Supplementary Information

Part 1 Sites and participants

According to the site-specific descriptions in ABIDE, we merged data from ABIDE I and II that originated from the same institution, used the same scanner, and employed identical scanning parameters, provided that the samples were non-overlapping. For example, data from the Kennedy Krieger Institute in ABIDE I were combined with those from the same institute in ABIDE II. This process yielded 27 independent sites. After excluding five sites consisting solely of male participants, we conducted preprocessing and quality assessment on the neuroimaging data from the remaining 22 sites. The site and participant screening process is illustrated in Figure S1.

A total of 492 participants were excluded due to failing quality control (QC) criteria: 11 participants (ASD: $n_{female} = 1$, $n_{male} = 7$; TD: $n_{female} = 1$, $n_{male} = 2$) lacked corresponding T1-weighted images; 289 participants exhibited excessive head motion (i.e., maximum displacement > 3 mm or 3 degrees, or mean framewise displacement > 0.5 mm; ASD: $n_{female} = 28$, $n_{male} = 137$; TD: $n_{female} = 25$, $n_{male} = 99$); and 192 participants showed abnormal spatial normalization (ASD: $n_{female} = 23$, $n_{male} = 90$; TD: $n_{female} = 30$, $n_{male} = 49$). Additionally, 93 individuals with a Full-Scale IQ (FSIQ) below 70 were excluded (ASD: $n_{female} = 6$, $n_{male} = 37$; TD: $n_{female} = 18$, $n_{male} = 32$). Finally, a second-level screening was conducted on QC-passed sites, and independent sites with fewer than five autistic females and five typically developing (TD) females were excluded from further analysis.

Seven hundred ninety-one participants were retained following rigorous selection

procedures. For the nodal- and edge-level analyses, 196 participants were selected—comprising 49 autistic females, 49 autistic males, 49 TD females, and 49 TD males—with groups matched on age, sex, and FSIQ (Table 1). For the state-level variability analysis, the four groups were additionally matched on repetition time (TR) alongside age, sex, and FSIQ (Table S1). In the validation analyses, all included participants (*N* = 791, Table S2) were subjected to nodal-level variability analyses to verify the robustness of the main findings.

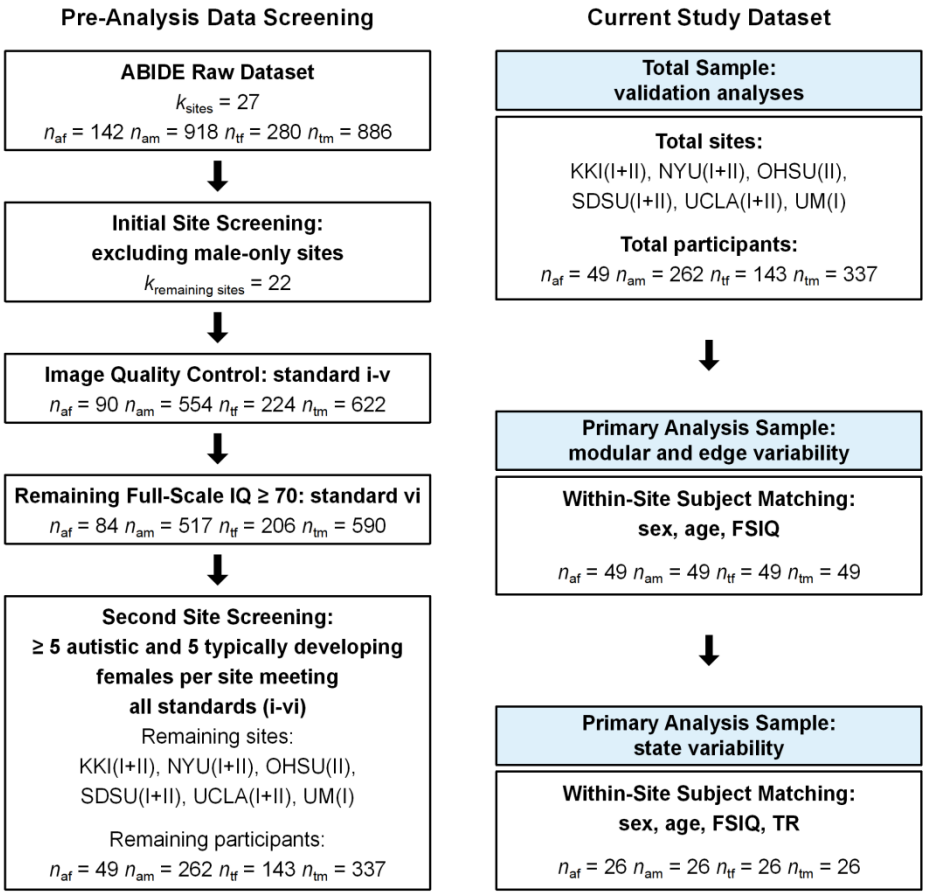

Figure S1 Flowchart of site and participant inclusion criteria. af: autistic female; am: autistic male; tf: TD female; tm: TD male; FSIQ: full-scale IQ; TR = repetition time; KKI: Kennedy Krieger Institute, TR = 2500 ms; NYU: NYU Langone Medical

Center, TR = 2000 ms; OHSU: Oregon Health and Science University, TR = 2500 ms; SDSU: San Diego State University, TR = 2000 ms; UCLA: University of California, Los Angeles, TR = 3000 ms; UM: University of Michigan, TR = 2000 ms.

Table S1 Demographic information, comparisons and core symptom performances of participants (N = 104).

	ASD		**TD**		**Contrasts**			
	female (n = 26)	male (n = 26)	female (n = 26)	male (n = 26)	am vs. af	tm vs. tf	am vs. tm	af vs. tf
Age (years)	15.94 (7.80)	15.37 (5.85)	15.07 (5.40)	15.76 (4.76)	0.766	0.629	0.794	0.641
range	5.22-38.76	6.70-29.98	8.04-29.13	6.36-30.08				
children (5-12 years)	11	9	9	7				
adolescents (13-17 years)	8	12	13	15				
adults (18+ years)	7	5	4	4				
FSIQ	102.54 (17.51)	101.77 (15.09)	104.17 (12.26)	102.02 (13.50)	0.866	0.550	0.950	0.698
Mean FD	0.17 (0.09)	0.16 (0.07)	0.13 (0.07)	0.15 (0.09)	0.500	0.365	0.693	0.045
ADOS-2 Total: af/am (n = 21/21)	11.24 (3.75)	12.19 (4.79)			0.478			
ADOS-2 Severity: af/am (n = 21/21)	6.48 (1.89)	6.90 (2.12)			0.493			
ADOS-2 Social Affect: af/am (n = 19/21)	8.00 (2.75)	9.24 (4.05)			0.270			
ADOS-2 RRB: af/am (n = 19/21)	2.79 (1.55)	2.95 (1.43)			0.731			
SRS Total: ASD/TD (n = 33/19)	101.18 (19.65)	85.81 (33.06)	20.75 (16.22)	16.91 (9.08)	0.120	0.518	< 0.001	< 0.001

Note. The number of participants in each age group is shown below the age range; the p-values of two-sample t-tests are presented in the 'Contrasts' columns. FSIQ: full-scale IQ; FD: framewise displacement; ADOS-2: Autism Diagnostic Observation Schedule, 2nd Edition; RRB:

Restricted and repetitive behaviors; SRS: Social Responsiveness Scale; af: autistic female; am: autistic male; tf: typically developing female; tm: typically developing male.

