: resources. KMK:
392 conceptualization, formal analysis, resources, writing - review & editing, visualization,
393 supervision, project administration, funding acquisition. RZ: conceptualization, resources,
394 writing - review & editing, visualization, supervision, project administration, funding
395 acquisition.

396 **Declaration of interests**

397 We declare no competing interests.

398 **Data sharing**

399 Individual participant data and the data dictionary defining each field in the set will not be
400 made available as all participants did not consent to have their data as a public resource. The
401 group-level data results as well as the data processing code which do not disclose the
402 participants' information will be available with publication from the corresponding authors
403 (KMK and RZ) on reasonable request (including a research proposal), subject to review.

404 **Acknowledgments**

405 This work was supported by Key Technological Projects of Guangdong Province (grant
406 number 2018B030335001 - KMK), Beijing Municipal Science & Technology Commission
407 (grant number Z181100001518005 - RZ), the National Key Research and Development
408 Program of China (2016YFC0105501 - RZ), Sichuan Science and Technology Program
409 (2021YJ0247 - XJ), National Natural Science Foundation of China (61976045 - XJ,
410 81801779 - SXJ). The funders played no role in the writing of the manuscript or the decision
411 where to submit it for publication. Furthermore, the authors would like to thank Dr. Ruo-Yan
412 Zhao for her GDS assessment. We appreciate Ms. Qing-Yun Wei of Sunshine Friendship
413 Rehabilitation Center, Beijing, China, and Ms. Meng-Lin Sun of Wucailu Rehabilitation
414 Center, Beijing, China, for their cooperation of children recruitment. Finally, we are grateful
415 to all the participant children and their families.

416 **References**

- 417 1. Hazlett HC, Poe MD, Gerig G, Styner M, Chappell C, Smith RG, Vachet C, Piven J.
418 Early brain overgrowth in autism associated with an increase in cortical surface area
419 before age 2 years. *Arch Gen Psychiatry*. 2011; **68(5)**:467-76.
- 420 2. Girault JB, Piven J. The neurodevelopment of autism from infancy through toddlerhood.
421 *Neuroimaging Clinical*. 2020; **30(1)**:97-114.
- 422 3. Catani M, DA Flavio, B Sanja, H Henrietta, TS Michel, et al. Frontal networks in
423 adults with autism spectrum disorder. *Brain*. 2016; **139(2)**:616-30.
- 424 4. Solso S, Xu R, Proudfoot J, et al. Diffusion tensor imaging provides evidence of
425 possible axonal overconnectivity in frontal lobes in autism spectrum disorder toddlers.
426 *Biological Psychiatry*. 2016; **79(8)**: 676-684.
- 427 5. Walker L, Gozzi M, Lenroot R, et al. Diffusion tensor imaging in young children with
428 autism: biological effects and potential confounds. *Biological Psychiatry*. 2012; **72**:
429 1043-51.
- 430 6. Ingalhalikar M, Parker D, Bloy L, et al. Diffusion based abnormality markers of
431 pathology: toward learned diagnostic prediction of ASD. *Neuroimage*. 2011;
432 **57(3)**:918-27.
- 433 7. Adluru N, Hinrichs C, Chung MK, et al. Classification in DTI using shapes of white
434 matter tracts. 2009 Annual International Conference of the IEEE Engineering in
435 Medicine and Biology Society. 2009 Sept 3-6; Minneapolis, MN, USA: IEEE; 2009
436 Nov. 2719 p.
- 437 8. Lord C, Elsabbagh M, Baird G, Veenstra-Vanderweele J. Autism spectrum disorder.
438 *Lancet*. 2018 Aug; **392(10146)**:508-20.
- 439 9. American Psychiatric Association. Diagnostic and statistical manual of mental
440 disorders: DSM-4. 4th edn. American Psychiantric Publishing; 2000.

