

## Research report

# No evidence for systematic white matter correlates of dyslexia: An Activation Likelihood Estimation meta-analysis

David Moreau\*, Josephine E. Stonyer, Nicole S. McKay, Karen E. Waldie

Centre for Brain Research, School of Psychology, University of Auckland, New Zealand

## ARTICLE INFO

## Article history:

Received 14 October 2017

Received in revised form 11 January 2018

Accepted 12 January 2018

## Keywords:

Neurodevelopmental disorders

Dyslexia

Diffusion tensor imaging

Fractional anisotropy

Axonal white matter

Voxel-based analyses

## ABSTRACT

Dyslexia is a prevalent neurodevelopmental disorder, characterized by reading and spelling difficulties. Beyond the behavioral and functional correlates of this condition, a growing number of studies have explored structural differences between individuals with dyslexia and typically developing individuals. To date, findings remain disparate – some studies suggest differences in fractional anisotropy (FA), an indirect measure of white matter integrity, whereas others do not identify significant disparities. Here, we synthesized the existing literature on this topic by conducting a meta-analysis of Diffusion Tensor Imaging (DTI) studies investigating white matter correlates of dyslexia via voxel-based analyses (VBA) of FA. Our results showed no reliable clusters underlying differences between dyslexics and typical individuals, after correcting for multiple comparisons (false discovery rate correction). Because group comparisons might be too coarse to yield subtle differences, we further explored differences in FA as a function of reading ability, measured on a continuous scale. Consistent with our initial findings, reading ability was not associated with reliable differences in white matter integrity. These findings nuance the current view of profound, structural differences underlying reading ability and its associated disorders, and suggest that their neural correlates might be more subtle than previously thought.

© 2018 Elsevier B.V. All rights reserved.

## 1. Introduction

Developmental dyslexia is a specific type of learning disability characterized by distinct reading and spelling difficulties. The disorder is typically diagnosed in childhood, affecting around 5–7% of school aged children, and can persist into adulthood (Lindgren et al., 1985; Lyon et al., 2003; Sally E. Shaywitz et al., 2008). With heritability estimated to range between 50 and 65% (Habib and Giraud, 2013), dyslexic reading difficulties occur despite appropriate learning environment and adequate resources, and are not attributable to sensory, neurological, psychiatric, intellectual or motivational issues or deficits (Habib and Giraud, 2013; Lyon et al., 2003).

Neuroimaging studies have investigated the neurobiological underpinnings of dyslexia, yielding three key left-hemisphere networks associated with impaired reading. The posterior temporoparietal network has been mainly linked to basic, phoneme level word analysis; the posterior occipitotemporal network, including the visual word form area (VWFA), is commonly associated with word form and fluent reading (Lyon et al., 2003;

Shaywitz et al., 2002; Shaywitz et al., 1998; Shaywitz et al., 2008); whereas the anterior network of the inferior frontal gyrus, including Broca's area, is involved in speech pronunciation (Shaywitz et al., 2008). Furthermore, studies by Shaywitz et al. (1998) and Shaywitz et al. (2002) reported underactivation in posterior temporoparietal and occipitotemporal regions while reading and performing phonological tasks in dyslexics, compared to typical readers. Numerous other functional imaging studies across cultures and stages of development have supported these findings (Brunswick et al., 1999; Horwitz et al., 1998; Paulesu et al., 2001; Rumsey et al., 1992; Simos et al., 2000). Shaywitz et al. (1998) and Shaywitz et al. (2002) also observed increased activation of the inferior frontal gyrus, involved in the anterior reading network, among dyslexic compared to typical readers. This hyperactivation is hypothesized to be a compensatory strategy: dyslexic readers use memorization of the structure of words—rather than phonological skills—to read, therefore overengaging frontal brain regions (Shaywitz et al., 2007; Shaywitz et al., 2003), though we should note that these findings have been debated in the literature (Hoefl et al., 2007; Norton et al., 2015; Richlan, 2014; Richlan et al., 2009).

Beyond functional differences, Diffusion Tensor Imaging (DTI) studies have demonstrated impairments in white matter cortical

\* Corresponding author at: The University of Auckland, Centre for Brain Research, Science Centre, 23 Symonds Street, Office 227, Auckland 1010, New Zealand.

E-mail address: [d.moreau@auckland.ac.nz](mailto:d.moreau@auckland.ac.nz) (D. Moreau).

connections between regions among dyslexic readers (Vandermosten et al., 2012). DTI allows probing the distance and direction of water molecule movement, producing form and orientation information about the underlying white matter structures (Assaf and Pasternak, 2008; Soares et al., 2013). In some cortical tissues, such as gray matter and cerebrospinal fluid, diffusion is isotropic; that is, water molecules disperse approximately equally in all directions. Conversely, white matter exhibits anisotropic water movement, with water molecules showing various degrees of diffusion in each direction (Assaf and Pasternak, 2008; Emsell et al., 2015; Soares et al., 2013). In typical DTI studies, diffusion images from at least six directions are analyzed using an ellipsoid tensor model—a symmetrical 3x3 matrix. Parallel and perpendicular diffusivities are then calculated and used to estimate properties of underlying tissues. Fractional anisotropy (FA) of the tissue is used most commonly (Assaf and Pasternak, 2008; Soares et al., 2013); FA is measured from 0, isotropic diffusion, to 1, anisotropic diffusion (Assaf and Pasternak, 2008). Other properties include the mean, axial and radial diffusivities (Soares et al., 2013).

Region of interest (ROI) and voxel-based analyses (VBA) can be conducted to compare DTI properties between groups or individuals. In ROI analyses, brain regions defined by *a priori* hypotheses are manually or automatically mapped onto brain images, before the DTI properties of the ROI are averaged within a region and compared across regions. These analyses, however, can be complex, time consuming, and subject to observer and selection biases (Soares et al., 2013; Van Hecke and Emsell, 2015). In contrast, VBA use brain images normalized to a standard brain atlas and smoothed, before computing and comparing DTI properties of each individual voxel. This approach greatly reduces the typical biases of ROI analyses, although this freedom comes at a cost—as VBA is typically less theoretically driven, more drastic corrections for multiple comparisons are often required (Soares et al., 2013; Van Hecke and Emsell, 2015).

Two main avenues of research have been pursued using DTI, employing both ROI and VBA approaches. First, studies have investigated significant differences in FA between dyslexic and typical readers. Two pioneer studies, Klingberg et al. (2000) and Deutsch et al. (2005), identified significant differences in FA in the temporoparietal regions of both hemispheres among small samples of dyslexic and typical reading adults and children, respectively. Lower FA values in the left temporoparietal region among dyslexics compared to typical readers have been further corroborated in subsequent studies (Carter et al., 2009; Rimrodt et al., 2010; Steinbrink et al., 2008), yet despite this apparent convergence, the reported differences within this region vary considerably (Vandermosten et al., 2012). More problematic perhaps, Keller and Just (2009) were unable to replicate these findings in an intervention study, instead reporting lower FA in an anterior region, the left anterior centrum semiovale. Similarly, Koerte et al. (2016) found no significant differences in FA when controlling for false positives adequately. Studies have also found a variety of significant differences in other brain regions, including the superior and inferior frontal regions, precuneus, insula and occipital region in the left hemisphere, superior corona radiata, splenium of the corpus callosum and throughout the right hemisphere (Carter et al., 2009; Deutsch et al., 2005; Frye et al., 2008; Niogi and McCandliss, 2006; Rimrodt et al., 2010; Steinbrink et al., 2008). In addition, Richards et al. (2008) found 45 clusters of significant FA differences between dyslexics and typical readers across the whole brain. Taken together, these findings highlight the wide discrepancies reported in the literature.