Table S2 Demographic information, comparisons and core symptom performances of participants (N = 791).

	ASD		**TD**		**Contrasts**			
	female (n = 49)	male (n = 262)	female (n = 143)	male (n = 337)	am vs. af	tm vs. tf	am vs. tm	af vs. tf
Age (years)	13.46 (6.46)	12.52 (4.40)	11.94 (3.94)	12.80 (4.43)	0.332	0.045	0.447	0.125
range	5.22-38.76	5.32-39.10	7.76-29.13	5.89-31.78				
children (5-12 years)	31	159	104	218				
adolescents (13-17 years)	11	87	29	90				
adults (18+ years)	7	16	10	29				
FSIQ	104.00 (16.35)	105.23 (15.95)	112.30 (11.80)	112.61 (12.12)	0.622	0.792	< 0.001	0.002
Mean FD	0.22 (0.12)	0.20 (0.10)	0.16 (0.10)	0.17 (0.09)	0.195	0.529	< 0.001	0.004
ADOS-2 Total: af/am (n = 36/216)	11.17 (3.24)	12.32 (4.47)			0.140			
ADOS-2 Severity: af/am (n = 36/216)	6.56 (1.65)	7.01 (2.02)			0.203			
ADOS-2 Social Affect: af/am (n = 34/213)	8.24 (2.58)	9.22 (3.71)			0.060			
ADOS-2 RRB: af/am (n = 34/213)	2.68 (1.45)	3.13 (1.74)			0.148			
SRS Total: ASD/TD (n = 209/291)	99.44 (21.30)	92.19 (29.57)	17.37 (11.68)	19.07 (11.73)	0.096	0.251	< 0.001	< 0.001

Note. The number of participants in each age group is shown below the age range; the p-values of two-sample t-tests are presented in the 'Contrasts' columns. FSIQ: full-scale IQ; FD: framewise displacement; ADOS-2: Autism Diagnostic Observation Schedule, 2nd Edition; RRB:

Restricted and repetitive behaviors; SRS: Social Responsiveness Scale; af: autistic female; am: autistic male; tf: typically developing female; tm: typically developing male.

Part 2 Supplementary results

2.1 Modular variability

ANCOVA results showed significantly higher modular variability in the right rolandic operculum, left medial superior frontal gyrus, right gyrus rectus, right postcentral gyrus, left angular gyrus, right Heschl's gyrus and right superior temporal gyrus of individuals with autism spectrum disorders (ASD) compared to their TD counterparts. However, the bilateral middle frontal gyrus, left orbital medial frontal gyrus, and bilateral inferior parietal lobes showed significantly lower modular variability in ASD than in TD. Notably, only the diagnosis effect on the modular variability of the left middle frontal gyrus remained significant after applying FDR correction (Figure S2, Table S3).

In addition, females exhibited greater modular variability in the left thalamus and right temporal pole of the superior temporal gyrus, but lower modular variability in the right parahippocampal gyrus and left cuneus compared to males. However, the sex effects on modular variability at all mentioned nodes did not remain significant after correcting for multiple comparisons (Figure S2, Table S3).

Lastly, the interaction effects between diagnosis and sex on modular variability were significant in three nodes. Specifically, autistic females and TD males showed higher modular variability in the right supramarginal gyrus, while demonstrating lower modular variability in the right rolandic operculum and right Heschl's gyrus compared to autistic males and TD females. However, these significant effects did not withstand FDR correction (Figure S2, Table S3).

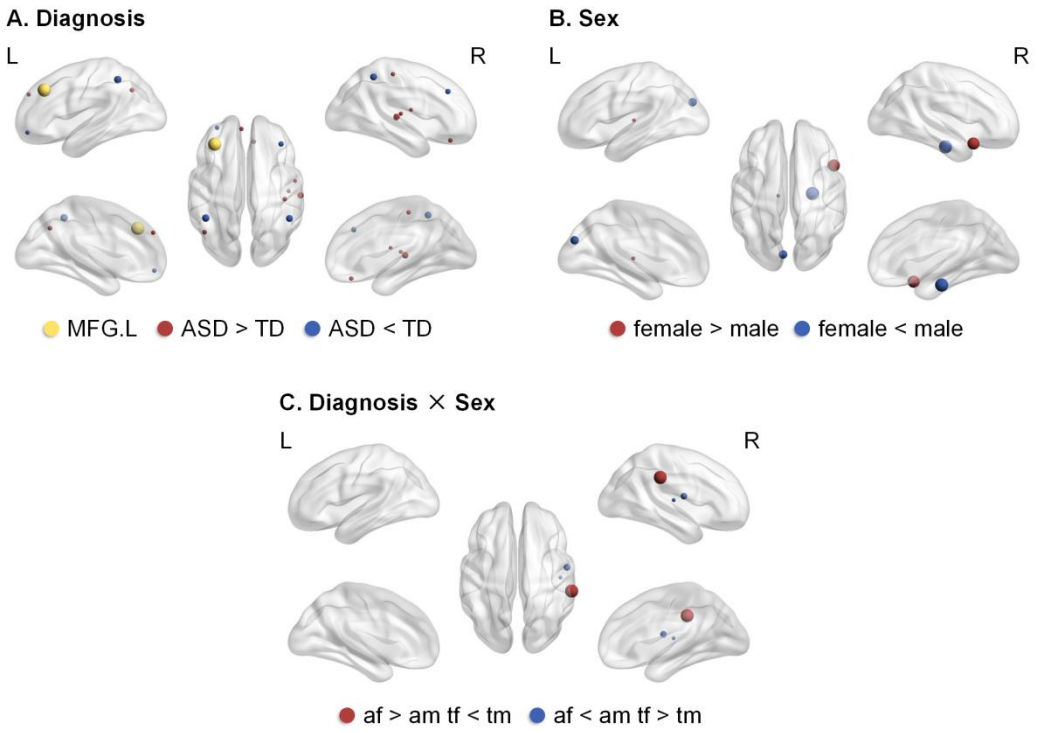

Figure S2 Brain maps illustrating the effects of diagnosis (A), sex (B), and sex-by-diagnosis interaction (C) on modular variability (N = 196). L: the left hemisphere; R: the right hemisphere; af: autistic female; am: autistic male; tf: typically developing female; tm: typically developing male.