- 441 10. American Psychiatric Association. Diagnostic and statistical manual of mental
442 disorders: DSM-5. 5th edn. American Psychiatric Publishing; 2013.
- 443 11. World Health Organization. International statistical classification of diseases and
444 related health problems. 10th revision, 2nd edn. World Health Organization; 2004.
- 445 12. Lord C, Rutter M, Goode S, et al. Autism diagnostic observation schedule: a
446 standardized observation of communicative and social behavior. *Journal of Autism and*
447 *Developmental Disorders*. 1989; **19(2)**:185-212.
- 448 13. Lord C, Rutter M, Pamela C, et al. ADOS: Autism diagnostic observation schedule.
449 Boston, MA: Hogrefe; 2008.
- 450 14. Jin X, Sun Y, Jiang F, et al. "Care for Development" intervention in rural China: a
451 prospective follow-up study. *Journal of Developmental Behavior and Pediatrics*. 2007;
452 **28(3)**:213-8.
- 453 15. Fan L, Li H, Zhuo J. The human brainnetome atlas: a new brain atlas based on
454 connectional architecture. *Cerebral Cortex*. 2016; **26(8)**:3508-26.
- 455 16. Diedrichsen J, Balsters JH, Flavell J, et al. A probabilistic MR atlas of the human
456 cerebellum. *Neuroimage*. 2009; **46(1)**:39-46.
- 457 17. Edlow BL, Takahashi E, Wu O, et al. Neuroanatomic connectivity of the human
458 ascending arousal system critical to consciousness and its disorders. *Journal of*
459 *Neuropathology & Experimental Neurology*. 2012; **71(6)**:531-46.
- 460 18. Fonov V, Evans AC, Botteron K, Almlí CR, McKinstry RC, Collins DL, & Brain
461 Development Cooperative Group. Unbiased average age-appropriate atlases for
462 pediatric studies. *NeuroImage*. 2011; **54(1)**:313-27.
- 463 19. Yeh FC, Verstynen TD, Wang Y, Fernández-Miranda JC, Tseng WYI. Deterministic
464 diffusion fiber tracking improved by quantitative anisotropy. *PloS One*. 2013; **8(11)**:
465 e80713.

- 466 20. Zalesky A, Fornito A, Bullmore ET. Network-based statistic: Identifying differences in
467 brain networks. *NeuroImage*. 2010; **53(4)**:1197-207.
- 468 21. Akiba T, Sano S, Yanase T, Ohta T, Koyama Masanori. Optuna: A next-generation
469 hyperparameter optimization framework. Proceedings of the 25th ACM SIGKDD
470 international conference on knowledge discovery & data mining; Anchorage AK USA,
471 NY USA: Association for Computing Machinery; 2019 Aug. 2623 p.
- 472 22. Hayes AF. PROCESS: A versatile computational tool for observed variable mediation,
473 moderation, and conditional process modeling. 2012; Retrieved from
474 <https://www.afhayes.com/public/process2012.pdf>
- 475 23. Laurens VDM, Hinton G. Visualizing Data using t-SNE. *Journal of Machine Learning*
476 *Research*. 2008; **9(2605)**:2579-605.
- 477 24. Casanova MF. The neuropathology of autism. *Brain pathology*. 2007; **17(4)**:422-33.
- 478 25. Middleton FA, Strick PL. Cerebellar projections to the prefrontal cortex of the primate.
479 *Journal of Neuroscience*. 2001; **21(2)**:700-12.
- 480 26. Johnson B, Stanley-Cary C, Fielding J, et al. *Cerebellum and the Psychopathology of*
481 *Autism and Asperger's Disorder*. New York: Springer; 2014. 845 p.
- 482 27. Kringelbach ML. The human orbitofrontal cortex: linking reward to hedonic experience.
483 *Nature Reviews Neuroscience*. 2005; **6**:691-702.
- 484 28. Jin P, Wang Y, Li Y, *et al*. The fair decision-making of children and adolescents with
485 high-functioning autism spectrum disorder from the perspective of dual-process
486 theories. *BMC Psychiatry*. 2020 Apr 6; **20(1)**:152.
- 487 29. Scott-Van Zeeland AA, Dapretto M, Ghahremani DG, et al. Reward processing in
488 autism. *Autism Research*. 2010; **3(2)**:53-67.
- 489 30. Badre D, Nee DE. Frontal cortex and the hierarchical control of behavior. *Trends in*
490 *Cognitive Science*. 2018; **22(2)**:170-88.

