Archival Report

Biological Psychiatry

Neuroanatomical Diversity of Corpus Callosum and Brain Volume in Autism: Meta-analysis, Analysis of the Autism Brain Imaging Data Exchange Project, and Simulation

Aline Lefebvre, Anita Beggiato, Thomas Bourgeron, and Roberto Toro

ABSTRACT

BACKGROUND: Patients with autism have been often reported to have a smaller corpus callosum (CC) than control subjects.

METHODS: We conducted a meta-analysis of the literature, analyzed the CC in 694 subjects of the Autism Brain Imaging Data Exchange project, and performed computer simulations to study the effect of different analysis strategies.

RESULTS: Our meta-analysis suggested a group difference in CC size; however, the studies were heavily underpowered (20% power to detect Cohen's $d = .3$). In contrast, we did not observe significant differences in the Autism Brain Imaging Data Exchange cohort, despite having achieved 99% power. However, we observed that CC scaled nonlinearly with brain volume (BV): large brains had a proportionally smaller CC. Our simulations showed that because of this nonlinearity, CC normalization could not control for eventual BV differences, but using BV as a covariate in a linear model would. We also observed a weaker correlation of IQ and BV in cases compared with control subjects. Our simulations showed that matching populations by IQ could then induce artifactual BV differences.

CONCLUSIONS: The lack of statistical power in the previous literature prevents us from establishing the reality of the claims of a smaller CC in autism, and our own analyses did not find any. However, the nonlinear relationship between CC and BV and the different correlation between BV and IQ in cases and control subjects may induce artifactual differences. Overall, our results highlight the necessity for open data sharing to provide a more solid ground for the discovery of neuroimaging biomarkers within the context of the wide human neuroanatomical diversity.

Keywords: Autism, Brain volume, Computational neuroanatomy, Corpus callosum, Meta-analysis, Statistical power

http://dx.doi.org/10.1016/j.biopsych.2015.02.010

Autism spectrum disorders (ASD) are pervasive developmental disorders with qualitative impairments in social interaction and communication, along with restricted, repetitive, and stereotyped patterns of behavior. Several cognitive studies have suggested that difficulty integrating multiple sources of stimulation may be a common characteristic of ASD, which has led, for example, to the influential weak central coherence hypothesis (1). The neural basis of these difficulties has been hypothesized to be an imbalance between local and distant connections: local overconnectivity and long-distance underconnectivity (2,3). The connectivity hypothesis has been a major subject of study and discussion in ASD research (4,5).

The corpus callosum (CC)—the largest commissure connecting the left and right hemispheres—appeared then as a natural candidate to look for evidence of connectivity abnormalities. The CC exists exclusively within eutherian mammals (kangaroos and other marsupials lack a CC) and has been suggested to play an important role in the evolution of lateralization. The number of callosal axons is disproportionally smaller in large-brain mammals (like humans) compared with small-brain mammals (like mice). A smaller number of callosal fibers, and their increased length, could hinder the formation of interhemispheric synchronous neuronal populations, thus facilitating local recruitment and leading to functional lateralization (6–8). The CC has been, as a consequence, one of the most studied white matter tracts in ASD. Numerous reports have indeed described significantly smaller CC among patients compared with control subjects, and a series of studies have suggested a higher incidence of ASD within cases of CC agenesis or callosotomy patients (9–11).

However, many of these analyses have relied on small cohorts (~30 patients, ~30 control subjects), without statistical power to find even effect sizes as large as .5 standard deviations between groups. Despite the lack of power, studies often report statistically significant differences. A solution to the methodological problems associated with small cohorts has been recently proposed by the Autism Brain Imaging Data Exchange project (Abide) (12). Large cohorts are difficult to gather and analyze by any single research group. Abide provides open access to

SEE COMMENTARY ARTICLE ON PAGE 81

behavioral and neuroimaging data for almost 600 patients and 600 age-, sex-, and IQ- matched control subjects from an international consortium of 17 research groups. Abide provides the research community with the statistical power necessary to detect even small differences in case/control designs, and to use more sophisticated analysis strategies, providing a wider perspective on neuroanatomical diversity.

Here, we first present a review of studies of CC size differences in ASD. We observed a general lack of statistical power (only 20% power to detect two-sided differences of .3 standard deviations at .05 level of significance), which contrasted with the frequent report of significant findings (9 out of 17 studies). Next, we present our analysis of the diversity of the CC in Abide and differences related to scanning sites, age, sex, brain size polymorphism, and diagnostic group. Even though previous studies have reported diagnostic group differences as large as .3 to .7 standard deviations and despite having analyzed a number of subjects comparable with the sum of all previously studied, we did not find significant differences between patients with ASD and control subjects. Finally, we discuss possible ways in which analysis strategies, such as the normalization of CC size by total brain volume (BV) or the matching of subjects by IQ scores, could lead to artifactual differences in BV and CC size between patients and control subjects.

METHODS AND MATERIALS

Meta-analysis

We included all studies from the recent review by Frazier and Hardan (13) and searched PubMed (http://www.ncbi.nlm.nih.gov/pubmed/) for additional studies reporting differences in CC between patients with ASD and control subjects: (autism OR PDD OR "pervasive developmental disorder") AND "corpus callosum". Table 1 describes the cohort in the articles meta-analyzed (see Figure S1 and Supplemental Materials in Supplement 1 for further details on the inclusion procedure).

Analysis of Abide

We used FreeSurfer v5.1 (http://surfer.nmr.mgh.harvard.edu/) to process the 1102 subjects in Abide (http://fcon_1000.projects.nitrc.org/indi/abide) with T1-weighted magnetic resonance imaging data available. We developed an open online tool to visually control the accuracy of the segmentations (http://siphonophore.org/corpuscallosum). Based on this quality control, we excluded 380 subjects, 331 of them because of a mislabeled fornix that was included in the middle segment of the CC (Figure 1; Figure S2 in Supplement 1). Not excluding the subjects with a mislabeled fornix, however, did not change our results. We only included subjects from 7.5 to 40 years old. The final sample consisted of 694 subjects: 328 patients (290 male subjects, 38 female subjects) and 366 control subjects (304 male subjects, 62 female subjects). Full IQ (FIQ) was available for 672 subjects and verbal IQ (VIQ) and performance IQ (PIQ) for 538 subjects. Table 2 describes the groups of subjects retained from each scanning site. Statistical analyses were performed using JMP Pro 10.0.2 (http://www.jmp.com), R (http://www.r-project.org/), iPython (http://ipython.org/), and G*Power (http://gpower.hhu.de/) (see Supplemental Materials in Supplement 1 for further details on the analysis).