Besides group differences contrasting dyslexics with typical readers, additional studies have identified regions where FA values significantly correlate with performance on reading tasks. Numerous studies report positive correlations between FA in the left tem-

poroparietal area of dyslexic or typical readers and reading ability, measured by a range of reading measures (e.g., word reading, pseudo word reading or phonological reading tasks; Beaulieu et al., 2005; Deutsch et al., 2005; Klingberg et al., 2000; Lebel et al., 2013; Nagy et al., 2004; Odegard et al., 2009; Steinbrink et al., 2008). Similar to the aforementioned literature on group contrasts, however, specific locations within these regions differ considerably between studies (Vandermosten et al., 2012). For example, positive correlations between reading ability and FA have been noted in the superior corona radiata, longitudinal fasciculi, external capsule, centrum semiovale and language areas of the left hemisphere, and bilateral inferior and temporofrontal regions, illustrating the wide variability in results (Deutsch et al., 2005; Keller and Just, 2009; Niogi and McCandliss, 2006; Rimrodt et al., 2010; Steinbrink et al., 2008; Zhang et al., 2014). Finally, negative correlations between reading ability and FA in the posterior/temporal corpus callosum have also been reported (Dougherty et al., 2007; Frye et al., 2008; Odegard et al., 2009; Zhang et al., 2014).

These discrepancies highlight the need to comprehensively examine the variability in brain regions linked to dyslexia. Using Activation Likelihood Estimation (ALE), a technique that determines convergence of activation probabilities across studies (Eickhoff et al., 2009; Eickhoff et al., 2012), Vandermosten et al. (2012) performed a meta-analysis and found that a large cluster (704 mm<sup>3</sup>) centered at  $-29, -17, 26$  near the left temporoparietal region, across three DTI studies of correlative and difference that employed VBA. A smaller cluster near the inferior frontal gyrus, centered at  $-26, 26, 18$  was also identified, although less reliably. However, this ALE meta-analysis only examined coordinates where significant differences between dyslexic and typical readers were identified, with particular combinations of studies that created difficulties in interpretation. The software the authors used, GingerALE, has also been updated since, including to correct problems that had a to increase the rate of false positives (Eickhoff et al., 2017). Lastly, further correlational research has been conducted since initial publication of the meta-analysis in 2012, suggesting a possible gap in the current literature, and a need to systematically summarize and quantify the relationship between developmental dyslexia and white matter connections. To address these limitations, we conducted a meta-analysis that consisted of two phases. In Phase 1, we focused on differences in FA, assessed via VBA, between dyslexic and typical readers. Phase 2 of the meta-analysis was restricted to correlations between reading ability and VBA studies of FA.

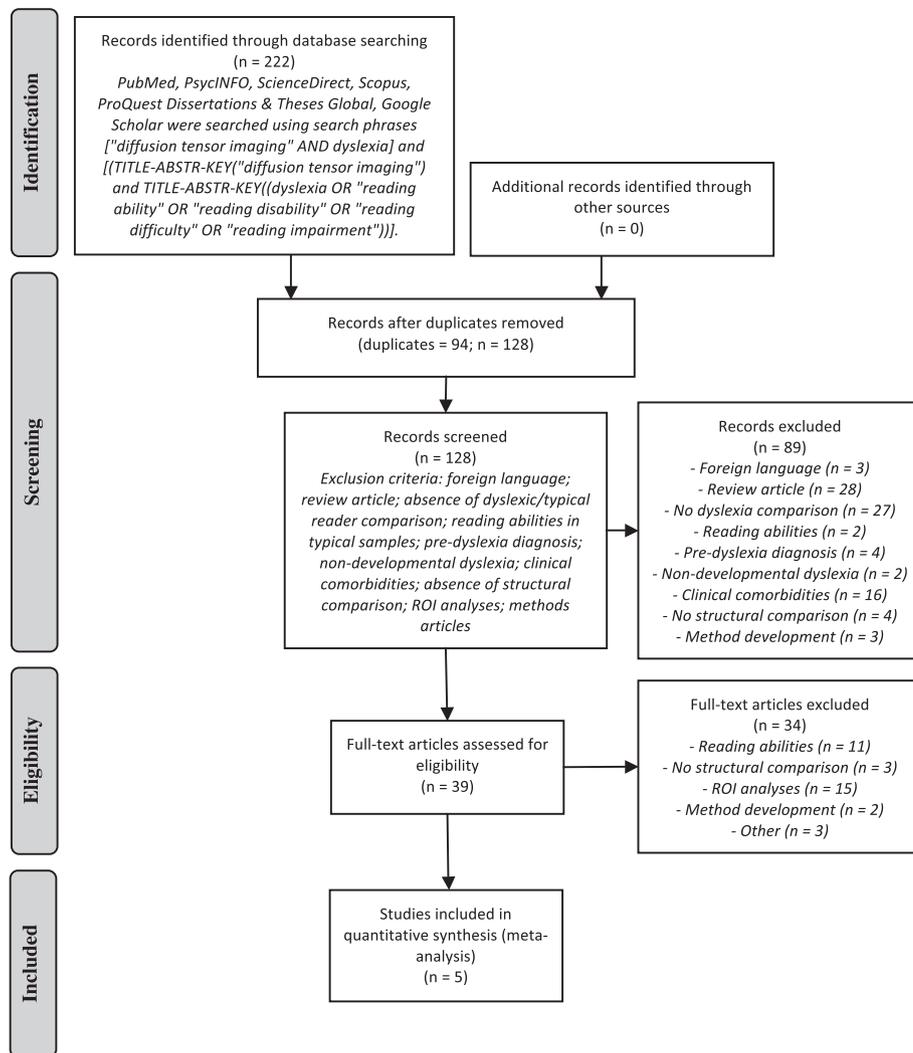
## 2. Results

### 2.1. Study selection and characteristics

All details regarding study selection are outlined in Fig. 1 (Phase 1) and Fig. 2 (Phase 2). Table 1 details the characteristics and demographics of participants included, and the findings of group differences in FA for each study (Phase 1). Table 2 reports the same information for correlations between FA and reading ability (Phase 2).

### 2.2. Synthesis of results

Two analyses were run in Phase 1. The analysis of 47 foci from 5 experiments (99 subjects), where FA was significantly greater in typical compared to dyslexic readers, yielded no significant clusters when using a FDR correction of 0.05. Similarly, the analysis of 17 foci from 2 experiments (52 subjects), where FA was significantly greater in dyslexic compared to typical readers, produced no significant clusters when using a FDR correction of 0.05.



**Fig. 1.** White matter tracts found to have greater fractional anisotropy in controls (A) and in dyslexics (B) across the studies included in Phase 1, uncorrected for multiple comparisons. These tracts are displayed on the skull stripped MNI152 standard brain template at 1 mm resolution using FSLview 4.0.1.