Table S3 The significant diagnosis, sex, and sex-by-diagnosis interaction effects on modular variability (N = 196).

	$F(1,189)$	p	p_{FDR}	partial η^2
Diagnosis				
MFG.L	**12.64**	**0.0005**	**0.043**	**0.063**
MFG.R	5.61	0.019	0.286	0.029
ORBmid.L	4.78	0.030	0.286	0.025
ROL.R	4.21	0.042	0.311	0.022
SFGmed.L	4.68	0.032	0.286	0.024
REC.R	4.85	0.029	0.286	0.025
PoCG.R	4.77	0.030	0.286	0.025
IPL.L	7.49	0.007	0.239	0.038
IPL.R	7.00	0.009	0.239	0.036
ANG.L	4.92	0.028	0.286	0.025
HES.R	4.40	0.037	0.304	0.023
STG.R	6.66	0.011	0.239	0.034
Sex				
PHG.R	6.18	0.014	0.582	0.032
CUN.L	5.25	0.023	0.582	0.027
THA.L	4.33	0.039	0.582	0.022
TPOsup.R	5.98	0.015	0.582	0.031
Diagnosis × Sex				
ROL.R	4.59	0.034	0.937	0.024
SMG.R	5.53	0.020	0.937	0.028
HES.R	4.19	0.042	0.937	0.022

Note. FDR: false discovery rate; MFG: middle frontal gyrus; ORBmid: orbital part of the middle frontal gyrus; ROL: rolandic operculum; SFGmed: medial superior frontal gyrus; REC: gyrus rectus; PoCG: postcentral gyrus; IPL: inferior parietal lobe; ANG: angular gyrus; HES: Heschl's gyrus; STG: superior temporal gyrus; PHG: parahippocampal gyrus; CUN: cuneus; THA: thalamus; TPOsup: temporal pole in the superior temporal gyrus; SMG: supramarginal gyrus; L: in the left hemisphere; R: in the right hemisphere.

2.2 Edge variability

Brain maps depicting the connectivity between the right olfactory cortex (OLF.R) and the right paracentral lobule (PCL.R), as well as the connectivity between the left amygdala (AMYG.L) and the right anterior cingulate and paracingulate gyri (ACG.R), are presented in Figure S3.

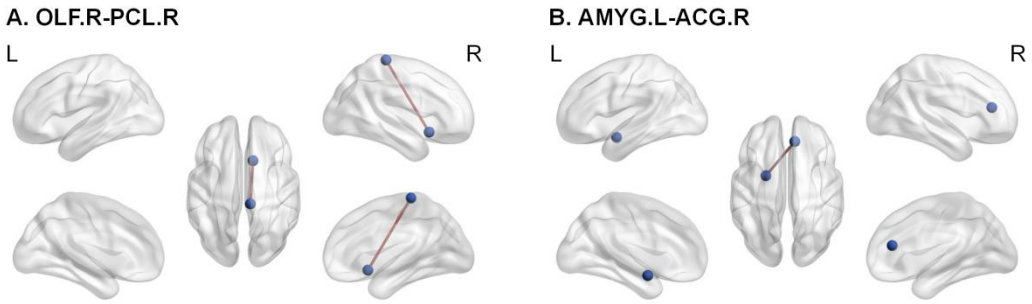

Figure S3 Brain maps of edge-connected regions with significant edge variability. (A) The functional connection between the right olfactory cortex (OLF.R) and the right paracentral lobule (PCL.R). (B) The functional connection between the left amygdala (AMYG.L) and the right anterior cingulate and paracingulate gyri (ACG.R). L: the left hemisphere; R: the right hemisphere.

2.3 State variability

Four distinct brain states were identified based on the functional connectivity matrices of all participants. The number and proportions of participants in each state, along with the median and interquartile range of the four time-varying indices are shown in Table S4 and Table S5.

We found two significant sex-by-diagnosis interaction effects on transition probability and fractional time. The first interaction effect was observed in the probability of the transition from State 3 to State 3 ($F = 11.74$, $p_{FDR} = 0.014$, power = 0.93). *Post hoc* comparisons indicated that autistic females exhibited a lower transition probability for State 3-3 than both autistic males ($p_{post\ hoc} = 0.008$) and TD females ($p_{post\ hoc} = 0.014$). Similar probabilities of State 3 transitions were observed in other contrasts (TD males vs. TD females: $p_{post\ hoc} = 0.067$; ASD males vs. TD males: $p_{post\ hoc} = 0.082$). Another significant sex-by-diagnosis interaction effect was found on the fractional time spent in State 3 ($F = 4.44$, $p = 0.038$), however, it did not remain significant after FDR correction ($p_{FDR} = 0.151$). No significant effects of diagnosis, sex, or sex-by-diagnosis interaction were observed for other state variability indices ($ps > 0.054$).

Moreover, a moderate negative correlation was observed in autistic males between the transition probability of State 3-3 and SRS total scores ($r = -0.560$, $p = 0.047$, $p_{FDR} = 0.233$). SRS total scores reflect the severity of autistic symptoms in individuals with ASD, and females diagnosed with ASD tend to present more severe symptoms than males with ASD. Since autistic females ($N = 17$, $M_{raw} = 101.18$) in our

study had higher SRS total scores compared to autistic males (N = 16, M_{raw} = 85.81), despite no statistically significant difference between the groups (t = 1.61, p = 0.120), we re-examined the sex-by-diagnosis interaction effect on the transition probability of State 3-3 by controlling for SRS total scores in a sample limited to individuals with both behavioral phenotypes and brain imaging (Table S1). The significant effect of the sex-by-diagnosis interaction on the maintenance probability of State 3 disappeared after controlling for SRS total scores (F = 3.83, p = 0.057, power = 0.53, Table S6).

Table S4 Descriptive information of brain states in different groups.