Simulations

Simulations were written in Python to analyze the effect on case/control comparisons produced by 1) normalization by

Table 1. Population in the Studies Included in the Meta-analysis

Reference	Age_{ASD} ± SD	Age_{Ctrl} ± SD	IQ Level (LF/HF)	N_{ASD} (F)	N_{Ctrl} (F)	Matching Strategy		
						Relative	GLM	IQ
Gaffney *et al.* 1987 (36)	11 ± 5	12 ± 5	LF/HF	13 (3)	35 (14)	No	No	No
Egaas *et al.* 1995 (37)	16 ± 10	16 ± 10	LF/HF	51 (6)	51 (6)	No	No	No
Piven *et al.* 1997 (38)	18 ± 5	20 ± 4	LF/HF (PIQ > 70)	35 (6)	36 (16)	No	Yes	No
Manes *et al.* 1999 (39)	14 ± 7	12 ± 5	LF	27 (5)	17 (6)	Yes	No	Yes
Elia *et al.* 2000 (40)	11 ± 4	11 ± 3	LF	22 (0)	11 (0)	No	No	No
Rice *et al.* 2005 (41)	12 ± 4	13 ± 4	HF/LF	12 (0)	8 (0)	No	Yes	No
Vidal *et al.* 2006 (42)	10 ± 3	11 ± 3	HF	24 (0)	26 (0)	Yes	Yes	Yes
Boger-Mediggo *et al.* 2006 (43)	4 ± 3	4 ± .5	–	45 (7)	26 (8)	Yes	Yes	No
Alexander 2007 (44)	16 ± 7	16 ± 6	HF (PIQ)	43 (–)	34 (–)	No	Yes	Yes
Just *et al.* 2007 (3)	27 ± 12	25 ± 10	HF	18 (1)	18 (3)	Yes	No	Yes
Hardan *et al.* 2009 (45)	11 ± 1	11 ± 1	HF	22 (0)	23 (0)	Yes	No	Yes
Freitag *et al.* 2009 (46)	18 ± 4	19 ± 1	HF	15 (2)	15 (2)	No	Yes	Yes
Keary 2009 (47)	20 ± 10	19 ± 9	HF	32 (2)	34 (2)	No	Yes	Yes
Anderson *et al.* 2011 (48)	22 ± 7	21 ± 7	HF/LF	53 (0)	39 (0)	No	No	Yes
Hong *et al.* 2011 (49)	9 ± 2	10 ± 2	HF	18 (0)	16 (0)	No	Yes	Yes
Frazier *et al.* 2012 (50)	11 (8–12)[a]	11 (7–13)[a]	HF	23 (0)	23 (0)	No	Yes	No
Prigge *et al.* 2013 (51)	14 ± 8	15 ± 7	(HF/HF PIQ > 70)	68 (0)	47 (0)	No	Yes	Yes

Matching strategy: Relative: differences in BV were accounted by dividing CC by BV; GLM: differences in BV were accounted by including BV as a covariate in a general linear model; IQ: differences in IQ were accounted for by matching groups by intelligence scores.

ASD, autism spectrum disorders; BV, total brain volume; CC, corpus callosum; Ctrl, control subjects; F, females; GLM, general linear model; HF, high functioning; LF, low functioning; PIQ, performance IQ.

[a]Age range.

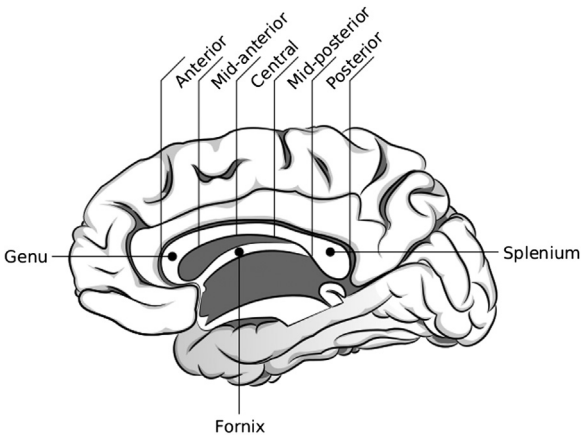

Figure 1. Segmentation of the corpus callosum. The corpus callosum was segmented into five regions from genu to splenium: anterior, mid-anterior, central, mid-posterior and posterior.

total BV, 2) covariation of total BV, and 3) IQ matching. The code for these simulations has been made available as iPython notebooks: http://siphonophore.org/corpuscallosum.

RESULTS

Meta-analysis

On October 8, 2013, our PubMed query returned 183 articles, among which 17 fulfilled our inclusion criteria. Combined, these articles provided data on a total of 980 subjects, 521 patients with ASD, and 459 control subjects: 77 female subjects, 903 male subjects; mean age 14.8 years old (data are summarized in Table 3).

The fixed and random effects meta-analyses provided an estimated global effect size of .5 standard deviations, without evidence of heterogeneity ($p = .87$, $\tau^2 < .0001$, $H = 1$, $I^2 = 0\%$) (see forest plot in Figure 2). A standard test for publication bias estimates the presence of a correlation between effect

Table 2. Population of the Abide Project Retained for Analysis per Scanning Site

Site	Institution	N_{ASD} Included/Total	N_{Ctrl} Included/Total	Age_{ASD} (Range)	Age_{Ctrl} (Range)	AQ (ASD/Ctrl)	SRS (ASD/Ctrl)	ADOS (ASD/Ctrl)
Caltech	California Institute of Technology, USA	7/19	8/19	26.3 (20–39)	25.2 (20–39)			19 (19/–)
CMU	Carnegie Mellon University, USA	6/14	5/13	27.5 (22–31)	31.2 (25–40)			14 (14/–)
KKI	Kennedy Krieger Institute, USA	20/22	32/33	9.9 (8–13)	10.2 (8–13)			22 (22/–)
Leuven	University of Leuven, Belgium	27/29	28/35	17.4 (12–29)	17.7 (12–28)	29 (14/15)	64 (29/35)	
MaxMun	Ludwig Maximilian University Munich, Germany	11/24	22/33	21.2 (8–35)	25.4 (10–35)			6 (6/–)
NYU	New York University Langone Medical Center, USA	54/79	75/105	14.7 (8–39)	16.3 (8–32)		149 (77/72)	79 (79/–)
OHSU	Oregon Health and Science University, USA	11/13	11/15	11.1 (8–14)	9.9 (8–12)			13 (13/–)
Olin	Olin, Institute of Living at Hartford Hospital, USA	16/20	14/16	17.4 (12–24)	16.9 (10–23)			20 (20/–)
Pitt	University of Pittsburg, School of Medicine, USA	30/30	27/27	18.9 (9–35)	18.9 (9–33)			26 (26/–)
SBL	Social Brain Lab, BCN NIC UMC Groningen and Netherlands Institute for Neurosciences, Netherlands	7/15	8/15	30.7 (27–35)	32.9 (26–39)	27 (14/13)		8 (8/–)
SDSU	San Diego State University, USA	8/14	12/22	15.2 (12–17)	14.3 (12–17)			13 (13/–)
Stanford	Stanford University, USA	8/20	5/20	9.6 (8–12)	11.2 (8–12)			19 (19/–)
Trinity	Trinity Centre for Health Sciences, Ireland	9/24	11/25	18.2 (14–23)	17.6 (12–26)			24 (24/–)
UCLA	University of California, Los Angeles, USA	32/62	35/47	13.3 (8–18)	12.9 (10–18)			61 (61/–)
UM	University of Michigan, USA	23/68	27/77	12.8 (9–19)	14.8 (9–29)			
USM	University of Utah, School of Medicine, USA	42/58	26/43	22.0 (11–38)	22.9 (9–39)		100 (58/42)	90 (58/32)
Yale	Yale Child Study Center, USA	17/28	20/28	13.1 (9–18)	12.9 (8–18)		52 (26/26)	1 (1/–)
	Total	328	366	16.6 (8–39)	17.0 (8–40)	56	365	415