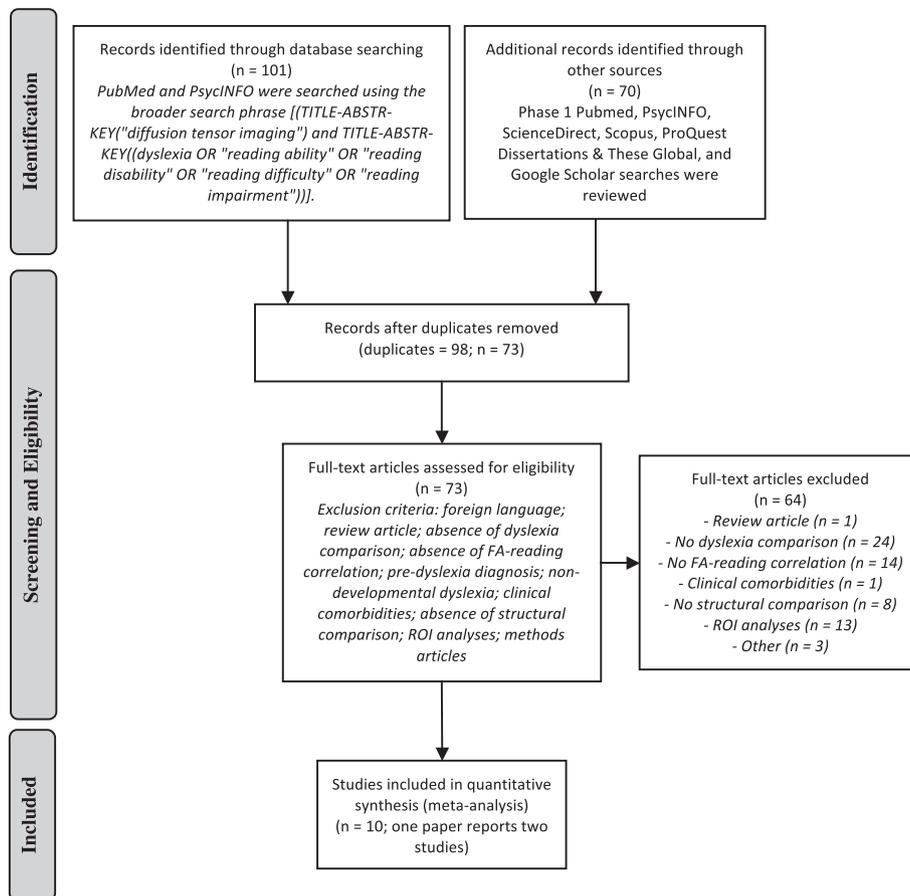
In Phase 2, two analyses were also undertaken. The analysis of 42 foci from 9 experiments (500 subjects), where reading ability was significantly positively correlated with FA generated no significant clusters when using a FDR correction of 0.05. Similarly, the analysis of 2 foci from 2 experiments (40 subjects) where reading ability was significantly negatively correlated with FA found no significant clusters when using a FDR correction of 0.05.

All analyses (Phases 1 and 2) were also run without correcting for multiple comparisons using an uncorrected  $p$ -value threshold of 0.05. Tables 3 and 4 present the coordinates of the clusters and their respective contributors for Phase 1 and Phase 2, respectively. Figures 3 and 4 present the white matter tracts corresponding to the uncorrected clusters, for each phase. These include clusters extracted from contrasting dyslexics and control individuals (Phase 1), and those extracted from correlating FA and reading ability (Phase 2). Probabilistic values were computed for each cluster, to determine the most likely white matter tracts associated with each cluster. For details about the clusters, foci and probabilistic estimates, see the Supplemental Material and online analyses outputs.

### 2.3. Additional analysis

One additional analysis was conducted in Phase 1 where studies including adults and children were analyzed separately. No significant clusters were produced when typical readers had significantly greater FA than dyslexic readers or when dyslexic readers had significantly greater FA than typical readers among studies of adults or children, using an FDR correction of 0.05.

In Phase 2, several additional analyses were undertaken. Firstly, studies including adult and child participants were again analyzed separately. No significant clusters were produced when reading ability was negatively correlated with FA among adult or child studies, or when reading ability was positively correlated with FA among adult studies, using an FDR correction of 0.05. One significant cluster—256 mm<sup>3</sup>, from MNI coordinates (−22, −10, 32) to (−16, −6, 38), centered at (−18.9, −8.34, 34.9)—was found when analyzing children studies reporting coordinates of positive correlations between reading and FA, but no significant cluster was found when results from Zhang et al. (2014) were excluded from analyses.



**Fig. 2.** White matter tracts associated with positive (A) and negative (B) correlations between fractional anisotropy and reading ability, uncorrected for multiple comparisons. These tracts are displayed on the skull stripped MNI152 standard brain template at 1 mm resolution using FSLview 4.0.1.

### 3. Discussion

We conducted a meta-analysis with two distinct but complementary phases—the first one sought to identify differences in FA between dyslexic and typical readers, whereas the aim of the second phase was to explore correlations between FA and reading ability. Our results showed no systematic differences in fractional anisotropy between dyslexic and typical readers, or as a function of reading ability, after correcting for multiple comparisons. As we argue henceforth, this finding may reflect the current lack of reliable and consistent structural differences in white matter tracts associated with reading ability, or reading disability.

We initially selected and analyzed DTI studies that used VBA to identify cortical coordinates where significant differences in FA existed between dyslexic and typical readers (Phase 1). All studies identified coordinates where typical readers showed greater FA than dyslexic readers, and two studies also reported coordinates where dyslexic readers had greater FA than typical readers. These coordinates were analyzed separately; in both analyses, no reliable clusters were found after correcting for multiple comparisons. That is, no reliable differences in white matter integrity, as measured by FA, were found between dyslexic and typical readers.

Because group comparisons may be too coarse to detect subtle differences, and given that several studies also analyzed the relationship between FA and reading ability on a continuum, we then focused on correlations (Phase 2). Specifically, we selected DTI studies that used VBA to locate cortical coordinates where FA significantly correlated with reading ability or performance on a reading-based task. From this search, nine studies reported coordi-

nates of significant positive FA-reading correlations and two studies reported significant negative correlations. Consistent with Phase 1 findings, no reliable clusters were detected in positive or negative correlation analyses after correcting for multiple comparisons. In other words, reading ability was not reliably associated with white matter integrity as measured by FA.

Our findings nuance the current view of profound differences in FA underlying reading ability and associated disorders such as dyslexia, and correlative relationships between reading ability and FA. We found no reliable differences in FA between dyslexic and typical readers, a finding that contrasts with early findings by Klingberg et al. (2000) and Deutsch et al. (2005) of a temporoparietal difference in FA between dyslexic and typical readers (see also Carter et al., 2009; Rimrodt et al., 2010; Steinbrink et al., 2008). Similarly, the null findings in Phase 2 of this meta-analysis do not support the idea of a relationship between reading ability and FA in the left temporoparietal region, and negative correlation between reading ability and FA in the corpus callosum (Beaulieu et al., 2005; Deutsch et al., 2005; Dougherty et al., 2007; Frye et al., 2008; Klingberg et al., 2000; Nagy et al., 2004).

In line with our findings, however, Vandermosten et al. (2012) noted that despite appearing consistent, each one of the studies they included in their meta-analysis produced coordinates at different locations within the temporoparietal region and corpus callosum; for example, some studies reported correlations with the temporal corpus callosum and others with the posterior corpus callosum. In fact, many studies have also reported differences and correlations in a range of other regions distributed widely throughout the cortex (Deutsch et al., 2005; Frye et al., 2008;