- 491 31. Fuster JM. The prefrontal cortex – an update: Time is of the essence. *Neuron*. 2011;
492 **30**:319-33.
- 493 32. Turken AU, Dronkers NF. The neural architecture of the language comprehension
494 network: converging evidence from lesion and connectivity analyses. *Front Syst*
495 *Neurosci*. 2011;**1**.
- 496 33. Harrison, L.A., Kats, A., Kilroy, E. et al. Motor and sensory features successfully
497 decode autism spectrum disorder and combine with the original RDoC framework to
498 boost diagnostic classification. *Scientific Reports*. 2021; **11**:7839.
- 499 34. Kjellmer L, Hedvall A, Holm A, et al. Language comprehension in preschoolers with
500 autism spectrum disorders without intellectual disability: Use of the Reynell
501 Developmental Language Scales. *Research in Autism Spectrum Disorders*. 2012; **6(3)**:
502 1119-25.
- 503 35. Vlamings P, Jonkman LM, Daalen EV, et al. Basic abnormalities in visual processing
504 affect face processing at an early age in autism spectrum disorder. *Biological*
505 *Psychiatry*. 2010; **68(12)**:1107-13.
- 506 36. Yi HL, Licari M, Spittle AJ, et al. Early motor function of children with autism
507 spectrum disorder: A systematic review. *Pediatrics*. 2021; **147(2)**:e2020011270.
- 508 37. Lynch CJ, Uddin LQ, Supekar K, et al. Default mode network in childhood autism:
509 posteromedial cortex heterogeneity and relationship with social deficits. *Biological*
510 *Psychiatry*. 2013; **74(3)**:212-9.
- 511 38. Souchay C, Ohlsson M, Zalla T. Autobiographical memory and theory of mind in
512 autism spectrum disorder. Atrium: WILEY Blackwell; 2018. 92 p. (Johnson JL,
513 Goodman GS, Mundy PC. *The Wiley handbook of memory, autism spectrum disorder,*
514 *and the law*; vol. 5).

- 515 39. Zhao W, Luo L, Li Q, Kendrick KM. What can psychiatric disorders tell us about
516 neural processing of the self. *Front Hum Neurosci.* 2013; **7**:485.
- 517 40. Livingston LA, Shah P, Milner V, et al. Quantifying compensatory strategies in adults
518 with and without diagnosed autism. *Molecular Autism.* 2020; **11**:15.
- 519

520 **Table 1.** Subject demographics and behavioral measures of the three datasets.

	TD			ASD		Statistics (<i>t</i> or χ^2)	Significance (<i>p</i>)
	N	Mean (SD)		N	Mean (SD)		
Discovery (Beijing) dataset							
Age, years	26	4.70 (0.46)		93	4.58 (0.55)	-1.07 ^a	0.29
Gender (boys: girls)	26	20:6		93	85:8	2.83 ^b	0.09
BMI	26	15.64 (1.68)		76	16.03 (1.77)	0.87 ^a	0.39
Handedness (R: D: L)	26	20:6:0		90	78:9:3	3.76 ^b	0.15
Head Circumference, cm	26	51.24 (1.48)		90	51.79 (1.72)	1.49 ^a	0.14
ADOS social interaction sub-scale score		86	9.88 (2.54)
ADOS communication sub-scale score		86	5.57 (1.87)
ADOS total score (social interaction + communication)		86	15.45 (4.02)
GDS total score	26	94.39 (10.1)		82	64.34 (17.1)	8.50 ^a	<0.001
ASD validation (Chengdu) dataset							
Age, years		12	4.98 (1.22)
Gender (boys: girls)		12	11:1
ADOS-2 total score (social affect + restricted and repetitive behavior)		12	18.58 (4.46)
TD validation (Nanjing) dataset							
Age, years	9	5.44 (0.84)	
Gender (boys: girls)	9	6:3	

521 TD: typically developing; ASD: autism spectrum disorder; BMI: body mass index; R: right handed; D:

522 double handed; L: left handed; ADOS: Autism Diagnostic Observation Schedule; ADOS-2:

523 Second Edition of ADOS; GDS: Gesell Developmental Scale; ^aIndependent two-sample *t* test, *t*

524 score; ^bChi-square test, χ^2 . SD: standard deviations.

525

526 **Table 2.** Profiles of the 33 structural connections with increased FA in children with ASD.