ADOS, Autism Diagnostic Observation Schedule; AQ, Autism Quotient; ASD, autism spectrum disorders; Ctrl, control subjects; SRS, Social Responsiveness Scale.

Table 3. Meta-analysis: Mean CC Size, Effect Size, Significance of the Difference, Statistical Power.

Reference	N_{ASD}	N_{Ctrl}	Mean CC_{ASD} ± SD (cm^2)	Mean CC_{Ctrl} ± SD (cm^2)	Effect Size	*p* Value (Two-Sided)	Power to Detect SD = .3 (Two-Sided)
Gaffney *et al.* 1987 (36)	13	35	5.89 ± 1.04	6.24 ± 1.37	−.27	.41	17.3%
Egaas *et al.* 1995 (37)	51	51	5.57 ± .99	5.89 ± .91	−.33	.097	32.3%
Piven *et al.* 1997 (38)	35	36	6.15 ± .83	6.40 ± .38	−.39	.11	24.1%
Manes *et al.* 1999 (39)	27	17	4.64 ± .99	5.71 ± .97	−1.07	.0011	16.2%
Elia *et al.* 2000 (40)	22	11	5.26 ± 1.00	5.41 ± .64	−.16	.67	12.7%
Rice *et al.* 2005 (41)	12	8	7.34 ± 1.11	7.75 ± 1.14	−.35	.45	9.3%
Vidal *et al.* 2006 (42)	24	26	6.06 ± 1.15	6.68 ± .79	−.62	**.033**	17.9%
Boger-Megiddo *et al.* 2006 (43)	45	26	4.59 ± .67	4.99 ± .72	−.57	**.022**	24.1%
Alexander *et al.* 2007 (44)	43	34	7.87 ± 1.99	9.32 ± .70	−.77	**.012**	25.2%
Just *et al.* 2007 (3)	18	18	6.40 ± .88	7.1 ± .88	−.78	**.025**	13.9%
Hardan *et al.* 2009 (45)	22	23	5.74 ± 1.13	6.58 ± 1.04	−.76	**.014**	16.2%
Freitag *et al.* 2009 (46)	15	15	6.22 ± .45	6.54 ± 1.24	−.34	.36	12.2%
Keary *et al.* 2009 (47)	32	34	6.19 ± 1.09	6.76 ± 1.10	−.51	**.040**	22.4%
Anderson *et al.* 2011 (48)	53	39	6.54 ± 1.20	7.05 ± .90	−.46	**.031**	29.6%
Hong *et al.* 2011 (49)	18	16	8.14 ± 1.31	8.27 ± 1.27	−.10	.78	13.3%
Frazier *et al.* 2012 (50)	23	2	6.30 ± 1.11	6.78 ± 1.08	−.43	.15	16.7%
Prigge *et al.* 2013 (51)	68	47	5.74 ± .91	6.24 ± .89	−.55	.0044	36.0%

The different values were scaled to provided measurements in cm^2 (this scaling does not affect our meta-analysis, which was performed on standardized mean differences).

ASD, autism spectrum disorders; CC, corpus callosum; Ctrl, control subjects.

size and standard error for each study (14). A correlation could be detected if small studies were more likely to be published when the effect size is large and then more likely to produce a statistically significant finding. Large studies, which approximate better the true effect size, should then provide a reference. The funnel plot (Figure 3) was symmetric, and we did not find a significant correlation between the effect size and the standard error of each study ($p = .87$, $t = .169$). However, it is difficult to establish whether there was a publication bias or not using this method because there were no really large studies in our sample: the largest statistical power to detect a .3 standard deviation was 36% (statistical power for each article is reported in Table 3). In general, the mean achieved power (power to detect the reported effect size) was 46.9%. A priori, the studies were very underpowered to detect a difference of .3 standard deviations (what is commonly considered to be a small difference): 20% power. Despite the general lack of power, however, 9 out of 17 studies reported significant results (Table 3), suggesting a selective publication of statistically significant findings.

Analysis of Abide

Behavioral Characteristics. Scores in the Autism Diagnostic Observation Schedule (ADOS) were available for 415 subjects: 95.3% of the patients had scores ≥7, a threshold suggestive of ASD; 70.5% had scores ≥10, suggestive of the most classic autism phenotype. Several sites provided individually calibrated severity ADOS scores (15): 81.7% of the patients had scores ≥6, suggestive of ASD. There were