**Table 1**  
Studies included in Phase 1 of the meta-analysis.

Authors and publication year	No. of participants (Gender split)	Participants' age in years: mean and/or range	Coordinates of group differences in FA where controls > dyslexics: MNI, mm <sup>3</sup>			Coordinates of group differences in FA where dyslexics > controls: MNI, mm <sup>3</sup>		
Rimrodt et al. (2010)	31 participants: • 14 dyslexics (7M, 7F) • 17 typically developing controls (9M, 8F)	• Overall range: 7–16 • Dyslexics: mean = 11.0 • Controls: mean = 11.0	–54	4	30	–44	0	–10
			–46	–32	18	50	–16	–28
			58	6	34	34	–32	52
						46	–70	–16
			18	20	–8			
			2	–4	–22			
			58	–48	10			
Steinbrink et al. (2008)	16 participants: • 8 dyslexics (6M, 2F) • 8 healthy controls (6M, 2F)	• Dyslexics: mean = 20.1 • Healthy controls: mean = 23.7	–21	27	6			
			–30	0	9			
			–33	–7	6			
			–24	–87	15			
			30	0	15			
Richards et al. (2008)	21 participants: • 14 dyslexics (14 M) • 7 healthy normal readers (7M)	• Overall: range = 30–45	–35	–7	38	–44	8	33
			52	0	21	–39	26	–14
			19	55	17	37	–41	–24
			36	36	17	–47	–23	45
			55	19	0	–54	2	–32
			38	43	1	61	–11	–16
			–47	38	–12	–45	–21	–33
			51	36	–8	41	–60	–39
			14	3	61	–42	–63	–48
			9	–94	12	–33	–52	–42
			–10	–89	–17			
			–20	–90	18			
			26	–76	26			
			42	–77	18			
			–20	–84	–19			
			41	–11	–32			
			–52	–19	30			
			53	–1	22			
			–14	–72	47			
			–30	–43	45			
			47	–47	40			
			–10	–72	48			
			–17	11	2			
–48	–33	8						
–23	15	–31						
–47	–34	8						
–60	–54	–19						
41	–9	–33						
–14	–1	36						
–17	11	2						
–16	–4	40						
–18	–47	33						
–46	–28	–12						
–37	–23	32						
–21	9	20						
Deutsch et al. (2005)	14 participants: • 7 poor readers with previous dyslexia diagnosis (6M, 1F) • 7 normal readers (6M, 1F)	• Overall: range = 7–13 • 7 Poor readers with previous dyslexia diagnosis: mean = 11.0 • Normal readers: mean = 10.4	–28.3	–25.14	23.88			
			30.96	–44.26	30.39			
			–22.15	24.29	20.46			
Klingberg et al. (2000)	17 participants: • 6 poor readers with history of dyslexia (5M, 1F) • 11 controls (6M, 5F)	• Poor readers with history of dyslexia: mean = 31.5 • Controls: mean = 23.1	–31.47	–29.84	16.45			

Niogi and McCandliss, 2006; Richards et al., 2008; Rimrodt et al., 2010; Steinbrink et al., 2008). Our null results reflect these within-region differences—when aggregated, the inconsistencies across published findings led to disparate clusters that could have emerged stochastically, given the number of comparisons performed in these analyses. We should also point out that the way we conducted the present meta-analysis, looking at aggregates but also at subsets via group splits (e.g., analyzing separately positive and negative correlations in Phase 2) is bound to inflate the false positive rate. This was deliberate on our part, because we

intended to show that *despite* these multiple, uncorrected comparisons, we still did not find evidence for reliable effects. Overall, our findings challenge the notion that dyslexic and typical readers show differences in FA that are sufficiently systematic to be ascribed to dyslexia (Phase 1), and that these differences reflect various degrees of reading performance (Phase 2).

Interestingly, Vandermosten et al. (2012) identified a significant left temporoparietal cluster in their meta-analysis, which included nine DTI studies using VBA to identify locations where reading ability correlated with FA. Phase 2 of this meta-analysis employed

**Table 2**  
Studies included in Phase 2 of the meta-analysis.

Authors and publication year	No. of participants	Participants' age in years: mean and/or range	Behavioral measure correlated with FA	Coordinates of positive correlations between behavioral measures and FA: MNI, mm3	Coordinates of negative correlations between behavioral measures and FA: MNI, mm3
Lebel et al. (2013)	136 participants (73 M, 63F)	<ul style="list-style-type: none"> <li>Overall: mean = 20.1, range = 15–28</li> </ul>	<ul style="list-style-type: none"> <li>Gray Oral Reading Test (GORT): fluency, speed and accuracy of reading text passages</li> <li>Woodcock–Johnson Word Identification test: untimed measure of word reading</li> <li>Woodcock–Johnson Word Attack test: untimed measure of pseudoword reading or decoding</li> </ul>	-10 -14 10 -20 42 10 18 38 6 26 18 13 32 -19 -11 -20 28 -6 5 -38 9 30 -18 40 35 -42 28	
Odegard et al. (2009)	17 participants: <ul style="list-style-type: none"> <li>10 dyslexics (7M, 3F)</li> <li>7 non-impaired readers (2M, 5F)</li> </ul>	<ul style="list-style-type: none"> <li>Overall: range = 10–14</li> <li>Dyslexics: mean = 12.20</li> <li>Non-impaired readers: mean = 12.32</li> </ul>	Decoding Skills Test: phonological decoding Wechsler Individual Achievement Test Second Edition basic reading test: word reading and recognition	-14 -9 46 -20 -8 36	-8 -35 22
Beaulieu et al. (2005)	32 participants (14M, 18F)	<ul style="list-style-type: none"> <li>Overall: mean = 11.1, range = 8.3–12.9</li> </ul>	Woodcock Reading Mastery Word Identification test: reading abilities	-28 -14 24 26 -14 24 -24 24 18 -22 -18 36 -18 -8 34	
Zhang et al. (2014)	93 American readers (39M, 54F)	<ul style="list-style-type: none"> <li>Overall: mean = 20.8, range 18–30</li> </ul>	<ul style="list-style-type: none"> <li>Test of Word Reading Efficiency Sight Word Efficiency (Form A) subtest (TOWRE-SWE): word reading accuracy and fluency</li> </ul>	-3 -24 -22 36 -5 -22 13 22 20 -11 24 18 9 30 6	
	89 Chinese readers (44M, 45F)	<ul style="list-style-type: none"> <li>Overall: mean = 21.7, range = 19–25</li> </ul>	<ul style="list-style-type: none"> <li>Chinese Character Reading Efficiency test: format of TOWRE-SWE, word reading accuracy and fluency in Chinese</li> </ul>	-23 -53 9 -40 -8 -23 -9 -3 -7 -17 -23 -6 -10 24 15 -14 -23 -3 20 -25 49 -12 -26 -5 27 -42 25 21 -29 48 -8 -21 1	
Nagy et al. (2004)	23 participants (14M, 9F)	<ul style="list-style-type: none"> <li>Overall: mean = 11.9, range = 7.8–18.5</li> </ul>	Unstandardized reading task: timed speed and accuracy reading task		-27 -28.4 5.22
Rimrodt et al. (2010)	31 participants <ul style="list-style-type: none"> <li>14 dyslexics (7M, 7F)</li> <li>17 typically developing controls (9M, 8F)</li> </ul>	<ul style="list-style-type: none"> <li>Overall range: 7–16</li> <li>Dyslexics: mean = 11.0</li> <li>Controls: mean = 11.0</li> </ul>	<ul style="list-style-type: none"> <li>Test of Word Reading Efficiency -composite of Sight Word Efficiency and Phonemic Decoding Efficiency subtests: word identification, decoding pseudowords, and speech of accurate articulation</li> </ul>	-54 4 30 54 26 20 60 8 10 -4 -94 -20 -6 72 58 -32 42 -20 -54 0 -22	
Deutsch et al. (2005)	14 participants: <ul style="list-style-type: none"> <li>7 poor readers with previous dyslexia diagnosis (6M, 1F)</li> <li>7 normal readers (6M, 1F)</li> </ul>	<ul style="list-style-type: none"> <li>Overall: range = 7–13</li> <li>Poor readers with previous dyslexia diagnosis: mean = 11.0</li> <li>Normal readers: mean = 10.4</li> </ul>	<ul style="list-style-type: none"> <li>Woodcock- Johnson Letter-Word Identification (WJ-LWID) and Word Attack (WJ-WA) subtests: word and pseudoword reading accuracy</li> <li>Woodcock- Johnson Spelling subtest: spelling</li> <li>Comprehensive Test of Phonological Processing (composite score): phonological awareness</li> <li>Woodcock- Johnson Basic Reading Composite test: average of the WJ-LWID WJ-WA subtests</li> </ul>	-28 -26 28	
Klingberg et al. (2000)	17 participants: <ul style="list-style-type: none"> <li>6 poor readers with history of dyslexia (5M, 1F)</li> <li>11 controls (6M, 5F)</li> </ul>	<ul style="list-style-type: none"> <li>Poor readers with history of dyslexia: mean = 31.5</li> <li>Controls: mean = 23.1</li> </ul>	<ul style="list-style-type: none"> <li>Woodcock Word ID test: word reading/decoding accuracy</li> <li>Woodcock Word attack test: pseudoword decoding</li> </ul>	-28 -18 28	
Keller and Just (2009)	72 participants: <ul style="list-style-type: none"> <li>47 poor readers</li> <li>25 good reading controls</li> </ul>	<ul style="list-style-type: none"> <li>Overall: range = 8–12</li> </ul>	<ul style="list-style-type: none"> <li>Woodcock Reading Mastery Word Attack test: word reading/decoding accuracy</li> </ul>	-16 12 34	

