Differences in the right inferior longitudinal fasciculus but no general disruption of white matter tracts in children with autism spectrum disorder

Kami Koldewyn[a,b,1], Anastasia Yendiki[c], Sarah Weigelt[a,d], Hyowon Gweon[a], Joshua Julian[a], Hilary Richardson[a], Caitlin Malloy[a], Rebecca Saxe[a], Bruce Fischl[c,e], and Nancy Kanwisher[a,1]

[a]Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139; [b]School of Psychology, Bangor University, Gwynedd LL57 2AS, United Kingdom; [c]Athinoula A. Martinos Center for Biomedical Imaging, Department of Radiology, Massachusetts General Hospital and Harvard Medical School, Charlestown, MA 02129; [d]Fakultät für Psychologie, Ruhr-Universität Bochum, D-44801 Bochum, Germany; and [e]Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139

Contributed by Nancy Kanwisher, December 24, 2013 (sent for review December 8, 2013)

One of the most widely cited features of the neural phenotype of autism is reduced "integrity" of long-range white matter tracts, a claim based primarily on diffusion imaging studies. However, many prior studies have small sample sizes and/or fail to address differences in data quality between those with autism spectrum disorder (ASD) and typical participants, and there is little consensus on which tracts are affected. To overcome these problems, we scanned a large sample of children with autism (n = 52) and typically developing children (n = 73). Data quality was variable, and worse in the ASD group, with some scans unusable because of head motion artifacts. When we follow standard data analysis practices (i.e., without matching head motion between groups), we replicate the finding of lower fractional anisotropy (FA) in multiple white matter tracts. However, when we carefully match data quality between groups, all these effects disappear except in one tract, the right inferior longitudinal fasciculus (ILF). Additional analyses showed the expected developmental increases in the FA of fiber tracts within ASD and typical groups individually, demonstrating that we had sufficient statistical power to detect known group differences. Our data challenge the widely claimed general disruption of white matter tracts in autism, instead implicating only one tract, the right ILF, in the ASD phenotype.

diffusion-weighted imaging | connectivity

What is the key difference in the brains of individuals with autism that accounts for the distinctive cognitive profile of this disorder? One of the most widely claimed brain signatures of autism spectrum disorder (ASD), reported in dozens of papers that used diffusion-weighted imaging (DWI), is reduced integrity of long-range fiber tracts (1). This finding has been taken as evidence that autism is fundamentally a "disconnection" syndrome, in which the core cognitive deficits result from reduced integration of information at the neural and cognitive levels (2–5). For example, it has been argued that the characteristic deficits in social cognition and language arise because these functions require rapid integration of information across spatially distant brain areas (3, 6, 7), which would likely be affected if major white matter tracts are compromised.

Evidence for a general reduction in the "integrity"* of white matter in autism has come primarily from diffusion imaging studies that report reduced directionality of the diffusion of water molecules, or fractional anisotropy (FA), and increased speed of diffusion, or mean diffusivity (MD) of many major fiber bundles. However, the literature reveals little actual agreement on the existence and direction of group differences in diffusion parameters (reviewed in ref. 1). White-matter differences have been reported in various brain regions in positive and negative directions. Possible reasons for these inconsistent findings include small sample sizes [mean of ∼20 in each group, with 40% of studies scanning 15 or fewer participants with ASD (1)], the heterogeneity of ASD itself, variations across studies in the age of the cohort tested, and the type of DTI analysis performed. Another potential problem that few diffusion studies of autism address or even mention is data quality. Indeed, to our knowledge, only two studies (9, 10) report quantitative analyses of the amount of motion in their DWI data. Group differences in head motion could be a serious confounding factor, given that head motion is likely to be greater in children with autism, and group differences in head motion can lead artifactually to just the effects most often reported: reduced FA in white matter tracts in ASD (11).

To address these concerns, we scanned a relatively large sample of children with and without ASD, and evaluated data quality from each participant by visual inspection of the data and quantification of head motion (11). We then excluded scans that did not reach our data quality criterion, and matched the remaining participants across groups for data quality. These data were used to determine whether people with autism do in fact show widespread differences in the known white matter tracts in ASD. We further tested the specific hypothesis that individuals with ASD show changes in one particular tract, the inferior longitudinal fasciculus (ILF), a white matter tract important for

Significance

One of the most accepted brain "signatures" of autism spectrum disorder (ASD) is a reduction in the integrity of long-range white-matter fiber tracts. Here, we assessed known white matter tracts in children with ASD by using diffusion-weighted imaging. In contrast to most prior studies, we carefully matched for head motion between groups. When data quality was matched, there was no evidence of widespread changes in white-matter tracts in the ASD group. Instead, differences were present in only one tract, the right inferior longitudinal fasciculus. These data challenge the idea that widespread changes in white-matter integrity are a signature of ASD and highlight the importance of matching for data quality in future diffusion studies of ASD and other clinical disorders.

Author contributions: K.K., A.Y., R.S., B.F., and N.K. designed research; K.K., S.W., H.G., J.J., H.R., and C.M. performed research; K.K. and A.Y. analyzed data; and K.K. and N.K. wrote the paper.

The authors declare no conflict of interest.

[1]To whom correspondence may be addressed. E-mail: k.koldewyn@bangor.ac.uk or ngk@mit.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1324037111/-/DCSupplemental.

*Although reductions in FA are often used to argue for reduced "integrity" of white matter tracts, the precise anatomical correlates of reduced FA are not clear (8). In the present paper, we therefore interpret reduced FA in ASD to show only differences in white matter, without assuming that such differences constitute reductions in the integrity of those tracts.

face recognition (12, 13), a mental function selectively disrupted in ASD (ref. 14; but see ref. 15).