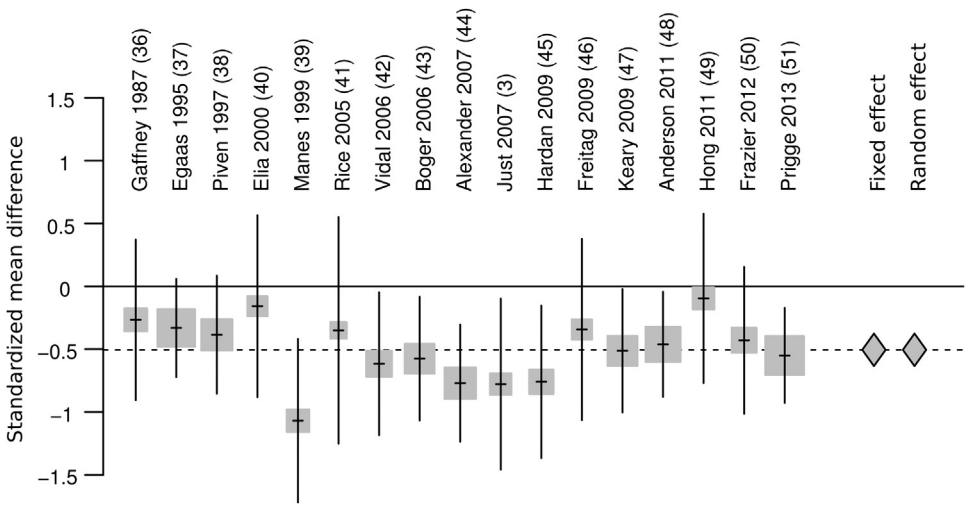

Figure 2. Standardized mean difference in corpus callosum size between patients with autism spectrum disorders and control subjects in the articles analyzed. The standardized effect sizes (mean differences) were not found to be heterogeneous, and the fixed and random effects meta-analyses provided the same estimates of the mean effect size, −.51 (95% confidence interval = −.63 to −.38, $Z = -7.642$, $p < .0001$). For each study, the size of the square corresponds with the sample size. The standardized mean difference is shown by diamonds (the estimation was the same for the fixed and random effects models).

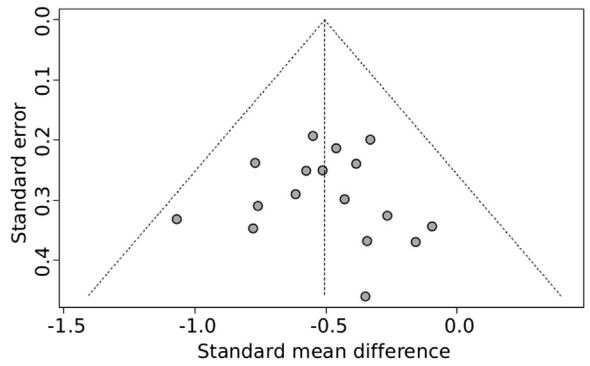

Figure 3. Standard error versus standard mean difference. The funnel plot does not show a bias among articles with small sample sizes (large standard errors) to overestimate the standard mean difference. However, the absence of studies with sufficiently large cohorts does not allow us to conclude an absence of publication bias. The dashed lines indicate the 95% confidence interval.

strongly statistically significant differences between patients and control subjects in the Social Responsiveness Scale ($p < .0001$) and Autism Quotient scores ($p < .0001$). Abide includes only high-functioning subjects. FIQ was statistically significantly lower among patients (6 points, $p < .0001$), mostly due to a statistically significantly lower VIQ (8 points, $p < .001$). There was no statistically significant difference in PIQ ($p = .19$). VIQ and PIQ were correlated at $\rho = .44$ (for further details on autistic traits and intelligence scores, see Supplemental Materials, Table S1, and Figure S3 in Supplement 1).

Neuroanatomical Diversity: Differences Related to Age, Sex, Scanning Site, and Diagnostic Group. There were significant effects of age, sex, and scanning site on CC, BV, and intracranial volume (ICV), but no statistically significant differences were related to diagnostic group (Table 4; Table S2 in Supplement 1). Subcallosal regions were not statistically significantly different either (Tables S3 and S4 in Supplement 1). Including subjects with mislabeled fornix did not change our results (Table S9 in Supplement 1). Total CC volume presented a small correlation with ICV ($\rho = .27$) and a medium correlation with BV ($\rho = .43$). The same was observed for individual callosal regions (Table S5 in Supplement 1). Indeed, the variability of callosal region size was strongly related to total CC size: the first principal component of the variance-covariance matrix of callosal region size captured 65% of the variance and correlated with $\rho = .98$ with total CC. A partial correlation analysis suggested that the variability of the CC was more directly related to BV than to ICV and that the observed correlation between ICV and CC was largely mediated by BV (Table S6 in Supplement 1). In consequence, our further analyses related CC variability with BV.

The size of many brain structures changes nonlinearly relative to total BV; for example, large brains tend to have disproportionately more cortical surface area than small brains (16,17). We estimated the way in which the proportion of region Y changes relative to region X by the slope β of the regression of log(Y) on log(X), known as the scaling coefficient. If β is $\sim$1, the volume of Y is proportional to the volume of X (called isometry); if $\beta < 1$, the proportion of Y decreases as X increases (negative allometry); and if $\beta > 1$, the proportion of Y is progressively larger as X increases (positive allometry). There was a statistically significant negative allometry of CC with BV, with scaling coefficient $\beta = .64$ (95% confidence interval = .54–.75), i.e., the proportion of CC in large brains should be smaller than in small brains (Figure 4; Figure S4 in Supplement 1). Including subjects with mislabeled fornix did not change our results ($\beta = .71$, 95% confidence interval = .62–.79). We observed the same negative scaling for each callosal subregion (average $\beta = .64$, range from .58 to .73; Table S7 in Supplement 1). The scaling coefficient for CC was not different between patients with ASD and control subjects —neither the diagnostic group effect nor the interaction between diagnostic group and BV were statistically significant (diagnostic group: $p = .43$, $F = .62$; diagnostic group $\times$ BV: $p = .85$, $F = .04$). Additionally, none of the callosal regions showed significant differences between patients and control subjects. Because of the negative allometry between CC and BV and due to the significant effect of sex on BV, we expected female subjects to have smaller CC than male subjects in absolute terms but larger than male subjects in relative terms (i.e., CC/BV). This was indeed the case: relative CC was statistically significantly larger in female than male subjects (age and site as covariates, sex effect: $p = .0095$, $F = 6.75$). The difference was, however, completely explained by the relationship between CC and BV: adding BV as a covariate made the sex effect not statistically significant ($p = .18$, $F = 1.8$), as it had been previously pointed out by Luders *et al.* (18).

Table 4. Site, Age, Sex, and Group Effects in ICV, BV, and CC

	ICV	BV	CC
Mean Size (cm^3) ± SD	1368 ± 231	1131 ± 130	3.16 ± .54
Site Effect			
F	32.8	10.6	10.7
p value	<.0001	<.0001	<.0001
R^2	43.7%	20.1%	20.2%
Age Effect			
Increase (cm^3/year)	4.3	2.3	.019
F	12.19	10.86	45
p value	.0005	.001	<.0001
R^2	1.7%	1.5%	6.1%
Sex Effect			
Percent difference (1-female/male)	9.4%	9.3%	7.4%
F	28.2	62.11	17.1
p value	<.0001	<.0001	<.0001
R^2	3.9%	8.2%	2.4%
Group Effect			
Difference (cm^3)	6.8	4	−.007
F	.26	.03	.35
p value	.61	.86	.56
R^2	.00	.00	.00
Variance Explained by the Full Model	46.9%	26.0%	22.0%

BV, total brain volume; CC, corpus callosum; ICV, intracranial volume.