No.	Network	Region name	MNI coordinates			Region name	MNI coordinates		
			X	Y	Z		X	Y	Z
1	1	STG_L	-55	-3	-10	PCun_L	-12	-67	25
2		IPL_L	-47	-65	26	PCun_L	-12	-67	25
3		SFG_L	-18	24	53	MVOcC_L	-13	-68	12
4		STG_L	-55	-3	-10	MVOcC_L	-13	-68	12
5	2	SFG_R	20	4	64	IFG_R	45	16	25
6		SFG_R	20	4	64	PoG_L	-21	-35	68
7		SFG_R	20	4	64	Right_IX	6	-54	-50
8	3	IFG_R	48	35	13	LOcC_R	32	-85	-12
9		IFG_R	54	24	12	LOcC_R	32	-85	-12
10		MFG_R	42	44	14	MVOcC_R	10	-85	-9
11		IFG_R	48	35	13	MVOcC_R	10	-85	-9
12		IFG_R	51	36	-1	MVOcC_R	10	-85	-9
13		IFG_R	51	36	-1	BG_R	22	8	-1
14		IFG_R	51	36	-1	BG_R	14	5	14
15		IFG_R	48	35	13	STG_R	47	12	-20
16		IFG_R	54	24	12	INS_R	36	18	1
17		SFG_R	7	-4	60	Tha_R	12	-14	1
18	IFG_R	48	35	13	Tha_R	12	-14	1	
19	4	MFG_L	-41	41	16	BG_L	-14	2	16
20		MFG_L	-41	41	16	Tha_L	-7	-12	5
21		IFG_L	-53	23	11	BG_L	-14	2	16
22		MFG_L	-41	41	16	CG_R	5	41	6
23		MFG_L	-26	60	-6	Tha_L	-7	-12	5
24		Tha_L	-7	-12	5	Left_IX	-8	-54	-48
25	5	OrG_R	23	36	-18	OrG_R	6	57	-16
26		OrG_R	23	36	-18	OrG_R	9	20	-19
27		OrG_L	-6	52	-19	OrG_L	-10	18	-19
28		OrG_R	6	47	-7	OrG_R	9	20	-19
29		OrG_L	-7	54	-7	CG_L	-4	39	-2
30		OrG_R	6	47	-7	CG_L	-4	39	-2
31		OrG_L	-10	18	-19	CG_L	-4	39	-2
32		OrG_R	6	47	-7	BG_R	15	14	-2
33		OrG_R	6	57	-16	BG_R	15	14	-2