	State 1	State 2	State 3	State 4
Number (Proportion)				
ASD	**47 (45.19%)**	**29 (27.88%)**	**32 (30.77%)**	**33 (31.73%)**
female	24 (23.08%)	14 (13.46%)	21 (20.19%)	17 (16.35%)
male	23 (22.12%)	15 (14.42%)	11 (10.58%)	16 (15.38%)
TD	**46 (44.23%)**	**30 (28.85%)**	**35 (33.65%)**	**41 (39.42%)**
female	22 (21.15%)	11 (10.58%)	15 (14.42%)	21 (20.19%)
male	24 (23.08%)	19 (18.27%)	20 (19.23%)	20 (19.23%)
Fractional Time				
ASD	**0.34 (0.51)**	**0.05 (0.29)**	**0.09 (0.29)**	**0.14 (0.36)**
female	0.40 (0.58)	0.03 (0.23)	0.21 (0.28)	0.11 (0.33)
male	0.31 (0.39)	0.12 (0.43)	0.00 (0.24)	0.17 (0.59)
TD	**0.31 (0.53)**	**0.05 (0.35)**	**0.13 (0.35)**	**0.23 (0.38)**
female	0.24 (0.50)	0.00 (0.14)	0.10 (0.38)	0.33 (0.49)
male	0.34 (0.52)	0.24 (0.38)	0.14 (0.34)	0.15 (0.32)
Mean Dwell Time				
ASD	**16.33 (26.95)**	**5.00 (16.95)**	**9.00 (22.38)**	**10.00 (25.75)**
female	16.92 (38.13)	3.50 (15.45)	13.33 (22.08)	10.00 (21.63)
male	15.63 (17.75)	7.25 (17.13)	0.00 (17.13)	8.50 (28.33)
TD	**17.00 (30.75)**	**6.00 (21.75)**	**9.75 (29.50)**	**15.50 (30.50)**
female	13.92 (35.69)	0.00 (10.75)	9.50 (30.08)	19.33 (33.50)
male	18.50 (28.38)	15.75 (29.50)	9.75 (20.17)	8.00 (30.13)

Note. This table comprises three sections providing information about brain states across different groups. The first section reports the number and proportions of participants in each state. The second and third sections respectively present the median and interquartile range of fractional time and mean dwell time for each state.

Table S5 The median and interquartile range for the number of transitions and the transition probability.

	ASD		TD	
Indices	**female**	**male**	**female**	**male**
Number of Transitions	4 (4)	5 (4)	3.5 (3)	4 (3)
Transition Probability				
State 1-1	0.97 (0.05)	0.95 (0.07)	0.96 (0.09)	0.97 (0.08)
State 1-2	0.00 (0.002)	0.00 (0.05)	0.00 (0.02)	0.00 (0.02)
State 1-3	0.01 (0.04)	0.00 (0.02)	0.00 (0.00)	0.00 (0.02)
State 1-4	0.00 (0.02)	0.00 (0.02)	0.00 (0.03)	0.00 (0.02)
State 2-1	0.005 (0.07)	0.00 (0.04)	0.00 (0.02)	0.00 (0.03)
State 2-2	0.98 (0.09)	0.98 (0.06)	1.00 (0.04)	0.97 (0.06)
State 2-3	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.02)
State 2-4	0.00 (0.00)	0.00 (0.01)	0.00 (0.00)	0.00 (0.004)
State 3-1	0.00 (0.04)	0.00 (0.00)	0.00 (0.01)	0.00 (0.03)
State 3-2	0.00 (0.01)	0.00 (0.00)	0.00 (0.00)	0.00 (0.02)
State 3-3	**0.97 (0.06)**	**1.00 (0.03)**	**0.99 (0.03)**	**0.97 (0.09)**
State 3-4	0.00 (0.01)	0.00 (0.00)	0.00 (0.02)	0.00 (0.00)
State 4-1	0.00 (0.03)	0.00 (0.03)	0.00 (0.02)	0.00 (0.03)
State 4-2	0.00 (0.00)	0.00 (0.00)	0.00 (0.01)	0.00 (0.00)
State 4-3	**0.00 (0.00)**	**0.00 (0.00)**	**0.00 (0.02)**	**0.00 (0.03)**
State 4-4	0.97 (0.06)	0.98 (0.04)	0.98 (0.05)	0.97 (0.13)

Note. This table consists of two sections that provide information on brain states across different groups. The first section reports the median and interquartile range of the number of transitions. The second section presents the median and interquartile range of transition probabilities between states, including transitions from one state to another or back to itself. State 1-1 = the transition from State 1 to State 1.

Table S6 Rank-based ANCOVA results for sex-by-diagnosis interaction effect on transition probability of State 3-3 with and without controlling for SRS total scores.

Transition Probability: State 3-3	**Diagnosis × Sex**			**Contrasts**			
	F	***p***	p_{FDR}	am vs. af	tm vs. tf	am vs. tm	af vs. tf
***N* = 104**							
Covariates: age, FSIQ, mean FD	11.74	0.001	0.014	**0.008**	0.067	0.082	**0.014**
***N* = 53**							
Covariates: age, FSIQ, mean FD	8.26	0.006	0.098	**0.026**	0.186	0.318	**0.026**
Covariates: age, FSIQ, mean FD, SRS	3.83	0.057	0.908	0.253	0.234	0.152	0.690

Note. The *p*-values of post hoc comparisons are presented in the Contrasts section. SRS: Social Responsiveness Scale; FSIQ: full-scale IQ; FD: framewise displacement; FDR: false discovery rate; af: autistic female; am: autistic male; tf: TD female; tm: TD male.

Part 3 Age-related effects on dynamic functional connectivity (dFC)

3.1 Method

We examined age-related effects on dFC in individuals with ASD at the nodal, edge, and brain-state levels, modeling age as both a continuous and a categorical variable. The dFC was estimated using a sliding-window approach with a window length of 60 seconds.

3.1.1 Age as a continuous variable

Three-way ANCOVAs were performed in R 4.2.2 to assess the main and interaction effects of diagnosis (ASD/TD), sex (female/male), and age on modularity (*Q*-value), nodal modular variability, and edge variability, with FSIQ and mean FD as covariates. Multiple comparisons were corrected using the FDR procedure. Simple effects of significant interactions surviving FDR correction ($p_{FDR} < 0.05$) were evaluated using the emmeans (v1.8.4-1) package (https://CRAN.R-project.org/package=emmeans), and *post hoc* (post-correction) statistical power was estimated with G*Power, version 3.1.9.7.

Rank-based ANCOVAs were performed using the npsm (v2.0.0) package (https://CRAN.R-project.org/package=npsm) to test the main and interaction effects of the categorical variables diagnosis and sex on state variability. As the npsm package does not support continuous predictors, when age was treated as a continuous variable, effects of FSIQ and mean FD were first regressed out from each state variability index. The residuals were then analyzed using rank-based ANOVAs and *post-hoc* comparisons via the ART (v1.0) package (https://CRAN.R-

project.org/package=ART). FDR correction and power analyses were conducted as described above.

3.1.2 Age as a categorical variable

Participants spanned a wide age range ([5.22, 38.76]; Table 1) in the current study, with children (5-12 years; $N = 114$) being the largest group, followed by adolescents (13-17 years; $N = 62$) and adults (18+ years; $N = 20$). Due to the small adult sample, adults and adolescents were combined into a single 'older' subgroup, while children were analyzed separately (i.e., developmental stage: children vs. older).

Modularity (Q-value), nodal modular variability, and edge variability were examined using three-way ANCOVAs in R 4.2.2 (diagnosis: ASD/TD × sex: female/male × age: children/older), with FSIQ and mean FD as covariates. For state variability indices, three-way rank-based ANCOVAs were conducted using the same factors, followed by *post hoc* comparisons. FDR correction, and statistical power estimation were performed as described above.