Results

A total of 21.3% of our diffusion scans on children with ASD and 11.3% of the scans on typical children do not meet even our more liberal threshold for usable data quality (*Methods* provides details on how data quality was assessed and quantified). Not surprisingly, the scans from children with ASD have significantly lower data quality than those from the typically developing (TD) children on each of our four measures of head motion/data quality (*P* values in Fig. 1). When these bad scans are excluded from the analysis (leaving $n = 40$ in the ASD group and $n = 71$ TD children; Fig. 1, "all scans without visible artifacts in FA maps"), and age, intelligence quotient (IQ), and sex are controlled for, we find significantly lower FA in many tracts (Fig. 2, column A), with a significant group-by-tract interaction [$F(17, 1{,}836) = 2.73$, $P = 0.001$] and a main effect of group just short of significance [$F(1, 108) = 3.10$, $P = 0.08$]. These results qualitatively resemble prior findings in the literature. However, data quality was significantly lower in the ASD group (Fig. 1), a situation we have shown can lead to spurious group differences in FA (11).

To equalize data quality across groups, we therefore identified two subsets of our data matched across groups for age, IQ, sex, and five measures of head motion and data quality. One subset of the data used a relatively liberal threshold for data inclusion (the absence of visible artifacts in the raw FA maps); this cohort enabled us to preserve a larger number of subjects (called from here on the main cohort). A second subset of the scans used a more stringent threshold for data quality (the absence of visible artifacts in raw DWI images) and a correspondingly smaller number of subjects (called from here on the stringent cohort). Fig. 1 provides details of the demographics in each group for each analysis.

Testing for Changes in FA, MD, and Radial Diffusivity in ASD in the Main Cohort. The main question of this study was whether widespread differences in major white matter tracts in autism are found when data quality is matched between ASD and typical participants (*Methods* provides details on how these white matter tracts were identified and assessed). The data from our main cohort (ASD, $n = 40$; TD, $n = 43$) provide no evidence for this hypothesis. A two-way ANOVA on FA as a function of group (ASD vs. TD) by tract found no main effect of group [$F(1, 81) = 0.591$, $P = 0.444$], and no group-by-tract interaction [$F(17, 1{,}377) = 1.358$, $P = 0.149$]. Post hoc comparisons of FA in each tract individually (Fig. 2) found that only one tract was significantly different between groups: the right ILF (rILF), which showed lower FA in the ASD group [$t(81) = 3.119$, $P = 0.003$].

	pIQ (K-bit)	Age (years)	SRS (total)	ADOS (CS)	Translation	Rotatation	Benner Score	% Bad Slices
All Scans before quality control measures								
ASD (n = 52) *(8 girls; 75 scans)*	107.4 (16.4)	8.88 (1.84)	78.9 (9.11)	7.35 (1.67)	1.179 (.6790)	.0112 (.0080)	.1185 (.1256)	.2635 (.5512)
TD (n = 73) *(15 girls;115 scans)*	117.4 (16.3)	8.64 (1.93)	46.7 (7.28)	-	.8494 (.4370)	.0068 (.0048)	.0605 (1001)	.0881 (.2139)
p value	*.001*	*.397*	*<.001*	-	*<.001*	*<.001*	*.001*	*.002*
All Scans without visible artifacts in FA maps								
ASD (n = 40) *(5 girls; 59 scans)*	106.9 (17.6)	8.98 (1.81)	78.9 (9.11)	7.35 (1.67)	.8691 (.0905)	.0075 (.0070)	.0801 (.1394)	.0921 (.2448)
TD (n = 71) *(14 girls; 102 scans)*	117.8 (14.7)	8.62 (1.94)	46.7 (7.28)	-	.7451 (.2752)	.0055 (.0028)	.0402 (.0829)	.0447 (.0991)
p value	*.001*	*.352*	*<.001*	-	*.013*	*<.001*	*.016*	*.016*
Matched Scans without visible artifacts in FA maps (Main Cohort)								
ASD (n = 40) *(5 girls)*	106.9 (17.6)	8.98 (1.81)	78.9 (9.11)	7.35 (1.67)	.8730 (.3639)	.0077 (.0044)	.0808 (.1298)	.0962 (.1567)
TD (n = 43) *(8 girls)*	110.8 (14.3)	8.90 (1.87)	46.2 (6.51)	-	.8285 (.3041)	.0067 (.0034)	.0730 (.1103)	.0699 (.1261)
p value	*.271*	*.854*	*<.001*	-	*.545*	*.263*	*.767*	*.401*
All Scans without visible artifacts in raw DWI images								
ASD (n = 17) *(5 girls,19 scans)*	107.6 (18.3)	9.32 (1.41)	78.9 (8.30)	7.00 (1.58)	.6553 (.1767)	.0048 (.0019)	.0065 (.0282)	.0037 (.0163)
TD (n = 42) *(8 girls, 50 scans)*	117.5 (12.8)	8.92 (2.03)	45.9 (6.82)	-	.6560 (.2428)	.0041 (.0013)	.0091 (.0233)	.0149 (.0452)
p value	*.021*	*.431*	*<.001*	-	*.990*	*.042*	*.489*	*.219*
Matched Scans without visible artifacts in raw DWI images (Stringent Cohort)								
ASD (n = 17) *(5 girls)*	107.6 (18.3)	9.22 (1.46)	78.9 (8.30)	7.00 (1.58)	.6649 (.1815)	.0050 (.0019)	.0072 (.0300)	.0042 (.0173)
TD (n = 21) *(5 girls)*	113.4 (12.3)	9.18 (2.11)	45.9 (6.94)	-	.6980 (.2826)	.0043 (.0014)	.0096 (.0265)	.0085 (.0201)
p value	*.260*	*.937*	*<.001*	-	*.677*	*.283*	*.801*	*.467*

Image Quality · Matching · Image Quality · Matching

Fig. 1. The steps taken to ensure acceptable data quality across all scans and data quality matching between groups. During "image quality" steps, subsets of data were selected based on visual inspection of each scan for artifacts in raw DWIs or FA maps. During "matching" steps, subsets of data were selected that enabled us to match ASD and TD groups for age, IQ, sex, and data quality measures. *P* values show the significance of the differences between the ASD and TD groups on the indicated dimensions for each column. Black-outlined rectangles denote the two cohorts of participants analyzed: the main cohort (*Middle*) and the stringent cohort (*Bottom*).