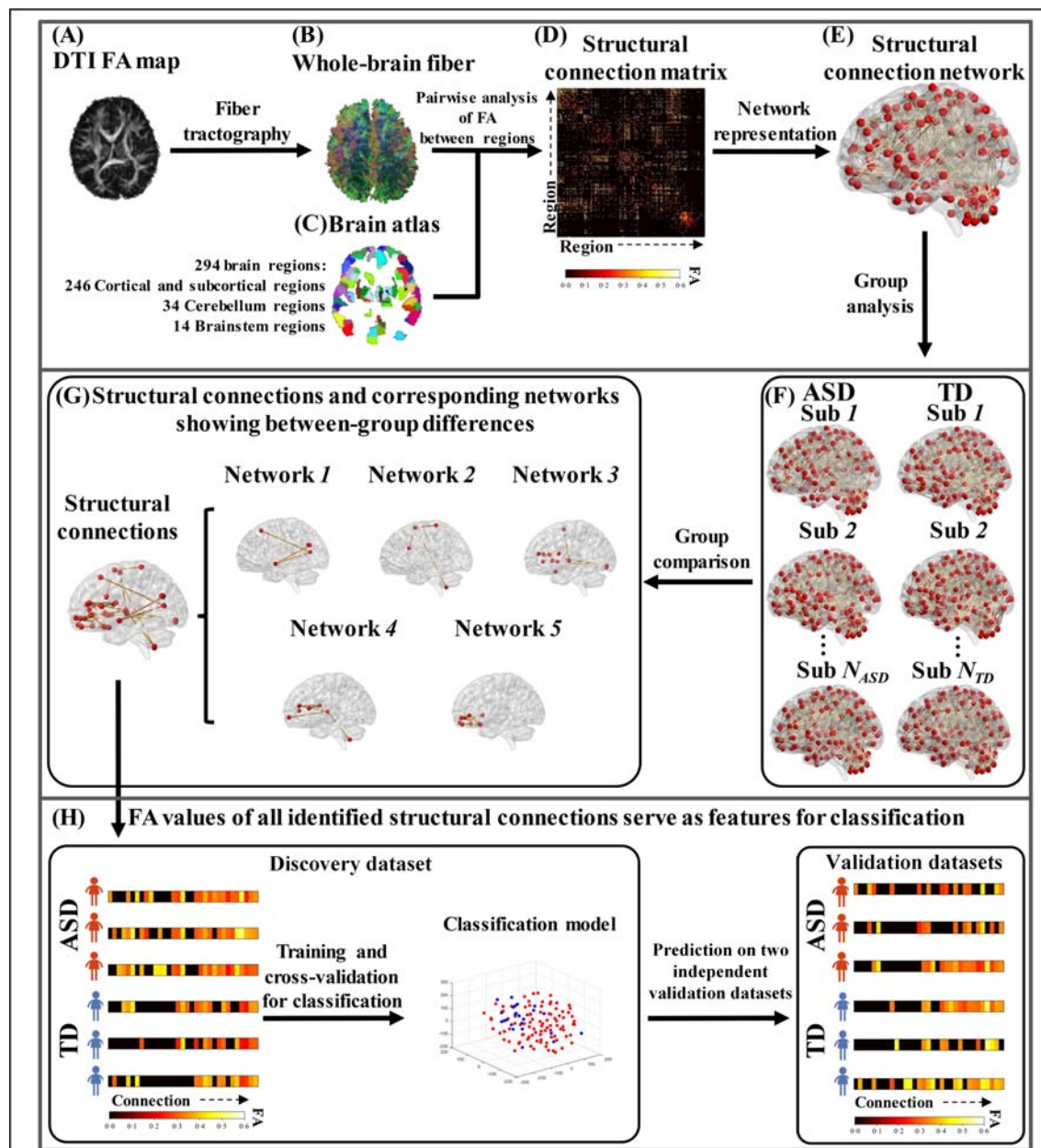
527 Abbreviations: L, left; R, right; STG, superior temporal gyrus; PCun, precuneus; IPL, inferior parietal

528 lobule; SFG, superior frontal gyrus; MVOcC, medioventral occipital cortex; IFG, inferior frontal

529 gyrus; PoG, postcentral gyrus; LOcC, lateral occipital cortex; Tha, thalamus; MFG, middle

530 frontal gyrus; OrG, orbitofrontal gyrus; CG, cingulate gyrus; BG, basal ganglia; INS, insula.

531



532

533

Figure 1 Flow chart of ASD identification framework. (A) The pre-processed DTI FA map. **(B)**

534

Reconstructed whole-brain streamline fibers. **(C)** Young children's brain atlas. **(D)** Structural

535

connection matrix for each participant. **(E)** The individual structural connection network. **(F)** All

536

individual structural connection networks in ASD and TD groups. **(G)** The structural connections and