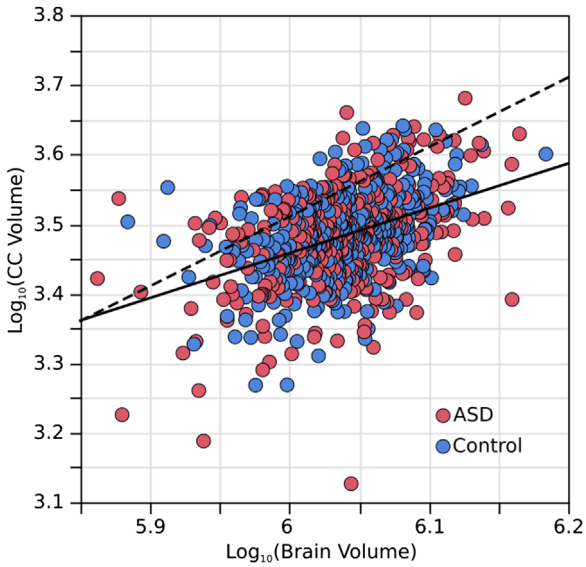

Figure 4. Relationship between corpus callosum (CC) size and total brain volume (BV). CC volume versus BV volume in logarithmic scale. The CC volume is proportionally smaller in large brains compared with small brains. The dashed line has a slope of 1, which would indicate a linear scaling of CC and BV. The solid line, showing the observed regression line, has a slope of .64, significantly smaller (95% confidence interval = .54–.75). ASD, autism spectrum disorders.

Correlation between BV and IQ. IQ is often used to match patient and control groups (8/17 articles in our meta-analysis used IQ matching). However, IQ and ICV or IQ and BV are correlated in the general population and matching subject groups by IQ may also affect the neuroanatomical composition of the groups. IQ appears to be sensitive also to environmental factors such as stress or socioeconomic status (19). It was not clear then that the relationship between IQ and brain size would be the same in a control group and a group of patients with ASD. Indeed, whereas we did observe the expected correlation between IQ and BV in the control group ($\rho = .23$, $p < .0001$, $F = 18.75$), the relationship was significantly weaker in the group of patients with ASD ($\rho = .04$, $p = .044$, $F = 4.10$). For equivalent BV, FIQ was lower in the patient group, and increases in BV resulted in smaller increases in FIQ in patients compared with control subjects (diagnostic group effect: $p < .0001$, $F = 29.33$; diagnostic group × BV interaction: $p = .0178$, $F = 5.64$; linear model for FIQ with site, age, and sex covariates and BV, diagnostic group, and diagnostic group × BV as main effects). Whereas in control subjects FIQ increased on average by 1 point every 31 cm^3, the same increase required 84 cm^3 in patients (Figure 5). Including subjects with mislabeled fornix did not change our results (diagnostic group × BV interaction: $p = .0016$ for FIQ).

Most of the difference in the increase of FIQ with BV between patients and control subjects was due to a difference in the correlation between VIQ and BV but not between PIQ and BV. VIQ and BV correlated with $\rho = .22$ in control subjects but with $\rho = .08$ in patients. By contrast, PIQ and BV correlated with $\rho = .18$ and $\rho = .17$ in control subjects and patients, respectively (Table S8 in Supplement 1 shows correlations at individual sites). The difference in the increase of VIQ with BV was statistically significant (diagnostic group × BV: $p = .035$, $F = 4.5$), but the difference in the increase of PIQ with BV was not statistically significant (diagnostic group × BV: $p = .60$, $F = .27$). The results were similar if subjects with mislabeled fornix were included (diagnostic group × BV: $p = .015$ for VIQ, $p = .19$ for PIQ).

Simulations: Effect of BV Normalization and IQ Matching on Diagnostic Group Comparisons

Total BV Normalization. Since the first observations by Kanner (20), many researchers have reported a higher incidence of macrocephaly (21) in ASD, as well as a larger average BV (22). Recent studies, however, have shed some doubts on this hypothesis (23,24). In any case, if head circumference or BV were larger among patients compared with the general population and because of the negative allometry between CC and BV, we would expect the proportion of the CC to be smaller in the ASD group than in the control group.

A frequent strategy for controlling BV effects is to divide CC by BV (5 of 17 articles in our meta-analysis used it). Our simulations showed that BV normalization did not eliminate the effect of a difference in BV between groups when this difference was large enough (Supplemental Materials and Figure S5 in Supplement 1). For example, BV normalization in a study comparing two groups of 50 subjects each with a mean BV difference of .65 standard deviations should detect a statistically significant difference in CC size 50% of the time. A mean BV difference of .25 standard deviations should suffice

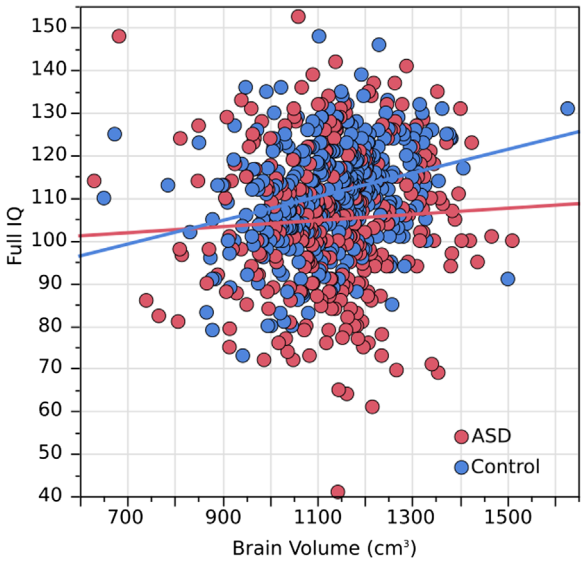

Figure 5. Relationship between full IQ (FIQ) and total brain volume (BV) in patients with autism spectrum disorders (ASD) and control subjects. The correlation between FIQ and BV was significantly weaker in ASD compared with control subjects (diagnostic group × BV interaction $p = .0178$). Whereas in control subjects FIQ increased on average by 1 point every 31 cm^3, the same 1 point increase in FIQ required an increase of 84 cm^3 in patients with ASD.

to produce an artifactual difference in CC size 20% of the time [higher than the average statistical power of many neuroimaging studies (25)].