Between-group differences in FA per tract in the Main Analysis

Tract	A	B	TD Mean	ASD Mean	95% C.I.	Cohen's d	Bayes Factor
f major	↓*	↓	.6505	.6462	-.020 - .029	.0768	5.63
f minor	↓*	↓	.5361	.5277	-.012 - .029	.1807	4.36
L ATR	↑	↑	.3789	.3790	-.015 - .015	.0031	5.96
R ATR	↑	↑	.3709	.3728	-.017 - .013	.0538	5.79
L CAB	↓*	↑	.3630	.3688	-.025 - .013	.1324	5.03
R CAB	↓	↓	.3774	.3538	-.000 - .047	.3661	2.15
L CCG	↓	↓	.4788	.4779	-.022 - .024	.0185	5.94
R CCG	↓	↓	.4450	.4443	-.021 - .022	.0148	5.95
L CST	↓	↓	.5286	.5275	-.018 - .021	.0246	5.92
R CST	↓	↓	.5169	.5151	-.018 - .022	.0389	5.87
L ILF	↓*	↓	.5212	.5063	-.004 - .034	.3359	2.03
R ILF	↓*	↓*	.5357	.5094	.010 - .043	.6840	0.08
L SLFP	↑	↑	.4393	.4436	-.019 - .010	.1318	5.05
R SLFP	↑	↑	.4284	.4292	-.015 - .019	.0259	5.92
L SLFT	↓	↓	.4555	.4526	-.010 - .016	.0985	5.43
R SLFT	↓	↓	.4328	.4280	-.009 - .019	.0261	4.79
L Unc	↑	↑	.4224	.4229	-.014 - .013	.0151	3.93
R Unc	↓	↑	.4120	.4160	-.019 - .011	.1156	5.24

Fig. 2. For each of the 18 major fiber tracts analyzed (left column), we show the direction and significance of the mean difference in FA between groups (an upward-pointing arrow means the ASD group showed higher FA in that tract than the TD group) both before matching groups on data quality (column A) and after matching groups (main cohort, column B). Also shown are (after matching) the means for each group, the 95% CI around the mean difference between groups, effect size (Cohen *d*) for this group difference in FA, and the Jeffreys–Zellner–Siow Bayes factor, reflecting the odds that there is no difference between groups vs. the alternative (a true difference between groups) in the form of odds-of-no-difference/alternative. ATR, anterior thalamic radiations; CAB, cingulum–angular bundle; CCG, cingulum–cingulate gyrus bundle; CST, corticospinal tract; f major, corpus callosum–forceps major; f minor, corpus callosum–forceps minor; SLFP, superior longitudinal fasciculus–parietal terminations; SLFT, superior longitudinal fasciculus–temporal terminations; UNC, uncinate fasciculus.

This specific finding was predicted before the data were analyzed, and also survives Bonferroni correction for the number of tracts tested (18).

Similarly, analyses of MD failed to find a significant group difference [$F(1, 81) = 1.63$, $P = 0.206$] or group-by-tract interaction [$F(17, 1{,}377) = 0.850$, $P = 0.635$]. Likewise, there was no main effect of group for radial diffusivity (RD) [$F(1, 81) = 1.388$, $P = 0.242$], and no group-by-tract interaction [$F(17, 1{,}377) = 1.392$, $P = 0.131$]. The rILF showed higher RD in ASD, although this difference did not survive correction for multiple comparisons [$t(81) = 2.46$, $P = 0.016$].

Prior reports suggest that differences between people with ASD and TD individuals may vary across development (16). Might differences in white matter tracts be present in just younger children, or just older children? Our data provide no support for this hypothesis: In a test run on just the 20 oldest children in each group in the main cohort (ASD mean age, 10.45 y; TD mean age, 10.4 y; matched for age and motion), we found no significant main effect of group on FA [$F(1, 38) = 0.02$, $P = 0.983$] and no significant group-by-tract interaction [$F(17, 646) = 1.34$, $P = 0.163$]. The same was true for MD (main effect of group, $F(1, 38) = 0.146$, $P = 0.705$; group-by-tract interaction, $F(17, 646) = 1.08$, $P = 0.367$] and RD [main effect of group, $F(1, 38) = 0.092$, $P = 0.764$; group-by-tract interaction, $F(17, 646) = 1.244$, $P = 0.224$]. FA in the rILF was still significantly different between ASD and TD groups in just this older cohort ($t = 2.515$, $P = 0.016$).

Similarly, our data provide no support for the hypothesis of white matter tract differences in only younger children with ASD. In a test run on just the 20 youngest children in each group in our main cohort (mean age for both groups, 7.5 y; matched for age and motion), we found no significant main effect of group on FA [$F(1, 38) = 2.34$, $P = 0.134$] and no group-by-tract interaction [$F(17, 646) = 0.565$, $P = 0.918$]. Similarly, for MD, we also found no main effect of group [$F(1, 38) = 1.048$, $P = 0.312$] and no group-by-tract interaction [$F(17, 646) = 1.23$, $P = 0.235$], and, for RD, we found no main effect of group [$F(1, 38) = 1.88$, $P = 0.181$] or group-by-tract interaction [$F(17, 646) = 0.801$, $P = 0.691$]. The younger children also showed a significant reduction in the FA of the rILF in the ASD group [$t(38) = 2.02$, $P = 0.05$]. Thus, with the one exception of the rILF, for which we specifically predicted a reduction of FA in ASD, our main cohort data provide no evidence for a widespread reduction of FA in ASD.