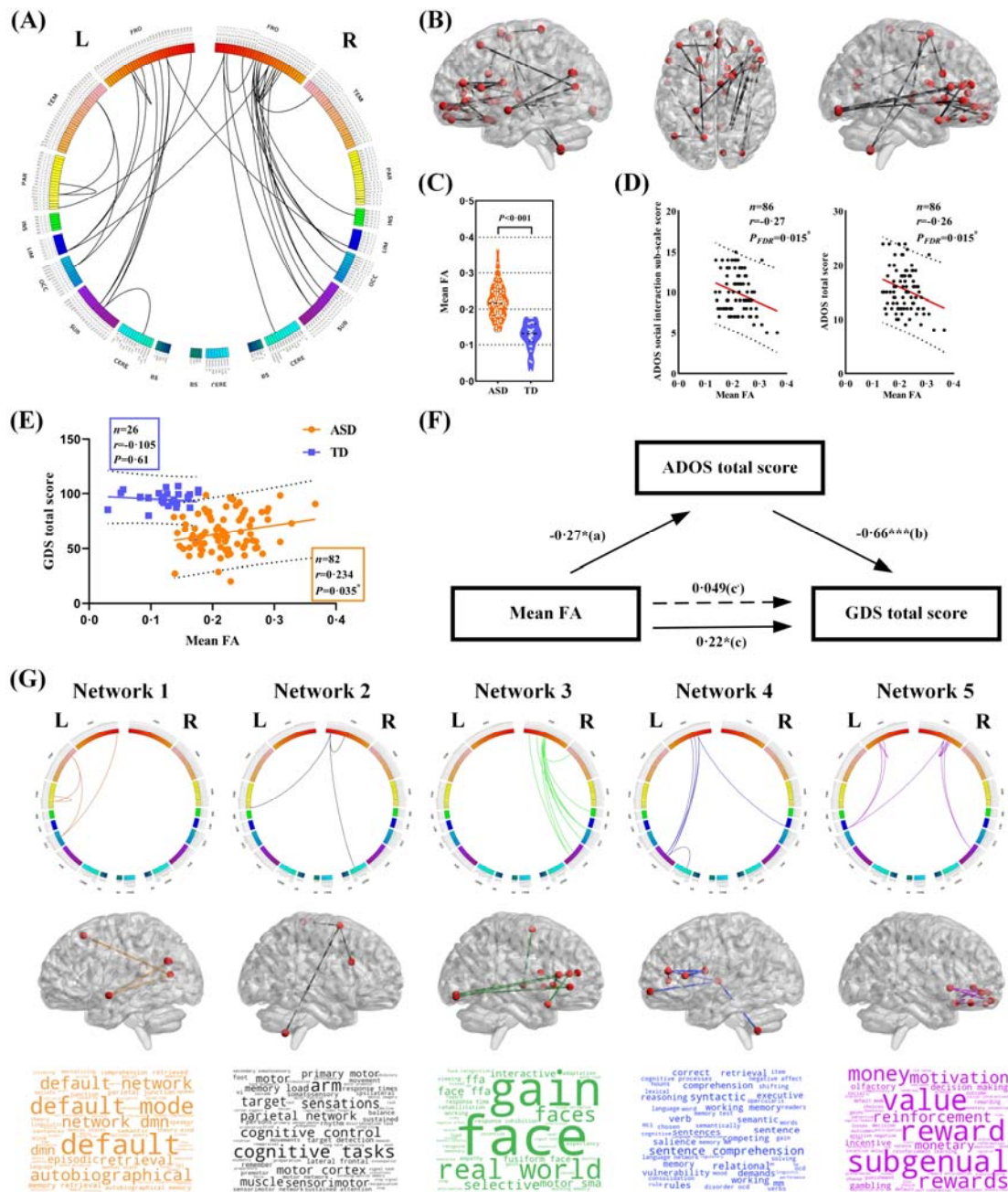
537

associated networks showing differences between ASD and TD groups. **(H)** Pattern classification

538

between ASD and TD using all structural connections showing between-group differences.