3.2 Results

At the nodal level, the diagnosis difference in modular variability of the MFG.L was not influenced by age (Table S7). Although no diagnosis differences survived FDR correction when age was treated as a categorical variable, the difference in MFG.L remained the most robust finding ($F(1,186) = 11.02$, partial $\eta^2 = 0.056$, $p = 0.001$, $p_{FDR} = 0.097$, power = 0.92). Furthermore, when age was modelled categorically, modularity in the connectomes of children was significantly lower than

in adolescents and adults ($p_{post\text{-}hoc}$ = 0.021, Table S7). This difference was no longer significant when age was treated as a continuous variable.

At the edge level, the sex-by-diagnosis interaction effect on edge variability of the AMYG.L-ACG.R connection was not affected by age (Table S8). When age was included as a predictor, the diagnosis difference in edge variability of the OLF.R-PCL.R connection did not survive FDR correction (age as a continuous variable: $F(1,186) = 14.46$, partial $\eta^2 = 0.072$, $p = 0.0002$, $p_{FDR} = 0.389$, power = 0.97; age as a categorical variable: $F(1,186) = 12.04$, partial $\eta^2 = 0.061$, $p = 0.0006$, $p_{FDR} = 0.601$, power = 0.94).

At the brain state level, when age was treated categorically, the sex-by-diagnosis interaction effect of State 3-3 transition probability was also unaffected (Table S9). When modeled continuously, this effect did not survive FDR correction ($F = 7.67$, $p = 0.007$, $p_{FDR} = 0.056$, power = 0.79), whereas significant interactions emerged for State 3-1 and State 3-4 transition probabilities (Table S9).

In summary, the effects of age on the core findings of this study were negligible.

Table S7 Age effects on modular variability after FDR correction.

Modular Variability	Age (Continuous)				DS (Chidren vs. Older)			
	$F(1,186)$	p	p_{FDR}	partial η^2	$F(1,186)$	p	p_{FDR}	partial η^2
Diagnosis								
MFG.L	13.31	0.0003	0.031	0.067	—			
Sex	—				—			
Age								
Q	—				5.38	0.021	0.021	0.028
Diagnosis × Sex	—				—			
Diagnosis × Age	—				—			
Sex × Age	—				—			
Diagnosis × Sex × Age	—				—			

Note. "—" indicates that no significant results survived FDR correction. FDR: false discovery rate; DS: developmental stage; MFG.L: the left middle frontal gyrus; Q: modularity.

Table S8 Age effects on edge variability after FDR correction.

Edge Variability	Age (Continuous)				DS (Chidren vs. Older)			
	$F(1,186)$	p	p_{FDR}	partial η^2	$F(1,186)$	p	p_{FDR}	partial η^2
Diagnosis	—				—			
Sex	—				—			
Age	—				—			
Diagnosis × Sex								
AMYG.L-ACG.R	23.93	< 0.001	0.009	0.114	23.29	< 0.001	0.012	0.111
Diagnosis × Age	—				—			
Sex × Age	—				—			
Diagnosis × Sex × Age	—				—			

Note. "—" indicates that no significant results survived FDR correction. FDR: false discovery rate; DS: developmental stage; AMYG: amygdala; ACG: anterior cingulate and paracingulate gyri; L: in the left hemisphere; R: in the right hemisphere.

Table S9 Age effects on state variability after FDR correction.

State Variability Indices	Age (Continuous)			DS (Chidren vs. Older)		
	F	*p*	p_{FDR}	*F*	*p*	p_{FDR}
Diagnosis	—			—		
Sex	—			—		
Age	—			—		
Diagnosis × Sex						
transition probability: State 3-1	11.65	0.0009	0.024	—		
transition probability: State 3-3	—			11.54	0.001	0.025
transition probability: State 3-4	9.47	0.003	0.034	—		
Diagnosis × Age	—			—		
Sex × Age	—			—		
Diagnosis × Sex × Age	—			—		

Note. "—" indicates that no significant results survived FDR correction. FDR: false discovery rate; DS: developmental stage.

Part 4 Validation analyses and results

4.1 Larger sample and varying window lengths

The main findings were validated across different sample sizes and window lengths. At the nodal level, we initially examined modular variability in a larger sample (N = 791) without age, sex and full-scale IQ matching. The dFC was constructed with a window length of 60 seconds. As expected, the results showed a significant diagnosis effect on the modular variability of the left middle frontal gyrus (MFG.L), which also passed the FDR correction (Figure S4, Table S10). The absence of sex and sex-by-diagnosis interaction effects on the modular variability of MFG.L reveals the sex similarities of dFC in autistic brain, consistent with our main results. Secondly, we sought to verify the significant diagnosis effect of MFG.L's modular variability through reconstructing dFC across three window lengths (45, 75 and 90 seconds). Although most results did not survive multiple comparisons, all of them showed significant differences across two samples and three window lengths (Table S11). These results demonstrate the relative robustness of the diagnosis effect on modular variability in MFG.L.

At the edge level, we validated the diagnosis effect on the variability of functional connectivity between the right olfactory cortex (OLF.R) and the right paracentral lobule (PCL.R), as well as the sex-by-diagnosis interaction effect on the variability of functional connectivity between the left amygdala (AMYG.L) and the right anterior cingulate and paracingulate gyri (ACG.R), by reconstructing dFC in three window lengths (45, 75 and 90 seconds). We not only verified the significant

sex-by-diagnosis interaction effect on the edge variability in AMYG.L-ACG.R across all three window lengths but also retained these results after the FDR correction (Table S12). Additionally, all validation findings concerning the edge variability of AMYG.L-ACG.R are consistent with the gender incoherence (GI) model (Table S13). However, regarding the edge variability of OLF.R-PCL.R, the significant diagnosis effect was validated only with a window length of 45 seconds (Table S12). Notably, this was the only window length at which the effect survived multiple comparisons, whereas other window lengths showed similar but uncorrected trends. This result provides evidence for sex similarities of dFC in the autistic brain, since no significant sex or sex-by-diagnosis interaction effects were found in the edge variability of OLF.R-PCL.R (Table S13). The verification findings regarding edge variability indicate that the atypical variability of AMYG.L-ACG.R is a more appropriate clinical biomarker for autistic diagnosis and intervention than the edge variability of OLF.R-PCL.R.

At the state level, we also sought to validate the main findings across three different window lengths (45, 75 and 90 seconds). Nevertheless, significant changes emerged during the identification of state centroids. While a window length of 45 seconds identified four state centroids, both 75-second and 90-second window lengths identified six centroids. These results emphasize the need for cautious interpretation of findings derived from clustering analysis based on sliding-window correlations.