Testing for Changes in FA, RD, and MD in ASD in the Stringent Cohort. Do the main findings described earlier remain when even more stringent criteria of data quality are applied? To address this question, we reran the aforementioned analyses on the stringent cohort (ASD, $n = 17$; TD, $n = 21$), who were not only matched for head motion across groups but who also showed no evidence of motion artifacts on raw diffusion images or FA maps when visually inspected by an expert. This analysis still provides no evidence of general differences in diffusion measures of white-matter tracts between children with ASD and TD children. A two-way ANOVA on FA as a function of group (ASD vs. TD) by tract found no main effect of group [$F(1, 36) = 1.58$, $P = 0.216$] and no group-by-tract interaction [$F(17, 612) = 0.926$, $P = 0.543$]. Post hoc comparisons of each tract individually found the rILF to have significantly lower FA in the ASD group [$t(36) = 2.17$, $P = 0.037$, uncorrected for multiple comparisons]. Parallel analyses failed to find significant group effects or group by tract interactions for MD [$F(1, 36) = 2.72$, $P = 0.108$], with no group-by-tract interaction [$F(17, 612) = 0.910$, $P = 0.562$], or RD [$F(1, 36) = 2.20$, $P = 0.147$], and no group-by-tract interaction [$F(17, 612) = 1.00$, $P = 0.456$].

Testing for Expected Developmental Change in White Matter Tracts. The preceding analyses failed to replicate the "standard" finding of a general reduction of FA in major fiber tracts in autism (except in the rILF). Can we be sure this failure to replicate previous findings is not a result of insufficient statistical power? To find out, we asked whether our data exhibit the well established increases in FA with age (17, 18). We therefore divided children into two groups, one younger and one older. It was impossible to match all four groups (older/younger children × ASD/TD) for head motion, so, for this analysis, we conducted separate ANOVAs for each group (ASD and TD), with tract as one factor and age group as the other factor, while matching head motion across the younger vs. older children within each group separately.

Main cohort, typical subjects. First, choosing from our main cohort of typical participants, we found 34 older (mean age, 10.27 y) and 34 younger (mean age, 6.99 y) children who could be matched for motion measures and IQ (all $P > 0.2$). A two-way ANOVA on FA as a function of age group–by–tract on these data found a main effect of group [$F(1, 66) = 25.10$, $P < 0.001$], as well as an age group–by–tract interaction [$F(17, 1{,}122) = 2.287$, $P = 0.002$]. These effects reflect widespread increases in the FA of white-matter tracts with age (Fig. S1, *Left*), with robust bilateral changes in tracts across the brain.

Main cohort, ASD subjects. In a parallel analysis on children with ASD, we contrasted 20 older (mean age, 10.62 y) and 20 younger (mean age, 7.27 y) children, matched for head motion and IQ. A two-way ANOVA on FA as a function of age group and tract found a main effect of group [$F(1, 38) = 6.44$, $P = 0.015$], but no

age group–by–tract interaction [$F(17, 646) = 0.929$, $P = 0.539$]. These effects reflect widespread increases in the FA of white-matter tracts with age, much like the findings seen in the typical group (Fig. S1, *Right*).

Stringent cohort. In the stringent cohort, we had 20 typical participants in each age group matched for head motion and IQ (mean ages, 10.29 y for older group and 6.96 y for younger group). A two-way ANOVA on FA as a function of age group and tract found a main effect of age group [$F(1, 38) = 4.209$, $P = 0.04$] but no age group–by–tract interaction [$F(17, 646) = 1.27$, $P = 0.208$]. The same analysis for ASD participants in the stringent cohort identified only eight children in each age group (older mean age, 10.34; younger mean age, 7.94) that could be matched for motion and IQ. Despite the small number of subjects, a two-way ANOVA on FA as a function of age group found a main effect of group [$F(1, 14) = 8.73$, $P = 0.010$], but no age group–by–tract interaction [$F(17, 646) = 0.717$, $P = 0.784$].

In sum, all four analyses show robust increases in FA with age. Evidently, our data are of sufficient quality, and our analyses of sufficient power, to detect known group differences in FA between groups. Note that half as much data went into the developmental analyses (because they were conducted within each group separately) compared with the main analysis comparing ASD vs. typical groups (across ages). Thus, we have substantially more power to detect differences in ASD if they exist compared with age differences, yet still we detected none (except for the predicted effect in the rILF). These results suggest that our failure to find differences between children with ASD and typical children in orthogonal analyses of the same data are unlikely to be a result of insufficient data quality or power, unless group differences in FA are substantially smaller than age differences. Note that, because no comparable prior study has reported effect sizes for the differences they report in FA between ASD and TD groups, it is not possible to determine whether our study had enough power to detect the effects reported in the prior literature.

Discussion

We used diffusion imaging in a large sample of children with and without autism to test the widespread claim that individuals with ASD show general disruption of long-range white matter tracts. Despite careful efforts to minimize head motion with the use of custom pediatric imaging coils and prior training in a mock scanner, a substantial percentage of the scans, and more in the autism group (21.3%) than the typical group (11.3%), were unusable because of head motion. When these bad scans were omitted, we replicated the standard finding that many white matter tracts show lower FA in ASD. However, because head motion strongly affects measures of diffusion (11, 19, 20), it is important to quantify this motion, exclude data not meeting a reasonable criterion of data quality, and match the remaining data for head motion/data quality across groups (10). By using these procedures (*Methods* and Fig. 1), we find no evidence for a general reduction in the integrity of white matter tracts in autism. This result is not likely caused by insufficient statistical power because our study included more subjects than most prior studies reporting such effects, and because we robustly detect the known increase in FA with age in the same data within ASD and typical groups. Further, we found a significant reduction in participants with autism in the diffusion anisotropy of the one major tract where we predicted this effect in advance: the rILF. These data argue against general changes in white matter tracts across the brain in autism, instead demonstrating a more specific effect on just the rILF.