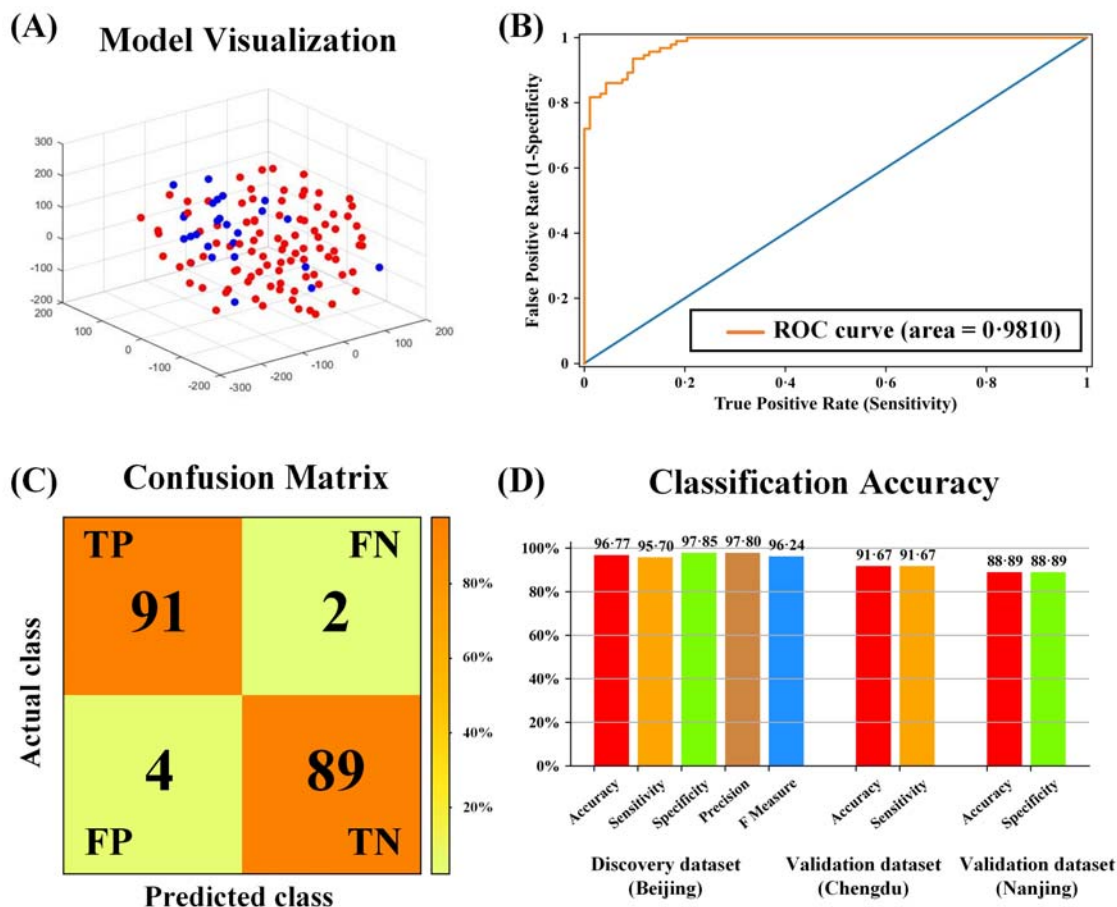
539



540

541 **Figure 2 Increased FA connections and networks in ASD compared to TD children.** (A) The 33
 542 increased FA values of structural connections in ASD in circos plot. Abbreviations: L, left hemisphere;
 543 R, right hemisphere; FRO, frontal lobe; TEM, temporal lobe; PAR, parietal lobe; INS, insular lobe;
 544 LIM, limbic lobe; OCC, occipital lobe; SUB, subcortical nuclei; CERE, cerebellum; BS, brain stem.
 545 (B) Locations of 33 connections on the cortical surface. (C) Averaged FA value of 33 connections in
 546 ASD and TD groups. (D) Correlations between averaged FA value of the 33 connections in ASD and
 547 ADOS ‘social interaction’ sub-scale score and total (‘communication’ + ‘social interaction’) score. (E)
 548 Correlations between averaged FA value of the 33 connections and ADOS social communication and
 549 total scores and GDS total score. (F) Mediation analysis between averaged FA, GDS total score and

550 ADOS total score (path $a=-0.27$, $p=0.018$; path $b=-0.66$, $p<0.001$; path $c=0.22$, $p=0.047$; path
551 $c'=0.049$, $p=0.58$). (G) The 33 increased connections were further categorized into 5 structural
552 networks. In each network, row 1-3 shows the structural connections in circo plot, locations of the
553 connections on cortical surface, and the word cloud of functional annotation via meta-analysis
554 respectively.
555



556

557 **Figure 3 Classification Accuracy between ASD and TD.** (A) Classification model in a 3-
 558 dimensional feature space after performing dimensionality reduction using the t-distributed stochastic
 559 neighbor embedding (t-SNE) algorithm. Red and blue dots represent ASD and TD subjects,
 560 respectively. (B) The Receiver Operating Characteristic (ROC) curve of the training model. (C) The
 561 confusion matrix of the training model. The colorbar represents the proportion of correctly classified
 562 subjects among all subjects. (D) The detailed classification accuracy metrics including accuracy,
 563 sensitivity, specificity, precision, and F measure in both discovery and validation datasets.

564