Total BV Covariation. An alternative strategy to control for BV effects is to include BV and diagnostic group × BV as covariates in a linear model (8 of 17 articles in our meta-analysis used it). Our simulations showed that this approach controlled better the effect of BV differences between groups, inducing detectable differences in CC size only for very large groups and very large differences in BV (Supplemental Materials and Figure S6 in Supplement 1).

IQ Matching. Intelligence scores are often used to match patients and control subjects (9 of 17 studies in our meta-analysis used it). Our analysis of Abide suggested, however, that IQ and brain size were correlated differently in patients and control subjects. If larger increases in BV were required among patients than control subjects to obtain similar increases in IQ, matching groups by IQ may bias the recruitment to either decrease the number of control subjects with large BV or increase the number of patients with large BV. This bias could, in turn, affect the assessment of group differences in CC size. We simulated the impact of matching by FIQ two groups with different correlations between FIQ and BV (Supplemental Materials in Supplement 1). The results (Figure S7 in Supplement 1) showed that only in groups with a strong relationship between IQ and BV did the strategy of matching groups by FIQ induce a reproducible artefactual difference in mean BV. For example, comparisons between two groups of 50 subjects, each with a correlation between IQ and BV or $\rho_{IB} = .5$, will detect a significant difference in BV 20% of the time.

DISCUSSION

Despite recurrent reports of smaller CC in ASD compared with control subjects, our analyses of Abide did not show any statistically significant difference. The difference that we observed was $<.02$ standard deviations for CC size and $<.04$ standard deviations for BV—practically inexistent. This contrasts starkly with the global effect size found in our meta-analysis: .5 standard deviations, more than 20 times larger. Studies using small cohorts can produce statistically significant results only if the effect sizes are large. Combined with a selective report of positive findings, this makes meta-analyses based on small studies tend to overestimate the true effect size (26).

The absence of difference does not seem to be due to a difference in the characteristics of Abide. Abide was comparable with the cohorts described in the literature in regard to the subjects' age, sex ratio, and IQ level. Additionally, patients with ASD in Abide do not seem to present with a milder autistic phenotype: ADOS, Autism Quotient, and Social Responsiveness Scale scores were all highly significantly different between patients and control subjects. Several of the previous reports, however, included low-functioning subjects either exclusively or in addition to high-functioning subjects, whereas Abide includes only high-functioning subjects. But differences in CC size had been equally reported among low-functioning and high-functioning subjects. Out of the nine studies in our meta-analysis that included exclusively high-functioning subjects, six reported significant differences in CC size as large as .78 standard deviations. High-functioning subjects represent, however, just one part of the autistic spectrum. Our results do not rule out the possibility that differences in CC size could be found in a sufficiently powered analysis of a cohort including subjects with low-functioning ASD.

The absence of difference does not seem to stem either from the multi-centric, nonharmonized nature of Abide. Several scanning sites within Abide included as many or even more subjects than in the previous literature (New York University Langone Medical Center: $n = 129$, University of Utah, School of Medicine: $n = 68$, University of California, Los Angeles: $n = 67$). In none of them did we detect a statistically significant difference. Finally, the computational neuroanatomy methods that we used to segment and measure CC, BV, and ICV are all standard and well validated. A recent analysis of Abide by Haar *et al.* (27) did find a small but statistically significant difference in the middle section of the CC; however, this is a region that was often mislabeled by FreeSurfer (30% of our cohort). We are making available our data tables and scripts, as well as a web interface with our quality control decisions, to allow the community to inspect and criticize our analyses (http://siphonophore.org/corpuscallosum). The scanning site effect in Abide was strongly significant and may have masked subtle differences. To detect them, future collaborative efforts will required large cohorts and methodological harmonization.

The major, clear, difference between our analysis and the previous ones was statistical power: we achieved 99% power to detect differences of .3 standard deviations, whereas the highest statistical power to detect this effect size among the studies in our meta-analysis was only 36%. Given the low statistical power of the previous reports, even if there were a real difference in CC of this size, there should be ~80% of negative reports. However, only 8 of the 17 articles analyzed reported nonsignificant differences. The scarcity of negative results is especially marked in autism research (28), a very damaging tendency that impedes us from deciding on which hypotheses are worth pursuing: our meta-analysis was well powered to detect even small differences; however, selective publication makes the results uninterpretable.

Despite the appeal of the hypothesis of a smaller CC in ASD, we need to consider the possibility that it may not be true. Does this falsify or weaken the underconnectivity theory? The underconnectivity theory states that autism is caused by insufficient integration circuitry (3). But whereas many articles in different subfields of autism research have indicated that their findings support (verify) the underconnectivity hypothesis, it is not clear what finding would be necessary to prove it wrong (falsify it). As pointed out by Braitenberg (29), most of the brain tissue could be considered as circuitry (myelinated and nonmyelinated axons, dendrites), the main role of which is undoubtedly some type of integration. It is not entirely surprising that autism, as other psychiatric disorders, can be interpreted as some type of insufficiency in integration circuitry. We believe that to progress, autism research requires a

theoretical framework with stronger, more clearly falsifiable predictions. In particular, a smaller CC has not been directly stated as a prediction of the underconnectivity theory. If our aim were to preserve the theory, we could simply add an ad hoc clause (autism is caused by an insufficiency in integration circuitry of a type that does not change the size of the corpus callosum), but the appropriateness of this approach has been criticized (30).

Our results suggest that nonlinear variations in CC size relative to BV present in the general population or different patterns of covariation of confounding factors could lead to some of the group differences reported in the literature. We found that 19% of the variance in CC size was captured by a relationship with BV, where progressively larger brains had a proportionally smaller CC. If for some reason BV in one population were different than in the other, this could lead to a group difference in CC size (for example, if female and male subjects were compared). Our analyses showed that normalization of CC by BV (which supposed isometric scaling) is insufficient to control for differences in BV. Including BV as a covariate provides a more reliable control and should be preferred to normalization. Besides a real difference in BV, we showed that IQ matching could, under certain circumstances, induce an artifactual BV difference that could be later observed as a CC size difference. Finally, besides the allometric scaling of CC and BV, similar nonlinear relationships have been also observed between total cortical surface and BV, between folding and BV, and between white matter volume and BV. Because of these nonlinear scaling relationships, it is expected that subjects with larger BV will have a larger cortical surface area, more folded particularly in the prefrontal cortex (16), and with a larger frontal white matter volume (17). These are exactly the findings that have been reported in several articles comparing patients with ASD and control subjects (3,31–33). The extent to which they arise from a difference in BV (real or artifactual) has yet to be evaluated.