How can our findings be reconciled with the prior literature suggesting general differences in white matter tracts in autism? One possibility is that many prior studies have been affected by head motion artifacts. Indeed, our own study found that, despite great effort to minimize head motion, it remains a substantial problem when scanning children, and is significantly worse for children with ASD. We see no reason to think that head motion would be less severe in prior diffusion studies in children with ASD. However, only a few diffusion imaging studies of autism even mention possible differences in head motion, let alone measure it. Only two papers report any quantitative analysis of the amount of motion present in the DWI scans, or report what motion threshold was used for excluding participants (9, 10). Ten studies gathered imaging data while some participants were under general anesthesia and an additional two while participants were sleeping naturally, presumably reducing head motion. Of these studies, however, only one imaged both ASD and typical groups under anesthesia (21). Indeed, one study that imaged participants with ASD under anesthesia and typical participants while asleep concluded that at least some of their group differences were likely caused by differences in motion (22). Thus, few prior studies have adequately dealt with possible artifacts of head motion.

If ASD and TD groups did in fact differ in head motion in prior studies, could these differences account for the reported differences in FA? Consistent with this possibility, the present study also finds widespread reductions in FA when data quality is not matched between groups. Further, a parallel analysis of the present data set found that FA is correlated with head motion, and differences in head motion are sufficient to produce spurious differences in FA between groups (11). Most strikingly, when a group of TD subjects was scanned twice each, a contrast of the higher-motion scan vs. the lower-motion scan within the very same children found significant differences in FA between "groups" (11). Thus, the FA differences between TD and ASD groups previously reported could be partly or entirely a result of differences in head motion.

Several papers published in the past year highlight these concerns. In particular, the other main line of neural evidence for reduced long-range connections in autism has come from studies in which reduced correlations are found in ASD between brain regions in the time course of the functional MRI signal at rest (reviewed in ref. 23). However, three recent papers have shown that head motion artifacts can produce functional connectivity patterns resembling those reported for autism, i.e., reduced long-range connectivity and increased local connectivity (24–26), and another recent paper in adults with ASD suggests there may be very little difference between groups in functional connectivity when head motion is carefully controlled (27). Thus, much of the prior evidence for reduced long-range connectivity in autism based on resting functional studies could also be an artifact of head motion (28). One very recent paper pooled functional connectivity scans from 17 different sites and 539 people with ASD across a wide age and IQ range and used data scrubbing techniques to try to mitigate residual artifacts from head motion. Although this study found statistically significant differences in functional connectivity between those with ASD and TD children (29), important questions for the future are (*i*) whether comparable differences in diffusion measures of connectivity would be found in similarly large samples, and (*ii*) whether effect sizes so small they can only be detected with extremely large samples are theoretically significant (30). In any event, the problem with head motion in functional correlation studies that use more standard sample sizes underlines the importance of matching for head motion in diffusion studies.

Beyond the widespread failure to control for head motion in past studies, several other factors could explain some of the differences between our results and the previous literature. Although we see no reason why this should be the case, differences in analysis methods (e.g., tract-based vs. voxel-wise methods) could in principle account for some of the differences between results among previous studies, or between our results and those of previous studies (1). One might wonder in particular whether

our method, which was originally developed for adult subjects, works as well with child participants. As described in *Methods*, our tractography method relies on the segmentation of each participant's T1-weighted structural scan within FreeSurfer, a process that was originally based on adult participants. However, previous work has demonstrated that this method is robust and unbiased when tested on children in the same age range as those in the preent study (31). Further, because our tractography method uses information from the atlas about which anatomical labels each tract passes through or next to, and not about the exact spatial location or shape of the tract, it does not require perfect spatial alignment of study subjects to the atlas. Thus, we have no reason to believe our method is less robust for the age range we are studying than for adults. Finally, it is conceivable that widespread reductions in FA are present in individuals with ASD who are younger, older, or lower-functioning than those we tested here. The literature, however, provides little consistent support for any of these possibilities, especially given the dearth of imaging studies in individuals with ASD and severe intellectual impairment.

The one tract that did show a reduction in FA in autism in our data, even in our most stringent contrast, was the rILF. This tract was the only one for which we had a specific prediction of reduced FA in ASD based on prior studies. In particular, because we and others have found a selective deficit in face recognition in autism (reviewed in ref. 13), and people with congenital face recognition deficits have reduced FA in the rILF (12), we predicted that individuals with ASD might have reduced FA in the rILF. It remains a puzzle why reduced FA in the rILF does not produce across-the-board impairments in object recognition (in congenital prosopagnosia or ASD), given that this tract carries information from many extrastriate visual areas, not just face processing regions. One possibility is that differences in the rILF are specific to projections from face-processing regions, a prediction that could be tested with the "connectivity fingerprint" method devised recently by Saygin et al. (32).

In sum, the present study finds no evidence for the widespread claim that individuals with autism have a general reduction in long-range connectivity of white matter tracts. Instead, we find evidence for reduced connectivity of only one fiber tract, the rILF. These anatomical findings parallel the uneven cognitive profile of autism observed behaviorally, in which individuals with autism do not show across-the-board cognitive deficits, but instead show consistent deficits in only a few cognitive functions, most notably social cognition. Given the possibility that some prior reports of white matter tract differences in autism may be artifacts of differential head motion, it will be important to match for data quality in future diffusion studies of autism and other clinical disorders.

Methods

Participants. A total of 57 children with ASD and 73 TD children were scanned with DWI. Children with ASD were recruited through the Simons Foundation Autism Research Initiative (SFARI) Simons Simplex Collection (SSC) database and the Boston Autism Consortium. Children with ASD diagnoses were carefully characterized, including confirmation of the *Diagnostic and Statistical Manual of Mental Disorders* (33) diagnosis and Autism Diagnostic Observation Schedule (34) administration by expert clinicians. TD children were recruited from the local community. Potential participants were excluded if they had any history of birth or brain trauma or a nonverbal IQ of less than 80. TD participants were further excluded if they scored higher than 11 on the Social Communication Questionnaire (35), had any developmental disorder or mental illness, or an immediate family history of ASD. Participants received modest monetary compensation and small motivating prizes for their participation. Every participant signed an assent form and a parent or guardian signed an informed consent approved by the Massachusetts Institute of Technology Committee on the Use of Humans as Experimental Subjects.