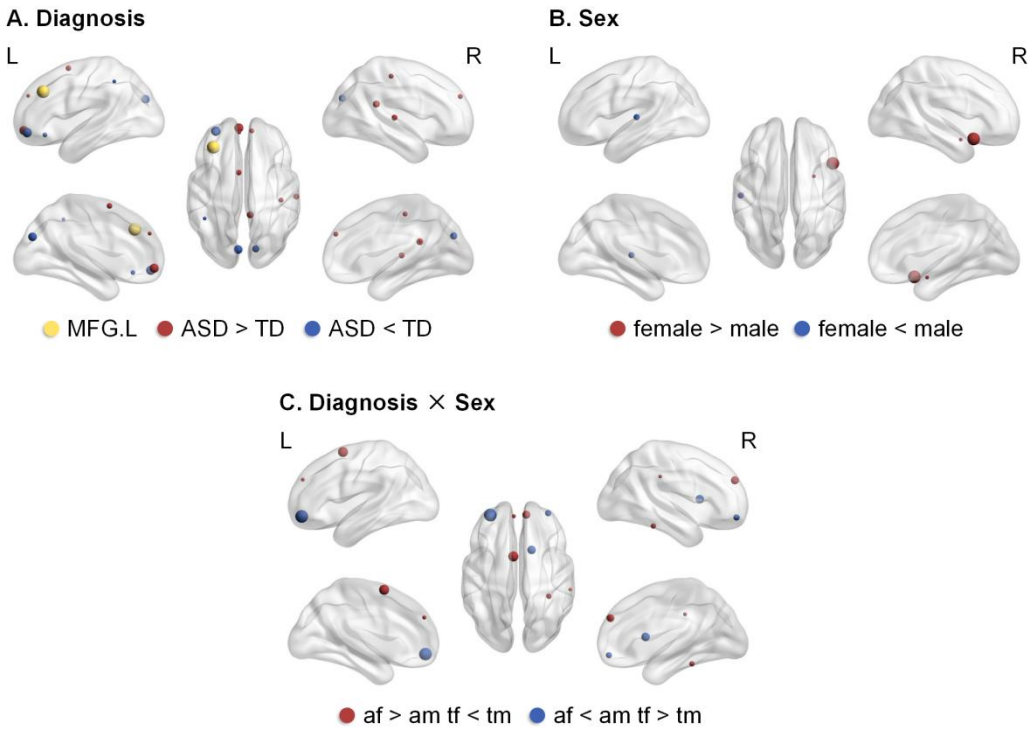

Figure S4 Brain maps illustrating the effects of diagnosis (A), sex (B), and sex-by-diagnosis interaction (C) on modular variability (N = 791). L: the left hemisphere; R: the right hemisphere; af: autistic female; am: autistic male; tf: typically developing female; tm: typically developing male.

Table S10 The significant diagnosis, sex, and sex-by-diagnosis interaction effects on modular variability (N = 791).

	$F(1,784)$	p	p_{FDR}	partial η^2
Diagnosis				
MFG.L	**12.58**	**0.0004**	**0.037**	**0.016**
ORBmid.L	9.49	0.002	0.072	0.012
ORBinf.L	5.02	0.025	0.207	0.006
SMA.L	5.91	0.015	0.172	0.007
SFGmed.L	4.28	0.039	0.292	0.005
SFGmed.R	5.23	0.022	0.202	0.007
ORBsupmed.L	9.14	0.003	0.072	0.012
PCG.R	7.13	0.008	0.116	0.009
CUN.L	8.74	0.003	0.072	0.011
CUN.R	7.28	0.007	0.116	0.009
PoCG.R	5.26	0.022	0.202	0.007
IPL.L	4.11	0.043	0.298	0.005
STG.R	6.20	0.013	0.167	0.008
Sex				
AMYG.R	4.32	0.038	0.714	0.005
STG.L	5.15	0.023	0.714	0.007
TPOsup.R	7.36	0.007	0.630	0.009
Diagnosis × Sex				
ORBmid.L	5.42	0.020	0.506	0.007
ORBmid.R	4.34	0.037	0.506	0.006
SMA.L	4.99	0.026	0.506	0.006
SFGmed.L	3.98	0.046	0.506	0.005
SFGmed.R	4.64	0.032	0.506	0.006
FFG.R	4.20	0.041	0.506	0.005
SMG.R	3.93	0.048	0.506	0.005
CAU.R	4.72	0.030	0.506	0.006

Note. FDR: false discovery rate; MFG: middle frontal gyrus; ORBmid: orbital part of the middle frontal gyrus; ORBinf: orbital part of the inferior frontal gyrus; SMA: supplementary motor area; SFGmed: medial superior frontal gyrus; ORBsupmed: medial orbital superior frontal gyrus; PCG: posterior cingulate gyrus; CUN: cuneus; PoCG: postcentral gyrus; IPL: inferior parietal lobe; STG: superior temporal gyrus; AMYG: amygdala; TPOsup: temporal pole in the superior temporal gyrus; FFG:

fusiform gyrus; SMG: supramarginal gyrus; CAU: caudate nucleus; L: in the left hemisphere; R: in the right hemisphere.

Table S11 ANCOVA results for diagnosis effects on modular variability in MFG.L across various window lengths and sample sizes.

MFG.L	***N* = 196**				***N* = 791**			
Window Length (s)	*F*	*p*	p_{FDR}	partial η^2	*F*	*p*	p_{FDR}	partial η^2
Validation: 45	7.95	0.005	0.150	0.040	6.20	0.013	0.210	0.008
Main: 60	**12.64**	**0.0005**	**0.043**	**0.063**	**12.58**	**0.0004**	**0.037**	**0.016**
Validation: 75	11.93	0.001	0.090	0.059	**12.38**	**0.0005**	**0.021**	**0.016**
Validation: 90	9.42	0.002	0.090	0.047	6.69	0.010	0.126	0.008

Note. MFG.L: the left middle frontal gyrus; FDR: false discovery rate.

Table S12 NBS results for diagnosis effects on edge variability in OLF.R-PCL.R and sex-by-diagnosis interaction effects on edge variability in AMYG.L-ACG.R across various window lengths.

Window Length (s)	***F*(1,189)**	p_{FDR}	**partial η^2**
OLF.R-PCL.R: diagnosis			
Validation: 45	17.61	< 0.001	0.085
Main: 60	14.61	< 0.001	0.072
Validation: 75	13.17	> 0.05	0.065
Validation: 90	12.04	> 0.05	0.060
AMYG.L-ACG.R: diagnosis × sex			
Validation: 45	19.29	< 0.001	0.093
Main: 60	24.08	< 0.001	0.113
Validation: 75	22.94	< 0.001	0.108
Validation: 90	22.17	< 0.001	0.105

Note. NBS: network-based statistics; FDR: false discovery rate; OLF: olfactory cortex; PCL: paracentral lobule; AMYG: amygdala; ACG: anterior cingulate and paracingulate gyri; L: in the left hemisphere; R: in the right hemisphere.