similar methods, yet results are ambiguous. It is worth pointing out that Vandermosten et al. (2012) used GingerALE 2.0.4, which

has since been updated to correct major flaws affecting analyses—initial errors made ALE analyses too lenient, therefore

**Table 3**  
Clusters volume and coordinates for contrasts between Controls and Dyslexics (Phase 1), with greater FA values for the former (Cont > Dys) and the latter (Dys > Cont). All coordinates are reported in standard MNI space.

Contrast	Cluster	Volume (mm <sup>3</sup> )	Weighted Center			Extrema			
			x	y	z	Value	x	y	z
Cont > Dys	1	17,592	-39.7	-24.7	23.2	0.009513951	-46	-32	18
Cont > Dys	1					0.008486417	-46	-34	8
Cont > Dys	1					0.008287956	-32	-30	16
Cont > Dys	1					0.008234489	-52	-20	30
Cont > Dys	1					0.008133247	-36	-24	32
Cont > Dys	1					0.007989156	-28	-26	24
Cont > Dys	1					0.007969191	-35.3	-7.3	38
Cont > Dys	2	15,232	-23.6	11.5	10.8	0.007985766	-22	10	20
Cont > Dys	2					0.007972456	-18	12	2
Cont > Dys	2					0.00765096	-30	0	10
Cont > Dys	2					0.00750154	-22	24	20
Cont > Dys	2					0.007428612	-22	26	6
Cont > Dys	2					0.007414908	-32	-6	6
Cont > Dys	3	7648	41.7	38.3	3.3	0.008234363	43.5	36	4.5
Cont > Dys	3					0.00796919	38	43	1
Cont > Dys	4	5312	54.9	2.4	27	0.009241275	58	6	34
Cont > Dys	4					0.008239598	52	0	22
Cont > Dys	5	5264	39.1	-45.6	35.3	0.007969638	46	-46	40
Cont > Dys	5					0.007299556	30	-44	30
Cont > Dys	6	5056	-24	-45	39	0.00796919	-24	-45	39
Cont > Dys	7	4712	-15	-86.5	-18	0.008234363	-20	-84	-19
Cont > Dys	7					0.00796919	-10	-89	-17
Cont > Dys	8	4312	-22	-88.5	16.5	0.012253704	-22	-88	16
Cont > Dys	9	3656	-15	-2.5	38	0.008508359	-16	-4	40
Cont > Dys	9					0.008234363	-14	-1	36
Cont > Dys	10	3408	-12	-72	47.5	0.008508359	-10	-72	48
Cont > Dys	10					0.008234363	-14	-72	47
Cont > Dys	11	2976	41	-10	-32.5	0.00796919	41	-11	-32
Cont > Dys	12	2568	30	0	15	0.007649369	30	0	16
Cont > Dys	13	2560	-60	-54	-19	0.008234363	-60	-54	-19
Cont > Dys	14	2560	-47	38	-12	0.008234363	-47	38	-12
Cont > Dys	15	2552	42	-77	18	0.008234363	42	-77	18
Cont > Dys	16	2528	55	19	0	0.00796919	55	19	0
Cont > Dys	17	2528	14	3	61	0.00796919	14	3	61
Cont > Dys	18	2440	-46	-28	-12	0.008508359	-46	-28	-12
Cont > Dys	19	2440	26	-76	26	0.008508359	26	-76	26
Cont > Dys	20	2440	-54	4	30	0.009234563	-54	4	30
Cont > Dys	21	2432	19	55	17	0.007712557	19	55	17
Cont > Dys	22	2416	9.1	-93.8	11.9	0.008234363	9	-94	12
Cont > Dys	23	2304	-23.2	15	-30.9	0.007712557	-23	15	-31
Dys > Cont	1	12,432	-49	1	-21	0.009234563	-44	0	-10
Dys > Cont	1					0.008508359	-54	2	-32
Dys > Cont	2	12,248	55.5	-13.5	-22	0.0092352	50	-16	-28
Dys > Cont	2					0.007970334	60	-12	-16
Dys > Cont	3	12,128	-37.5	-57.5	-45	0.008234363	-37.5	-57.5	-45
Dys > Cont	4	6720	41	-60	-39	0.00796919	41	-60	-39
Dys > Cont	5	6720	37	-41	-24	0.00796919	37	-41	-24
Dys > Cont	6	6720	-47	-23	45	0.007712557	-47	-23	45
Dys > Cont	7	6560	-39	26	-14	0.008234363	-39	26	-14
Dys > Cont	8	6560	-44	8	33	0.008234363	-44	8	33
Dys > Cont	9	6008	18	20	-8	0.009234563	18	20	-8
Dys > Cont	10	6008	34	-32	52	0.009234563	34	-32	52
Dys > Cont	11	5960	58	-48	10	0.009234563	58	-48	10
Dys > Cont	12	5632	45.9	-69.9	-15.9	0.009234563	46	-70	-16
Dys > Cont	13	5408	-44.9	-20.5	-32.2	0.007712557	-45	-21	-33
Dys > Cont	14	4616	2.3	-4.8	-21.1	0.009234563	2	-4	-22

inadequately controlling for spurious findings (Turkeltaub et al., 2012). In this study, we used the latest update, GingerALE 2.3.6. The importance of adequate correction for multiple comparisons is striking when visually examining Figs. 3 and 4 – virtually all white matter tracts are associated either with comparisons between dyslexics and controls, or as a function of reading ability. This illustrates the heterogeneity of results in the literature, which undoubtedly underlies the absence of robust effect after adequate corrections. More generally, the methods implemented in GingerALE are bound to underestimate inconsistencies between studies (Eickhoff et al., 2009), which indicates that caution should be used when interpreting positive results. In any case, this provides addi-

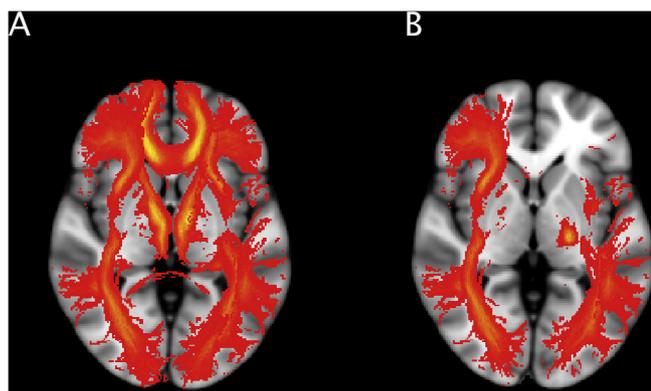
tional strength to the results we report herein – our overall analyses yielded null results despite potential biases toward positive results.