Five children with ASD were later removed from the data set because they did not meet criteria for ASD on the ADOS ($n = 3$) or did not complete ADOS testing ($n = 2$). Some children were imaged on two occasions so the initial data set included 75 DWI scans from 52 children with ASD and 115 scans from 73 TD children. Demographic characteristics of these subjects are in Fig. 1.

Procedure. Two weeks before their visit, participants received a CD and illustrated booklet that introduced the experimenters children would meet, described the MRI procedure, and included recordings of scanner sounds. Earbuds similar to those they would wear in the scanner were also included so that children could become accustomed to them. Parents were encouraged to review all materials with their children and asked to help them practice lying still while listening to the noises of the scanner. Immediately before their scanning session, all children were trained for 15 to 30 min in a "mock" scanner, designed to simulate the appearance, noise, and confinement of the actual scanner. During these training sessions, children practiced lying still while watching a movie. The movie was turned off by a motion tracking system any time children moved too much in order to teach them how still they had to be to get "good brain pictures." During the diffusion scan, children watched a movie of their choice and were reminded to keep their heads absolutely still. Some children were scanned twice over two separate scan sessions (Fig. 1).

Imaging Parameters/Data Acquisition. Scanning was performed in a 3.0-T Siemens Tim Trio Scanner at the A.A. Martinos Imaging Center at the McGovern Institute for Brain Research at the Massachusetts Institute of Technology. Images were acquired by using one of two custom-made 32-channel phased-array head coils sized to fit younger (5–8 y) or older (9–11 y) children, or the standard Siemens head coil. The diffusion scan was acquired as part of a longer protocol that included other structural scans and functional imaging. DWI data were acquired by using standard echo-planar imaging. Slices numbered between 52 and 74, chosen to allow full-brain coverage, with echo time 84 ms, repetition time from 8.04 s to 10.39 s depending on slice number and the child's weight, and bandwidth 1,395 Hz per pixel with a generalized autocalibrating partial parallel acquisition acceleration factor of 2. The voxel size was $2 \times 2 \times 2$ mm with a 128×128 base resolution, and diffusion weighting was applied along 30 directions with a b-value of 700 s·mm^{-2}. In addition, 10 images were collected without diffusion weighting. A motion-compensated high-resolution (1 mm^3) 3D magnetization-prepared rapid acquisition with gradient echo (36) scan was also acquired on each participant. We then registered all images in a scan to the first $b0$ image (37) by using a standard function in Functional MRI of the Brain Software Library (eddy_correct) to correct distortions caused by eddy currents and misalignment caused by head motion. We reoriented each of the diffusion gradient vectors to match the rotation applied in the corresponding diffusion-weighted image (38).

Quantifying Head Motion, Data Quality Thresholds, and Matching Across Groups. We quantified data quality with four different measures. For the first two measures, we obtained frame-to-frame translation and rotation from the affine registration matrix of each frame to the first. We then averaged the frame-to-frame measures over all frames in a scan to calculate the mean translation and rotation caused by head motion for each scan. For the third and fourth measures, we computed the intensity dropout score proposed by Benner et al. (20) for each slice in each volume. This measure quantifies the attenuation of image intensities in each slice with respect to the corresponding slice in the reference ($b = 0$) volume, capturing the effect of within-slice head motion on intensity values. Slices whose score was greater than 1 were considered to have suspect signal dropout. We then calculated (*i*) the average signal dropout score for those slices with scores greater than 1 (the "Benner score") and (*ii*) the percentage of slices with a score greater than 1 across the scan ("percentage of bad slices"). Thus, our four motion measures capture global frame-to-frame motion as well as the frequency and severity of rapid slice-to-slice motion.

Next, all scans were assessed for image artifacts in the FA maps by a trained expert (A.Y.). Scans whose quality was too poor for subsequent inclusion—16 scans from children with ASD (21.3% of the data) and 13 scans (11.3% of the data) from TD children—were removed from subsequent analysis. As we had two scans from some children and were removing scans rather than excluding subjects, this left 40 children with ASD and 71 TD children in the analysis (only one scan from a single individual was used). As differences in motion between groups can greatly affect differences detected in DWI measures (11), we then excluded 28 additional TD children to match the two groups on all four motion measures (ASD, $n = 40$; TD, $n = 43$). Our general strategy was to preserve as much ASD data as possible and drop TD data until the two groups could be mean-matched on all data quality measures as well as age, sex, and IQ. One set of the analyses described in the results section was conducted on these matched groups, the main cohort.

Finally, to be sure that residual motion artifacts were not affecting our results, we conducted a second visual assessment of the raw DWI images,

removing from the data any scan with any images showing visible motion artifacts (even if the FA maps did not). This much more stringent data quality control removed an additional 40 scans from children with ASD and 52 scans from TD children, removing for this analysis 75.6% of the original sample from participants with ASD and 56.5% of the data from TD children. The remaining scans (ASD, $n = 17$; TD, $n = 42$) were then again explicitly matched on IQ, age, sex, and all four motion measures before the second analysis of the stringent cohort was conducted. Fig. 1 includes details on the demographics and matching of groups for both of these analyses.