Table S13 Validation of edge variability in OLF.R-PCL.R revealing the sex similarities and edge variability in AMYG.L-ACG.R supporting the GI model across various window lengths.

	OLF.R-PCL.R			**AMYG.L-ACG.R**		
Window Length (s)	diagnosis	sex	diagnosis × sex	diagnosis	sex	diagnosis × sex
Validation: 45	√	×	×	×	×	√
Main: 60	√	×	×	×	×	√
Validation: 75	×	×	×	×	×	√
Validation: 90	×	×	×	×	×	√

Note. GI: gender incoherence; OLF: olfactory cortex; PCL: paracentral lobule; AMYG: amygdala; ACG: anterior cingulate and paracingulate gyri; L: in the left hemisphere; R: in the right hemisphere.

4.2 Alternative dFC construction method

The present findings were derived from dFC constructed using the sliding-window approach. To verify the robustness of our core results, we further validated them using the flexible least squares (FLS) method. FLS estimates dFC within a time-varying parameter regression framework (Kalaba & Tesfatsion, 1989), emphasizing the temporal continuity of connectivity changes without relying on window length or step size. For each pair of brain regions, the BOLD signal of one region was modelled as a linear function of the other, with the regression coefficient allowed to vary smoothly over time. The algorithm jointly optimizes model fit and temporal smoothness by minimizing the sum of squared residuals and the penalty on parameter changes across consecutive time points. This approach provides a more flexible and objective characterization of dynamic interactions.

The FLS analysis was implemented in the DynamicBC (v2.2) toolbox (https://github.com/guorongwu/DynamicBC/; Liao et al., 2014) using the AAL-90 atlas, with the smoothing parameter (μ) fixed at 100 to balance model fit and temporal smoothness. Considering the potential instability of cluster centroid identification in clustering analyses, only the nodal- and edge-level findings were subjected to validation. All statistical analyses were performed following the same procedures as those described in the main text.

Using the FLS approach to construct dFC, we also observed significant diagnosis effects in modular variability of the MFG.L and in edge variability of the OLF.R-PCL.R connection, as well as a significant sex-by-diagnosis interaction effect in edge

variability of the AMYG.L-ACG.R connection (Table S14). However, none of these effects survived FDR correction, and their effect sizes were smaller than those obtained with the sliding-window method. Overall, these findings suggest that the results demonstrate a certain degree of stability across different dFC construction methods.

Table S14 Validation of core findings using dynamic functional connectivity constructed by the flexible least squares method at the nodal and edge levels.

Dynamic Functional Connectivity	Sliding-Window (WL = 60s)				Flexible Least Squares			
	$F(1,189)$	p	p_{FDR}	partial η^2	$F(1,189)$	p	p_{FDR}	partial η^2
Modular Variability								
MFG.L: diagnosis	12.64	< 0.001	0.043	0.063	10.02	0.002	0.109	0.050
Edge Variability								
OLF.R-PCL.R: diagnosis	14.61	< 0.001	< 0.001	0.072	11.28	0.001	> 0.05	0.056
AMYG.L-ACG.R: diagnosis × sex	24.08	< 0.001	< 0.001	0.113	12.64	< 0.001	> 0.05	0.063

Note. WL: window length; FDR: false discovery rate; MFG: middle frontal gyrus; OLF: olfactory cortex; PCL: paracentral lobule; AMYG: amygdala; ACG: anterior cingulate and paracingulate gyri; L: in the left hemisphere; R: in the right hemisphere.

4.3 Alternative atlases

In the primary analyses, dFC was constructed using the widely adopted AAL-90 atlas. Despite its extensive use, this atlas has limited spatial resolution and anatomical precision. Moreover, it does not include cerebellar regions, which have been shown to exhibit atypical functional coupling with the cerebral cortex in ASD (e.g., Kelly et al., 2020; Khan et al., 2015). To address these limitations and test the robustness of the main findings, we reconstructed dFC using two alternative atlases: the finer-grained, functionally informed Brainnetome-246 (BN246) atlas, and the Brainnetome-274 (BN274) atlas that includes cerebellar regions (Fan et al., 2016). Given the inherent variability of clustering in identifying cluster centroids, validation was restricted to nodal and edge levels. All other methods and analytical procedures were identical to those described in the main text.

The results indicated that, at the nodal level, no Brainnetome regions showed significant diagnosis, sex, or sex-by-diagnosis effects on modular variability after multiple-comparison correction. Nevertheless, the diagnosis effect on the modular variability of the MFG.L was consistently observed across both atlases, primarily localized to its ventrolateral area (i.e., MFG.L part 5 [area 8]; Table S15, consistent with prior findings on localization of quantitative diagnosis effects across MFG subregions (Belmonte et al., 2010)), although the effect size was reduced relative to that in the AAL-90 atlas.

At the edge level, several connections exhibited significant edge variability surviving FDR correction (Table S16-S17), including two involving cerebellar

regions. (Table S17). Specifically, in the dFC analysis based on the BN274 atlas, males showed greater edge variability between the left superior frontal gyrus (medial area 10) and the right cerebellar lobule V than females ($p_{post\text{-}hoc} < 0.001$). Moreover, the direction of sex differences in edge variability between the right superior frontal gyrus (lateral area 9) and the right cerebellar lobule X was reversed in individuals with ASD (autistic male > autistic female; $p_{post\text{-}hoc} = 0.002$) compared with their TD peers (TD male < TD female; $p_{post\text{-}hoc} = 0.016$). However, the diagnosis effect on edge variability between OLF.R and PCL.R, as well as the sex-by-diagnosis interaction between AMYG.L and ACG.R, was not replicated across the two atlases. Notably, although the first 246 regions of the BN246 and BN274 atlases are identical, the reproducibility of edge-level results between them was extremely low, suggesting that higher-resolution parcellations may be more susceptible to noise in edge variability analyses.

Taken together, only a subset of findings exhibited cross-parcellation consistency, as further discussed in the main text.

Table S15 Validation of core findings using different atlas at the nodal and edge levels.

Dynamic Functional Connectivity	AAL-90				BN-246				BN-274			
	$F(1,189)$	p	p_{FDR}	partial η^2	$F(1,189)$	p	p_{FDR}	partial η^2	$F(1,183)$[a]	p	p_{FDR}	partial η^2
Modular Variability												
MFG.L[b]: diagnosis	12.64	< 0.001	0.043	0.063	8.32	0.004	0.180	0.042	10.83	0.001	0.141	0.056
Edge Variability												
OLF.R-PCL.R: diagnosis	14.61	< 0.001	< 0.001	0.072	—				—			
AMYG.L-ACG.R: diagnosis × sex	24.08	< 0.001	< 0.001	0.113	—				—			

Note. AAL: anatomical automatic labeling; BN: brainnetome; FDR: false discovery rate; MFG: middle frontal gyrus; OLF: olfactory cortex; PCL: paracentral lobule; AMYG: amygdala; ACG: anterior cingulate and paracingulate gyri; L: in the left hemisphere; R: in the right hemisphere.

a. Data from six participants (one autistic female, four autistic males, and one typically developing male) were excluded due to incomplete cerebellar coverage in the resting-state fMRI scans.

b. In both BN-246 and BN-274 atlases, the significant diagnosis effect on MFG.L modular variability was located in the ventrolateral area of the MFG.L.