There were other differences in statistical analyses between the present meta-analysis and the study by Vandermosten et al. (2012)—these authors did not analyze studies reporting significant differences in FA between dyslexic and typical readers; rather, coordinates of significant differences were integrated with coordinates of significant correlations between reading and FA. Although the intent was to increase statistical power, combining disparate studies can considerably blur the interpretation of the findings. This reluctance to aggregate studies that are fundamentally

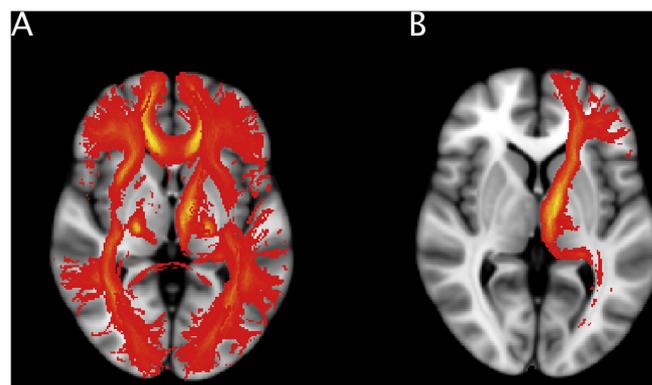
**Table 4**

Clusters volume and coordinates for positive and negative correlations between FA and reading ability (Phase 2). All coordinates are reported in standard MNI space.

Correlation	Cluster	Volume (mm <sup>3</sup> )	Weighted Center			Extrema			
			x	y	z	Value	x	y	z
Positive	1	17,696	-21.9	-11.2	33	0.015478318	-18	-8	34
Positive	1					0.01356399	-28	-16	26
Positive	1					0.010248792	-16	12	34
Positive	1					0.009979409	-22	-18	36
Positive	1					0.00782307	-14	-8	46
Positive	2	10,616	16.5	27	11.2	0.010700151	18	38	6
Positive	2					0.010260232	26	18	14
Positive	2					0.010130377	10	30	6
Positive	2					0.010052992	14	22	20
Positive	2					0.010657542	30	-18	40
Positive	3	9624	24.8	-20.7	39.1	0.009664648	20	-24	48
Positive	3					0.009661209	22	-28	48
Positive	3					0.009285999	26	-14	24
Positive	3					0.010723425	-10	-14	10
Positive	3					0.010027994	-12	-26	-6
Positive	4	8632	-11.6	-20.5	0.7	0.009861533	-8	-20	2
Positive	4					0.009661506	-14	-22	-2
Positive	4					0.018599521	-10	24	16
Positive	4					0.009308879	-24	24	18
Positive	4					0.010029106	-40	-8	-22
Positive	5	6768	-14.8	24	17	0.009235151	-54	0	-22
Positive	5					0.01004936	36	-6	-22
Positive	5					0.009875091	32	-18	-12
Positive	5					0.011431893	34	-42	28
Positive	5					0.011065371	28	-42	26
Positive	6	5720	-47	-4	-22.5	0.009234563	60	8	10
Positive	6					0.009234563	54	26	20
Positive	6					0.009234563	-54	4	30
Positive	6					0.009874855	5	-38	9
Positive	6					0.009234563	-32	42	-20
Positive	7	5600	34	-12.1	-16.4	0.010655258	-20	42	10
Positive	7					0.01065525	-20	42	10
Positive	7					0.01065525	-20	28	-6
Positive	7					0.010049283	-4	-24	-22
Positive	7					0.009304394	-10	-4	-6
Positive	8	4840	31	-42	26.5	0.00930438	-23	-53	9
Positive	8					0.009234563	-4	-94	-20
Positive	8					0.008335105	-28	-28	6
Positive	8					0.00930438	-23	-53	9
Positive	8					0.009234563	-4	-94	-20
Positive	9	2920	60	8	10	0.009234563	54	26	20
Positive	9					0.009234563	54	26	20
Positive	9					0.009234563	-54	4	30
Positive	9					0.009874855	5	-38	9
Positive	9					0.009234563	-32	42	-20
Positive	10	2920	54	26	20	0.010655258	-20	42	10
Positive	10					0.010655258	-20	42	10
Positive	10					0.01065525	-20	28	-6
Positive	10					0.010049283	-4	-24	-22
Positive	10					0.009304394	-10	-4	-6
Positive	11	2920	-54	4	30	0.00930438	-23	-53	9
Positive	11					0.009234563	-4	-94	-20
Positive	11					0.008335105	-28	-28	6
Positive	11					0.00930438	-23	-53	9
Positive	11					0.009234563	-4	-94	-20
Positive	12	2848	5	-38	9	0.009234563	54	26	20
Positive	12					0.009234563	54	26	20
Positive	12					0.009234563	-54	4	30
Positive	12					0.009874855	5	-38	9
Positive	12					0.009234563	-32	42	-20
Positive	13	2800	-32	42	-19.9	0.010655258	-20	42	10
Positive	13					0.010655258	-20	42	10
Positive	13					0.01065525	-20	28	-6
Positive	13					0.010049283	-4	-24	-22
Positive	13					0.009304394	-10	-4	-6
Positive	14	2736	-20	42	10	0.00930438	-23	-53	9
Positive	14					0.009234563	-4	-94	-20
Positive	14					0.008335105	-28	-28	6
Positive	14					0.00930438	-23	-53	9
Positive	14					0.009234563	-4	-94	-20
Positive	15	2728	-20	28	-6	0.009234563	54	26	20
Positive	15					0.009234563	54	26	20
Positive	15					0.009234563	-54	4	30
Positive	15					0.009874855	5	-38	9
Positive	15					0.009234563	-32	42	-20
Positive	16	2704	-3	-24	-22	0.010655258	-20	42	10
Positive	16					0.010655258	-20	42	10
Positive	16					0.01065525	-20	28	-6
Positive	16					0.010049283	-4	-24	-22
Positive	16					0.009304394	-10	-4	-6
Positive	17	2464	-9	-3	-7	0.00930438	-23	-53	9
Positive	17					0.009234563	-4	-94	-20
Positive	17					0.008335105	-28	-28	6
Positive	17					0.00930438	-23	-53	9
Positive	17					0.009234563	-4	-94	-20
Positive	18	2432	-23	-53	9	0.009234563	54	26	20
Positive	18					0.009234563	54	26	20
Positive	18					0.009234563	-54	4	30
Positive	18					0.009874855	5	-38	9
Positive	18					0.009234563	-32	42	-20
Positive	19	1792	-5.2	-93.3	-18.7	0.010655258	-20	42	10
Positive	19					0.010655258	-20	42	10
Positive	19					0.01065525	-20	28	-6
Positive	19					0.010049283	-4	-24	-22
Positive	19					0.009304394	-10	-4	-6
Negative	1	28,928	-17.6	-31.7	13.6	0.008335105	-28	-28	6
Negative	1					0.007787312	-8	-35	22



**Fig. 3.** Phase 1 flow diagram.



**Fig. 4.** Phase 2 flow diagram.

different was our rationale for distinct analyses. Another point of difference with the [Vandermosten et al. \(2012\)](#) meta-analysis was in our inclusion of two additional studies published after 2012, which in total contained three groups of significant coordinates. This last point in itself does not explain the failed replication, however, given that leaving these studies out still did not allow us to exactly reproduce the findings by [Vandermosten et al. \(2012\)](#).