We found that intelligence scores do not covary with BV in the same way in ASD and control subjects. This is not completely surprising, as FIQ—and VIQ in particular—are known to be affected by environmental factors such as stress or socioeconomic status (19). Psychiatric disorders such as ASD do not only impose to the patients the cognitive challenges that we most often use to define them but also a daily confrontation with various comorbidities and various degrees of social, educational, and daily life difficulties depending on our society's ability to integrate them. This additional burden is very likely to leave physiological traces, in particular neuroanatomical. Finding the biological markers of the causes of ASD may then require us to disentangle them not only from risk factors but also from the social effects of being different (handicap, disadvantage, etc.).

Research suggests today that the etiology of ASD is highly heterogeneous, with hundreds of genetic mutations associated with it (34,35), as well as many environmental factors. The patient's phenotypes are also so diverse, with the presence of such large numbers of different comorbidities and wide spectrum of cognitive abilities, that one could wonder about the relevance of looking for neuroanatomical traits common to all of them. It is important, however, to remember that the nervous system is strongly self-regulated and capable of striking plasticity: the processes leading to the formation of a mammalian nervous system are very resilient and able to produce viable cognitive function under the most extreme circumstances. The phenomena that we may be able to observe at the scale of the complete nervous system will be more likely the trace of this shared mechanism of developmental canalization and compensation than a direct reflection of the heterogeneous etiology. In this sense, the comprehension of the normal response of the nervous system to perturbation and the way in which the diversity of this response is regulated by the complete individual's genetic background may turn out to be as important as the direct study of the pathologic cause. The study of large cohorts with extensive behavioral, genetic, and neuroimaging data will be of fundamental relevance for this objective.

ACKNOWLEDGMENTS AND DISCLOSURES

This work was funded by the Institut Pasteur, CNRS, INSERM, AP-HP, University Paris Diderot, the Bettencourt-Schueller foundation, the Orange foundation, the FondaMental foundation, the Conny-Maeva foundation, the Cognacq-Jay foundation, the ANR (SynDivAutism), Neuron-ERANET (EUHF-AUTISM). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. All authors report no biomedical financial interests or potential conflicts of interest.

ARTICLE INFORMATION

From the Human Genetics and Cognitive Functions Unit (AL, AB, TB, RT), Institut Pasteur, Paris; Unité Mixte de Recherche 3571, Genes, Synapses and Cognition (TB, RT), Centre National de la Recherche Scientifique, Institut Pasteur, Paris; Human Genetics and Cognitive Functions (TB, RT), University Paris Diderot, Sorbonne Paris Cité, Paris; Department of Child and Adolescent Psychiatry (AL, AB), Assistance Publique-Hôpitaux de Paris, Robert Debré Hospital; and Foundation Fondamentale (TB), Créteil, France.

Address correspondence to Roberto Toro, Ph.D., Institut Pasteur, Human Genetics and Cognitive Functions, 25, rue du Docteur Roux, Paris 75015, France; E-mail: rto@pasteur.fr.

Received Aug 27, 2014; revised Jan 31, 2015; accepted Feb 2, 2015.

Supplementary material cited in this article is available online at http://dx.doi.org/10.1016/j.biopsych.2015.02.010.

REFERENCES

1. Frith U (2003): Autism: Explaining the Enigma, 2nd ed. Malden, MA: Blackwell Publishers.
2. Courchesne E, Pierce K (2005): Why the frontal cortex in autism might be talking only to itself: Local over-connectivity but long-distance disconnection. Curr Opin Neurobiol 15:225–230.
3. Just MA, Cherkassky VL, Keller TA, Kana RK, Minshew NJ (2007): Functional and anatomical cortical underconnectivity in autism: Evidence from an fMRI study of an executive function task and corpus callosum morphometry. Cereb Cortex 17:951–961.
4. Mengotti P, Brambilla P (2014): Developmental trajectory of white matter and connectivity maturation in autism. In: Patel VB, Preedy VR, Martin CR, editors. The Comprehensive Guide to Autism. New York. New York: Springer, 911–927.
5. Hughes JR (2009): Update on autism: A review of 1300 reports published in 2008. Epilepsy Behav 16:569–589.
6. Ringo JL, Doty RW, Demeter S, Simard PY (1994): Time is of the essence: A conjecture that hemispheric specialization arises from interhemispheric conduction delay. Cereb Cortex 4:331–343.
7. Rilling JK, Insel TR (1999): Differential expansion of neural projection systems in primate brain evolution. Neuroreport 10:1453–1459.