Analysis of Diffusion Measures. Automated parcellation of the T1-weighted images were performed in FreeSurfer 5.1 (39) to identify gray and white matter volumes and to define specific cortical and subcortical regions in each individual (40, 41). Automated segmentation results were reviewed for quality and corrected by trained experts when necessary and then registered to each individual's diffusion images. These segmentations were then used as part of the diffusion analysis. Anatomically constrained probabilistic diffusion tractography was carried out by using the Tracts Constrained by UnderLying Anatomy (TRACULA) tool within FreeSurfer (42). This algorithm for automated global probabilistic tractography estimates the posterior probability of each of 18 white-matter pathways given the diffusion-weighted MRI data of each participant. This posterior probability is modeled as the product of two terms: (*i*) a data likelihood term, which uses the "ball-and-stick" model of diffusion; and (*ii*) a pathway prior term, which incorporates prior anatomical knowledge about the pathways from a set of training subjects. There is no assumption that the pathways have the same shape in the study subjects as in the training subjects, and thus TRACULA does not rely on perfect alignment between study and training subjects. The work of Yendiki et al. (42) provides more details on this method, as well as information on its accuracy in healthy participants and those with schizophrenia. Mean values for FA, MD, RD, and axial diffusivity (AD) were obtained for each of the tracts reconstructed by TRACULA. These mean values were computed by thresholding the pathway distributions at 20% of their maximum value, and FA, MD, RD, and AD values at each voxel in the tract were weighted by the pathway probability at that voxel. Analyses run with DWI measures from just the center of each tract did not change the results in any substantive way.

ACKNOWLEDGMENTS. The authors thank the study participants and their families, including those participating in the SFARI Simplex Collection and the Autism Consortium, as well as the principle investigators at SFARI SSC sites. We also thank the team at the Athinoula A. Martinos Imaging Center at McGovern Institute for Brain Research, Massachusetts Institute of Technology for their excellent technical support. This study was supported by the Ellison Medical Foundation; a grant from the Simons Foundation to the Simons Center for the Social Brain at Massachusetts Institute of Technology; National Center for Research Resources Grant U24 RR021382; National Institute for Biomedical Imaging and Bioengineering Grants 5P41EB015896-15 and R01EB006758; National Institute on Aging Grants AG022381 and 5R01AG008122-22; resources provided by Shared Instrumentation Grants 1S10RR023401, 1S10RR019307, and 1S10RR023043; and National Institutes of Health Blueprint for Neuroscience Research Grant 5U01-MH093765, part of the multiinstitutional Human Connectome Project.

1. Travers BG, et al. (2012) Diffusion tensor imaging in autism spectrum disorder: A review. *Autism Res* 5(5):289–313.
2. Hetzler BE, Griffin JL (1981) Infantile autism and the temporal lobe of the brain. *J Autism Dev Disord* 11(3):317–330.
3. Just MA, Cherkassky VL, Keller TA, Minshew NJ (2004) Cortical activation and synchronization during sentence comprehension in high-functioning autism: Evidence of underconnectivity. *Brain* 127(pt 8):1811–1821.
4. Mottron L, Dawson M, Soulières I, Hubert B, Burack J (2006) Enhanced perceptual functioning in autism: An update, and eight principles of autistic perception. *J Autism Dev Disord* 36(1):27–43.
5. Minshew NJ, Williams DL (2007) The new neurobiology of autism: Cortex, connectivity, and neuronal organization. *Arch Neurol* 64(7):945–950.
6. Herbert MR (2004) Neuroimaging in disorders of social and emotional functioning: What is the question? *J Child Neurol* 19(10):772–784.
7. Lewis JD, Elman JL (2008) Growth-related neural reorganization and the autism phenotype: A test of the hypothesis that altered brain growth leads to altered connectivity. *Dev Sci* 11(1):135–155.
8. Jones DK, Knösche TR, Turner R (2013) White matter integrity, fiber count, and other fallacies: The do's and don'ts of diffusion MRI. *Neuroimage* 73:239–254.
9. Groen WB, Buitelaar JK, van der Gaag RJ, Zwiers MP (2011) Pervasive microstructural abnormalities in autism: A DTI study. *J Psychiatry Neurosci* 36(1):32–40.
10. Shukla DK, Keehn B, Müller R-A (2011) Tract-specific analyses of diffusion tensor imaging show widespread white matter compromise in autism spectrum disorder. *J Child Psychol Psychiatry* 52(3):286–295.
11. Yendiki A, Koldewyn K, Kakunoori S, Kanwisher N, Fischl B (2013) Spurious group differences due to head motion in a diffusion tractography study. *Neuroimage*, 10.1016/j.neuroimage.2013.11.027.
12. Thomas C, et al. (2009) Reduced structural connectivity in ventral visual cortex in congenital prosopagnosia. *Nat Neurosci* 12(1):29–31.
13. Weigelt S, Koldewyn K, Kanwisher N (2012) Face identity recognition in autism spectrum disorders: A review of behavioral studies. *Neurosci Biobehav Rev* 36(3): 1060–1084.
14. Weigelt S, et al. (2014) Domain-specific development of face memory but not face perception. *Dev Sci* 17(1):47–58.
15. Ewing L, Pellicano E, Rhodes G (2013) Reevaluating the selectivity of face-processing difficulties in children and adolescents with autism. *J Exp Child Psychol* 115(2): 342–355.
16. Amaral DG, Schumann CM, Nordahl CW (2008) Neuroanatomy of autism. *Trends Neurosci* 31(3):137–145.
17. Yakovlev P, Lecours IR (1967) *Regional Development of the Brain in Early Life*, ed Minkowski A (Blackwell, Oxford).
18. Lebel C, Beaulieu C (2011) Longitudinal development of human brain wiring continues from childhood into adulthood. *J Neurosci* 31(30):10937–10947.
19. Anderson AW, Gore JC (1994) Analysis and correction of motion artifacts in diffusion weighted imaging. *Magn Reson Med* 32(3):379–387.
20. Benner T, van der Kouwe AJW, Sorensen AG (2011) Diffusion imaging with prospective motion correction and reacquisition. *Magn Reson Med* 66(1):154–167.
21. Weinstein M, et al. (2011) Abnormal white matter integrity in young children with autism. *Hum Brain Mapp* 32(4):534–543.
22. Walker L, et al. (2012) Diffusion tensor imaging in young children with autism: Biological effects and potential confounds. *Biol Psychiatry* 72(12):1043–1051.
23. Schipul SE, Keller TA, Just MA (2011) Inter-regional brain communication and its disturbance in autism. *Front Syst Neurosci* 5:10.
24. Power JD, Barnes KA, Snyder AZ, Schlaggar BL, Petersen SE (2012) Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *Neuroimage* 59(3):2142–2154.
25. Satterthwaite TD, et al. (2012) Impact of in-scanner head motion on multiple measures of functional connectivity: Relevance for studies of neurodevelopment in youth. *Neuroimage* 60(1):623–632.
26. Van Dijk KRA, Sabuncu MR, Buckner RL (2012) The influence of head motion on intrinsic functional connectivity MRI. *Neuroimage* 59(1):431–438.
27. Tyszka JM, Kennedy DP, Paul LK, Adolphs R (2013) Largely typical patterns of resting-state functional connectivity in high-functioning adults with autism. *Cereb Cortex*, 10.1093/cercor/bht040.
28. Deen B, Pelphrey K (2012) Perspective: Brain scans need a rethink. *Nature* 491(7422):S20.
29. Di Martino A, et al. (2013) The autism brain imaging data exchange: Towards a large-scale evaluation of the intrinsic brain architecture in autism. *Mol Psychiatry*, 10.1038/mp.2013.78.
30. Friston K (2012) Ten ironic rules for non-statistical reviewers. *Neuroimage* 61(4): 1300–1310.
31. Ghosh SS, et al. (2010) Evaluating the validity of volume-based and surface-based brain image registration for developmental cognitive neuroscience studies in children 4 to 11 years of age. *Neuroimage* 53(1):85–93.
32. Saygin ZM, et al. (2012) Anatomical connectivity patterns predict face selectivity in the fusiform gyrus. *Nat Neurosci* 15(2):321–327.
33. American Psychiatric Association (1994) *Diagnostic and Statistical Manual of Mental Disorders* (American Psychiatric Association, Washington, DC), 4th Ed.
34. Lord C, et al. (2000) The autism diagnostic observation schedule-generic: A standard measure of social and communication deficits associated with the spectrum of autism. *J Autism Dev Disord* 30(3):205–223.
35. Rutter M, Bailey A, Lord C (2003) *SCQ: Social Communication Questionnaire* (Western Psychological Services, Los Angeles).
36. Tisdall MD, et al. (2012) Volumetric navigators for prospective motion correction and selective reacquisition in neuroanatomical MRI. *Magn Reson Med* 68(2):389–399.
37. Jenkinson M, Bannister P, Brady M, Smith S (2002) Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage* 17(2):825–841.
38. Leemans A, Jones DK (2009) The B-matrix must be rotated when correcting for subject motion in DTI data. *Magn Reson Med* 61(6):1336–1349.
39. Fischl B (2012) FreeSurfer. *Neuroimage* 62(2):774–781.
40. Fischl B, et al. (2002) Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain. *Neuron* 33(3):341–355.
41. Fischl B, et al. (2004) Automatically parcellating the human cerebral cortex. *Cereb Cortex* 14(1):11–22.
42. Yendiki A, et al. (2011) Automated probabilistic reconstruction of white-matter pathways in health and disease using an atlas of the underlying anatomy. *Front Neuroinform* 5:23.