Our findings are also consistent with a recent study by [Koerte et al. \(2016\)](#), which failed to detect significant differences in FA between dyslexic and typical readers, after correcting for multiple

comparisons. The authors proposed that the strict multiple comparison correction employed in their analyses contributed to null results; however, the procedure they used is the recommended one given the high rate of false positives when failing to correct for multiple comparisons ([Laird et al., 2005](#); [Moreau et al., 2016](#); [Richlan, 2014](#); [Richlan et al., 2009](#); [Turkeltaub et al., 2012](#)). Together with the present findings, this suggests that differences in other VBA studies may be based on unique characteristics of each individual sample (e.g., demographics, study designs), or false positives, and does not support the idea of systematic differences in FA between dyslexic and typical readers. At the very least, the

present findings call for caution when interpreting significant voxels or clusters in a given comparison or correlational analysis, and emphasize the need for statistically sound procedures in future work.

There are several limitations to the present findings. First, we focused on VBA studies, given the consequent body of literature that has shown differences between dyslexics and typical readers based on this method. However, it is important to note that recent studies have used ROI-based approaches, and that these were not included in our analyses. ROI-based approaches allow analyses that can be more theory-driven than whole-brain analyses, yet they are still prone to many issues observed in the VBA literature on dyslexia. For example, the inherent noise of both the data acquisition process and the DTI model (e.g., Zhao et al., 2016), together with flexible pipelines of analysis (Carp, 2012) and failures to correct adequately for multiple comparisons, are all components of the developmental dyslexia literature that potentially exacerbate differences, or could even be responsible for a multitude of spurious findings (see for a recent review Ramus et al., 2018). Our focus on VBA studies was deliberate, both for internal consistency and because this type of analysis, applied to developmental dyslexia, has yielded a large number of findings with very little published negative results (Ramus et al., 2018). However, we do not exclude the possibility that more fine-grained methods and analyses could further document structural differences between dyslexics and typical readers (Zhao et al., 2016), although more precise and valid measurements have also been shown to provide additional evidence for a lack of structural differences between groups (Vanderauwera et al., 2015).

In addition, we analyzed together studies that included children and those focusing on adult populations in both phases of this meta-analysis. This was meant to allow generalizations to a disorder, dyslexia, rather than to the same disorder *given specific developmental characteristics*. Averaging these populations together, however, may have reduced the sensitivity of our analyses. For example, age-related reading differences could influence outcomes in children, whereas cortical differences in adult populations could reflect the poorer reading exposure, both quantitative and qualitative, commonly experienced by dyslexic readers throughout their lives (Ben-Shachar et al., 2007). Aware of this potential limitation, we ran additional analyses separating adults and children to verify that age-related factors did not influence results (see online material for details). In Phase 1, no significant clusters were identified in these separate analyses of children or adults, corroborating initial conclusions. In Phase 2, no significant clusters were found upon analysis of adult studies; however, analysis of positive correlations between reading and FA in five studies including children identified one significant cluster located at MNI coordinates (−18, −8, 34). This suggests a possible positive correlation between reading, as measured by performance on reading-related tasks, and white matter integrity, as measured by FA. It is difficult to confidently relate this finding with dyslexia *per se*, however, given that in three of these studies, participants included a range of typical readers with diverse reading abilities, and the “poor readers” were not explicitly diagnosed with dyslexia (Beaulieu et al., 2005; Nagy et al., 2004; and Keller and Just, 2009).

It is also possible that specific characteristics of the selected studies disproportionately influenced the results. For example, Zhang et al. (2014) reported coordinates where significant correlations between FA and performance on a Chinese Character Reading Efficiency Test were found among a Chinese readers, yet research has demonstrated that when reading Chinese, a character-based language system, different cortical processing are employed compared to reading English, an alphabetic language system (Siok et al., 2008). To ensure that our conclusions were not excessively influenced by the study by Zhang et al. (2014), we re-ran our anal-

ysis excluding it. No significant differences were detected in this additional analysis, substantiating our initial conclusions. On a similar note, Nagy et al. (2004) employed unstandardized tests to quantify participants' reading ability. These may provide less reliable measures of reading ability compared to standardized tests used in other studies included in this meta-analysis. Excluding this study in an additional analysis to determine if results were affected was not possible as the significant coordinate found by Nagy et al. (2004) made up one of two foci in the meta-analysis of negative correlations. Thus, the small number of foci and unstandardized testing may have limited the validity of our results for the negative correlation analysis.

Finally, the number of studies included in our final sample, both in Phase 1 and Phase 2, was rather small. Although this is a legitimate concern, we deliberately favored consistency in the methods reported across studies over sample size, given that small, precise meta-analyses are typically more informative (see for example Turner et al., 2013). We should also point out that the overall search initially generated a large body of 222 studies, and that Phase 1 included a final sample of 99 and 52 participants, respectively for the two contrasts, whereas Phase 2 included a final sample of 500 and 40 participants, respectively for each correlation analysis. The report of uncorrected clusters for both phases in Figs. 3 and 4, as well as the online publication of details for all analyses, are additional attempts to make transparent the process and materials of the present study. Although VBA findings to date do not seem to support fundamental differences in FA associated with reading disability, the addition of future studies might, and these can be easily incorporated into the analyses we reported. Gathering scientific evidence is a cumulative process, and by no means do we imply that the picture we present is definitive, nor do we make specific predictions about future studies. Rather, the present study highlights the lack of current evidence for consistent differences in FA as a function of reading ability, including extreme cases of reading disability.

## 4. Methods and materials

This section includes two related but distinct *meta-analytic* searches. The first search (Phase 1) corresponds to the initial contrast between dyslexics and typical readers. It was meant to provide an assessment of FA differences between these two groups. The second search (Phase 2) was intended to explore the relationship between reading ability and FA, within samples of typical readers. The Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) protocol guided the process and reporting of this meta-analysis in both phases (Moher et al., 2009). Data, including details about all papers reviewed, spatial coordinates, and all scripts for the reported and additional analyses are freely available online at [https://github.com/davidmoreau/2018\\_Brain\\_Research](https://github.com/davidmoreau/2018_Brain_Research)

### 4.1. Phase 1

#### 4.1.1. Search

Six databases—PubMed, PsycINFO, ScienceDirect, Scopus, ProQuest Dissertations & Theses Global, and Google Scholar—were searched from 17th–21st November 2016 for articles containing relevant material (see criteria below). The search on Google Scholar and ProQuest Dissertation and Theses Global database—which include unpublished dissertations and theses—were intended to minimize publication bias. As all relevant articles were accessible, no additional contact with the authors was required. We used two main search phrases. Firstly, we used the combined search phrases “diffusion tensor imaging and “dyslexia” in PubMed and PsycINFO

databases. These search phrases generated too many results when inputted into the remaining databases, ScienceDirect, Scopus, ProQuest Dissertations & Theses Global, and Google Scholar. Therefore, the search phrase was limited to only the title, abstract and keywords for the remaining database. Given this search limitation, synonyms for dyslexia were included. The search phrases used were: "diffusion tensor imaging", combined with each of the following: "dyslexia", "reading ability", "reading disability", "reading difficulty", and "reading impairment".

#### 4.1.2. Eligibility

Several eligibility criteria were employed to eliminate irrelevant articles from database searches when initially scanning the abstract, and later, the full text of articles (see flow diagram, Fig. 1). Exclusion criteria included: duplicates, articles in a foreign language, review articles, absence of a comparison between dyslexics and typical readers, studies focusing on reading abilities in typical samples, studies focusing on non-developmental dyslexia, clinical comorbidities, absence of structural comparison, ROI analyses, and methods articles. We detail these criteria hereafter.