Table S16 The significant diagnosis, sex, and sex-by-diagnosis interaction effects on edge variability after FDR correction in the Brainnetome-246 atlas (N = 196).

Brainnetome-246	$F(1,189)$	p	p_{FDR}	partial η^2
Diagnosis				
IFG_L_6_5_44op-STG_R_6_2_41/42	14.48	0.0002	< 0.05	0.071
STG_L_6_5_38l-FuG_L_3_3_37vl	12.66	0.0005	< 0.05	0.063
PrG_R_6_3_4ul-PoG_L_4_1_1/2/3ulhf	16.36	0.0001	< 0.05	0.080
IFG_L_6_4_45r-Str_R_6_5_dCa	16.68	0.0001	< 0.05	0.081
Sex				
MFG_L_7_2_IFJ-STG_L_6_6_22r	16.32	0.0001	< 0.05	0.079
IFG_R_6_6_44v-PCun_R_4_4_31	16.65	0.0001	< 0.05	0.081
SPL_R_5_4_7pc-Tha_L_8_2_mPMtha	14.04	0.0002	< 0.05	0.069
Diagnosis × Sex				
ITG_L_7_1_20iv-pSTS_R_2_2_cpSTS	13.20	0.0004	< 0.05	0.065
PhG_R_6_5_TI-PoG_L_4_2_1/2/3tl	17.65	0.00004	< 0.05	0.085

Note. FDR: false discovery rate; IFG: inferior frontal gyrus; STG: superior temporal gyrus; FuG: fusiform gyrus; PrG: precentral gyrus; PoG: postcentral gyrus; Str: striatum; MFG: middle frontal gyrus; PCun: precuneus; SPL: superior parietal lobule; Tha: thalamus; ITG: inferior temporal gyrus; pSTS: posterior superior temporal sulcus; PhG: parahippocampal gyrus; L: in the left hemisphere; R: in the right hemisphere; op: opercular; l: lateral; vl: ventrolateral; ul: upper limb; r: rostral; dCa: dorsal caudate; IFJ: inferior frontal junction; v: ventral; pc: postcentral; mPMtha: medial premotor thalamus; iv: intermediate ventral; cpSTS: caudoposterior superior temporal sulcus; tl: tongue and larynx.

Table S17 The significant diagnosis, sex, and sex-by-diagnosis interaction effects on edge variability after FDR correction in the Brainnetome-274 atlas (N = 190).

Brainnetome-274	$F(1,183)$	p	p_{FDR}	partial η^2
Diagnosis				
IFG_R_6_5_44op-ITG_L_7_5_37vl	16.83	0.0001	< 0.05	0.084
PrG_R_6_3_4ul-PoG_L_4_1_1/2/3ulhf	19.39	0.00002	< 0.05	0.096
Sex				
MFG_L_7_3_46-IFG_L_6_4_45r	18.15	0.00003	< 0.05	0.090
MFG_R_7_6_6vl-MTG_L_4_4_aSTS	13.02	0.0004	< 0.05	0.066
MTG_L_4_2_21r-ITG_L_7_7_20cv	16.33	0.0001	< 0.05	0.082
OrG_R_6_6_12/47l-IPL_R_6_3_40rd	13.93	0.0003	< 0.05	0.071
SFG_L_7_7_10m-Cb_Right_V	15.22	0.0001	< 0.05	0.077
Diagnosis × Sex				
pSTS_R_2_2_cpSTS-PCun_R_4_2_5m	16.62	0.0001	< 0.05	0.083
SFG_R_7_3_9l-Cb_Left_X	15.16	0.0001	< 0.05	0.076

Note. FDR: false discovery rate; IFG: inferior frontal gyrus; ITG: inferior temporal gyrus; PrG: precentral gyrus; PoG: postcentral gyrus; MFG: middle frontal gyrus; MTG: middle temporal gyrus; OrG: orbital gyrus; IPL: inferior parietal lobule; SFG: superior frontal gyrus; Cb: cerebellar; pSTS: posterior superior temporal sulcus; PCun: precuneus; L: in the left hemisphere; R: in the right hemisphere; op: opercular; vl: ventrolateral; ul: upper limb; r: rostral; aSTS: anterior superior temporal sulcus; cv: caudoventral; l: lateral; rd: rostrodorsal; m: medial; V: cerebellar lobule V; cpSTS: caudoposterior superior temporal sulcus; X: cerebellar lobule X.

Reference

Belmonte, M. K., Gomot, M., & Baron-Cohen, S. (2010). Visual attention in autism families: 'unaffected' sibs share atypical frontal activation. *Journal of Child Psychology and Psychiatry, 51*(3), 259–276. https://doi.org/10.1111/j.1469-7610.2009.02153.x

Fan, L., Li, H., Zhuo, J., Zhang, Y., Wang, J., Chen, L., ... Jiang, T. (2016). The human Brainnetome atlas: A new brain atlas based on connectional architecture. *Cerebral Cortex*, *26*(8), 3508–3526. https://doi.org/10.1093/cercor/bhw157

Kalaba, R., & Tesfatsion, L. (1989). Time-varying linear regression via flexible least squares. *Computers & Mathematics with Applications*, *17*(8-9), 1215–1245. https://doi.org/10.1016/0898-1221(89)90091-6

Kelly, E., Meng, F., Fujita, H., Morgado, F., Kazemi, Y., Rice, L. C., ... Tsai, P. T. (2020). Regulation of autism-relevant behaviors by cerebellar-prefrontal cortical circuits. *Nature Neuroscience*, *23*(9), 1102–1110. https://doi.org/10.1038/s41593-020-0665-z

Khan, A. J., Nair, A., Keown, C. L., Datko, M. C., Lincoln, A. J., & Müller, R.-A. (2015). Cerebro-cerebellar resting-state functional connectivity in children and adolescents with autism spectrum disorder. *Biological Psychiatry*, *78*(9), 625–634. http://doi.org/10.1016/j.biopsych.2015.03.024

Liao, W., Wu, G.-R., Xu, Q., Ji, G.-J., Zhang, Z., Zang, Y.-F., & Lu, G. (2014). DynamicBC: A MATLAB toolbox for dynamic brain connectome analysis. *Brain Connectivity*, *4*(10), 780–790. https://doi.org/10.1089/brain.2014.0253