Supporting Information

Koldewyn et al. 10.1073/pnas.1324037111

	FA (TD Group: n = 68)					FA (ASD Group: n = 40)				
Tract	Dir.	Mean 1	Mean 2	p-value	Cohen's d	Dir.	Mean 1	Mean 2	p-value	Cohen's d
f major	↑*	.6487	.6743	.031	.5335	↑	.6231	.6584	.062	.6080
f minor	↓	.5353	.5334	.868	.0406	↓	.5292	.5225	.625	.1557
L ATR	↑*	.3688	.3882	.005	.6987	↑	.3709	.3859	.180	.4323
R ATR	↑*	.3607	.3751	.044	.4980	↑	.3664	.3780	.300	.3323
L CAB	↑	.3476	.3660	.126	.3756	↑	.3626	.3768	.322	.3176
R CAB	↑*	.3517	.3856	.005	.7061	↑	.3451	.3602	.386	.2774
L CCG	↑*	.4553	.5102	<.001	1.182	↑	.4677	.4860	.216	.3976
R CCG	↑*	.4186	.4676	<.001	1.036	↑*	.4240	.4640	.005	.9408
L CST	↑*	.5172	.5401	.046	.4928	↑*	.5135	.5392	.050	.6384
R CST	↑*	.4916	.5290	.001	.8467	↑*	.5005	.5285	.036	.6881
L ILF	↑	.5048	.5235	.060	.4637	↑	.5016	.5122	.463	.2345
R ILF	↑*	.5182	.5372	.037	.5159	↑	.5021	.5124	.421	.2570
L SLFP	↑*	.4245	.4403	.047	.4901	↑	.4350	.4511	.092	.5463
R SLFP	↑	.4165	.4295	.132	.3697	↑	.4207	.4389	.077	.5741
L SLFT	↑*	.4436	.4614	.024	.5616	↑	.4453	.4581	.117	.5074
R SLFT	↑*	.4200	.4402	.009	.6548	↑	.4172	.4356	.071	.5880
L Unc	↑*	.4019	.4282	.003	.7463	↑*	.4103	.4340	.007	.8995
R Unc	↑*	.3970	.4232	.003	.7418	↑*	.4081	.4209	.024	.4322

Fig. S1. Fractional anisotropy (FA) for each tract between younger (mean 1) and older children (mean 2) from the main cohort for typical (*Left*) and autism spectrum disorder (ASD; *Right*) groups separately. Arrows indicate the direction of the difference from younger to older children (an upward arrow indicates older children have higher FA than younger children). *P* value (uncorrected for number of tracts) and Cohen *d* refer to comparisons across age groups for each tract individually; asterisks denote significant between-group differences at $P < 0.05$, uncorrected.