8. Aboitiz F, López J, Montiel J (2003): Long distance communication in the human brain: Timing constraints for inter-hemispheric synchrony and the origin of brain lateralization. Biol Res 36:89–99.
9. Booth R, Wallace GL, Happé F (2011): Connectivity and the corpus callosum in autism spectrum conditions: Insights from comparison of autism and callosal agenesis. Prog Brain Res 189:303–317.
10. Paul LK, Brown WS, Adolphs R, Tyszka JM, Richards LJ, Mukherjee P, Sherr EH (2007): Agenesis of the corpus callosum: Genetic, developmental and functional aspects of connectivity. Nat Rev Neurosci 8:287–299.
11. Paul LK, Corsello C, Kennedy DP, Adolphs R (2014): Agenesis of the corpus callosum and autism: A comprehensive comparison. Brain 137:1813–1829.
12. Di Martino A, Yan C-G, Li Q, Denio E, Castellanos FX, Alaerts K, *et al*. (2014): The autism brain imaging data exchange: Towards a large-scale evaluation of the intrinsic brain architecture in autism. Mol Psychiatry 19:659–667.
13. Frazier TW, Hardan AY (2009): A meta-analysis of the corpus callosum in autism. Biol Psychiatry 66:935–941.
14. Rothstein H, Sutton AJ, Borenstein M (2005): Publication bias in meta-analysis: Prevention, assessment and adjustments. Chichester, England: Wiley.
15. Gotham K, Pickles A, Lord C (2009): Standardizing ADOS Scores for a measure of severity in autism spectrum disorders. J Autism Dev Disord 39:693–705.
16. Toro R, Perron M, Pike B, Richer L, Veillette S, Pausova Z, Paus T (2008): Brain size and folding of the human cerebral cortex. Cereb Cortex 18:2352–2357.
17. Toro R, Chupin M, Garnero L, Leonard G, Perron M, Pike B, *et al*. (2009): Brain volumes and Val66Met polymorphism of the BDNF gene: Local or global effects? Brain Struct Funct 213:501–509.
18. Luders E, Toga AW, Thompson PM (2014): Why size matters: Differences in brain volume account for apparent sex differences in callosal anatomy: The sexual dimorphism of the corpus callosum. Neuroimage 84:820–824.
19. Hanscombe KB, Trzaskowski M, Haworth CMA, Davis OSP, Dale PS, Plomin R (2012): Socioeconomic status (SES) and children's intelligence (IQ): In a UK-representative sample SES moderates the environmental, not genetic, effect on IQ. PloS One 7:e30320.
20. Kanner L (1943): Autistic disturbances of affective contact. Nervous Child 2:217–250.
21. Lainhart JE, Bigler ED, Bocian M, Coon H, Dinh E, Dawson G, *et al*. (2006): Head circumference and height in autism: A study by the Collaborative Program of Excellence in Autism. Am J Med Genet A 140:2257–2274.
22. Courchesne E, Campbell K, Solso S (2011): Brain growth across the life span in autism: Age-specific changes in anatomical pathology. Brain Res 1380:138–145.
23. Chaste P, Klei L, Sanders SJ, Murtha MT, Hus V, Lowe JK, *et al*. (2013): Adjusting head circumference for covariates in autism: Clinical correlates of a highly heritable continuous trait. Biol Psychiatry 74:576–584.
24. Raznahan A, Wallace GL, Antezana L, Greenstein D, Lenroot R, Thurm A, *et al*. (2013): Compared to what? Early brain overgrowth in autism and the perils of population norms. Biol Psychiatry 74: 563–575.
25. Button KS, Ioannidis JPA, Mokrysz C, Nosek BA, Flint J, Robinson ESJ, Munafò MR (2013): Power failure: Why small sample size undermines the reliability of neuroscience. Nat Rev Neurosci 14: 365–376.
26. Ioannidis JPA (2005): Why most published research findings are false. PLoS Med 2:0020124.
27. Haar S, Berman S, Behrmann M, Dinstein I. (2014): Anatomical abnormalities in autism? [published online ahead of print October 14]. Cereb Cortex.
28. Joober R, Schmitz N, Annable L, Boksa P (2012): Publication bias: What are the challenges and can they be overcome? J Psychiatry Neurosci 37:149–152.
29. Braitenberg V (2001): Brain size and number of neurons: An exercise in synthetic neuroanatomy. J Comput Neurosci 10:71–77.
30. Popper KR (2000): The Logic of Scientific Discovery. London: Routledge.
31. Carper RA, Moses P, Tigue ZD, Courchesne E (2002): Cerebral lobes in autism: Early hyperplasia and abnormal age effects. Neuroimage 16:1038–1051.
32. Herbert MR, Ziegler DA, Makris N, Filipek PA, Kemper TL, Normandin JJ, *et al*. (2004): Localization of white matter volume increase in autism and developmental language disorder. Ann Neurol 55:530–540.
33. Amaral DG, Schumann CM, Nordahl CW (2008): Neuroanatomy of autism. Trends Neurosci 31:137–145.
34. Toro R, Konyukh M, Delorme R, Leblond C, Chaste P, Fauchereau F, *et al*. (2010): Key role for gene dosage and synaptic homeostasis in autism spectrum disorders. Trends Genet 26:363–372.
35. Huguet G, Ey E, Bourgeron T (2013): The genetic landscapes of autism spectrum disorders. Annu Rev Genomics Hum Genet 14: 191–213.
36. Gaffney GR, Kuperman S, Tsai LY, Minchin S, Hassanein KM (1987): Midsagittal magnetic resonance imaging of autism. Br J Psychiatry 151:831–833.
37. Egaas B, Courchesne E, Saitoh O (1995): Reduced size of corpus callosum in autism. Arch Neurol 52:794–801.
38. Piven J, Bailey J, Ranson BJ, Arndt S (1997): An MRI study of the corpus callosum in autism. Am J Psychiatry 154:1051–1056.
39. Manes F, Piven J, Vrancic D, Nanclares V, Plebst C, Starkstein SE (1999): An MRI study of the corpus callosum and cerebellum in mentally retarded autistic individuals. J Neuropsychiatry Clin Neurosci 11:470–474.
40. Elia M, Ferri R, Musumeci SA, Panerai S, Bottitta M, Scuderi C (2000): Clinical correlates of brain morphometric features of subjects with low-functioning autistic disorder. J Child Neurol 15:504–508.
41. Rice SA, Bigler ED, Cleavinger HB, Tate DF, Sayer J, McMahon W, *et al*. (2005): Macrocephaly, corpus callosum morphology, and autism. J Child Neurol 20:34–41.
42. Vidal CN, Nicolson R, DeVito TJ, Hayashi KM, Geaga JA, Drost DJ, *et al*. (2006): Mapping corpus callosum deficits in autism: An index of aberrant cortical connectivity. Biol Psychiatry 60:218–225.
43. Boger-Megiddo I, Shaw DWW, Friedman SD, Sparks BF, Artru AA, Giedd JN, *et al*. (2006): Corpus callosum morphometrics in young children with autism spectrum disorder. J Autism Dev Disord 36:733–739.
44. Alexander AL, Lee JE, Lazar M, Boudos R, DuBray MB, Oakes TR, *et al*. (2007): Diffusion tensor imaging of the corpus callosum in autism. Neuroimage 34:61–73.
45. Hardan AY, Pabalan M, Gupta N, Bansal R, Melhem NM, Fedorov S, *et al*. (2009): Corpus callosum volume in children with autism. Psychiatry Res 174:57–61.
46. Freitag CM, Luders E, Hulst HE, Narr KL, Thompson PM, Toga AW, *et al*. (2009): Total brain volume and corpus callosum size in medication-naïve adolescents and young adults with autism spectrum disorder. Biol Psychiatry 66:316–319.
47. Keary CJ, Minshew NJ, Bansal R, Goradia D, Fedorov S, Keshavan MS, Hardan AY (2009): Corpus callosum volume and neurocognition in autism. J Autism Dev Disord 39:834–841.
48. Anderson JS, Druzgal TJ, Froehlich A, DuBray MB, Lange N, Alexander AL, *et al*. (2011): Decreased interhemispheric functional connectivity in autism. Cereb Cortex 21:1134–1146.
49. Hong S, Ke X, Tang T, Hang Y, Chu K, Huang H, *et al*. (2011): Detecting abnormalities of corpus callosum connectivity in autism using magnetic resonance imaging and diffusion tensor tractography. Psychiatry Res 194:333–339.
50. Frazier TW, Keshavan MS, Minshew NJ, Hardan AY (2012): A two-year longitudinal MRI study of the corpus callosum in autism. J Autism Dev Disord 42:2312–2322.
51. Prigge MBD, Lange N, Bigler ED, Merkley TL, Neeley ES, Abildskov TJ, *et al*. (2013): Corpus callosum area in children and adults with autism. Res Autism Spectr Disord 7:221–234.