Characteristics of participants included in the studies were thoroughly assessed. Given the aforementioned aim of this meta-analysis, studies that did not include a comparison group of dyslexic individuals were excluded. Similarly, we excluded studies in which dyslexia was not explicitly mentioned (e.g., poor reading ability), or those in which participants were not formally diagnosed with dyslexia (e.g., high risk factors but not diagnosis). Furthermore, individuals with pure neglect dyslexia or with spelling impairments only were excluded, as white matter correlates may differ from those of individuals with developmental dyslexia. Studies where dyslexic participants experienced clinical comorbidities, such as TBI, leukemia, pre-term birth or brain tumors, were also excluded because these may confound or provide alternative explanations for any relationships between FA and dyslexia.

Next, studies for which no structural DTI comparisons between dyslexic and control participants, or only including ROI DTI analyses, were excluded. For example, many studies only reported the correlations of FA differences to behavioral measures, such as performance on reading tasks. Methodological studies, such as scanning protocols for children and mathematical modeling, were also excluded. Finally, a handful of studies were excluded because they failed to report spatial coordinates. Details of this procedure are presented in Fig. 1.

Overall, 222 articles were identified in the six database searches. Ninety-four were duplicates and were thus removed. The remaining 128 articles were screened based on the abstract, with 89 excluded using the eligibility criteria. The full-text of the remaining 39 articles was assessed for relevance. Thirty-four studies were excluded at this stage, and the remaining five studies were included in the meta-analysis (see Fig. 1). More information about studies inclusion/exclusion can be found in the online material.

#### 4.1.3. Data collection and analyses

We extracted participant characteristics for each study. This included sample size, gender ratio, groups, mean age and age range, and any other relevant characteristics, such as native language. Then, we logged the coordinates of significant group differences in FA, with the corresponding reference space. The intended summary measures to quantify the outcomes of primary interest were cortical coordinates of voxels/clusters where FA significantly differed between dyslexic and typical reading individuals.

In the studies we selected, coordinates where FA significantly differed were reported in either MNI or Talairach space. Because we conducted the meta-analysis in MNI space, we transformed studies that reported coordinates in Talairach space (2) to MNI space, with the Lancaster et al. (2007) transformation integrated

in GingerALE 2.3.6 (Eickhoff et al., 2009, 2012; Turkeltaub et al., 2012). We used GingerALE 2.3.6 software (Eickhoff et al., 2009, 2012; Turkeltaub et al., 2012), which allows ALE implementation. ALE is the most common statistical technique for coordinate-based meta-analyses; the technique converts activation foci into probability distributions centered at specific coordinates. Activation foci are successively centered following a Gaussian probability distribution to generate single modeled activation maps for each study. These maps are then combined in a random-effects model, which provides flexible modeling of uncertainty across studies. Details of the ALE procedure in GingerALE can be found in the online documentation (<http://brainmap.org/ale/manual.pdf>) and in additional studies (Eickhoff et al., 2009, 2012; Lancaster et al., 2007; Turkeltaub et al., 2012).

Analyses were run separately for coordinates where controls had significantly greater FA than dyslexics (47 foci across 5 experiments), and where dyslexics had significantly greater FA than controls (17 foci across 2 experiments). In both analyses, a false discovery rate (FDR) threshold of 0.05 was applied (Laird et al., 2005). In addition, given that previous literature has suggested that the reduced and poorer quality reading exposure dyslexic readers likely experience throughout their life may contribute to apparent cortical differences between dyslexic and typical adult readers (Ben-Shachar et al., 2007), we ran separate additional analyses for children and adults. In analyses including coordinates where controls had greater FA than dyslexic readers, three studies included adults and two studies included children. One study including adults and one study including children reported coordinates where dyslexic readers had greater FA than controls.

## 4.2. Phase 2

Phase 1 included studies that specifically explored differences between dyslexic and typical readers. However, it is plausible that differences in FA are subtler and cannot be detected by simply contrasting groups, but could appear when considering a continuum to represent reading ability. With this rationale, many studies have investigated the association between reading ability and FA, regardless of a particular diagnosis of reading disability or dyslexia. To further investigate this relationship, the initial meta-analysis was complemented by a second search, detailed hereafter.

#### 4.2.1. Search

The same six databases used in Phase 1 were accessed and searched during Phase 2 from 24th to 26th January 2017. In Phase 2 of the study, the initial eligibility criteria were broadened to allow the inclusion of studies comparing dyslexic readers to typical readers via continuous measures spectrum, as well as studies exploring levels of reading ability among non-impaired readers. Eligibility criteria were identical to those of Phase 1, except for the exclusion of studies that did not include dyslexic participants, relevant in the previous search but not in the follow-up. In addition, studies that did not correlate FA with appropriate reading measures were excluded.

In Phase 2, alterations were made to the Phase 1 database searches to reflect the changes in the Phase 2 eligibility criteria. The Phase 1 searches in PubMed and PsycINFO used the narrow keyword of "dyslexia" only, which is unlikely to have generated articles with a range of reading abilities that did not include dyslexia comparisons; thus, these searches were insufficient to generate all required studies for Phase 2. Consequently, the searches in these databases were run again using the broader search phrase "diffusion tensor imaging" combined with each of the following: "dyslexia", "reading ability", "reading disability", "reading difficulty", and "reading impairment".

#### 4.2.2. Eligibility

Eligibility criteria were identical to those used in Phase 1, except for the substitution of group comparisons with FA-reading correlations. Firstly, when repeating the searches in PubMed and PsycINFO, the same process as Phase 1 of this study was undertaken. Secondly, studies from Phase 1 searches in ScienceDirect, Scopus, ProQuest Dissertations & Theses Global and Google Scholar that were eliminated because they did not include a dyslexic group, had no structural comparison, or were exclusively ROI analyses were reviewed for inclusion in Phase 2, for a total of 70 studies.

Overall, 171 articles were identified in the six database searches and reviews. Ninety-eight were duplicates and therefore excluded. The remaining 73 articles were screened by reading the abstract and full text where required. A further 64 were excluded based on the exclusion criteria. The remaining nine articles (ten studies in total) were included in the meta-analysis (see Fig. 2). More information about studies inclusion/exclusion can be found in the online material.

#### 4.2.3. Data collection and analyses

The same participant characteristics and reference space information as in Phase 1 were extracted from the articles. Details of reading task used to measure reading performance, and coordinates where this significantly correlated with FA, were also recorded. The intended summary effect measure to quantify the outcomes of primary interest were cortical coordinates of voxels/clusters where FA significantly correlated with reading ability, as measured by performance on reading tasks.

Coordinates of three studies required transformation from Talairach to MNI space using the Lancaster et al. (2007) transformation in GingerALE 2.3.6 (Eickhoff et al., 2009, 2012; Turkeltaub et al., 2012). Four studies reported findings of significant correlations between FA and multiple reading measures. When coordinates were not directly available, we performed a meta-analysis within individual studies to identify coordinates where FA significantly correlated with overall reading ability, as measured by all reading tasks in the particular study (Eickhoff et al., 2009, 2012; Turkeltaub et al., 2012). Analyses were run separately for coordinates where FA positively correlated with reading ability, and negative correlations. A FDR threshold of 0.05 was applied, given positive dependence of foci and common conventions (Laird et al., 2005).

Overall meta-analyses were conducted using GingerALE 2.3.6 software (Eickhoff et al., 2009, 2012; Turkeltaub et al., 2012). The overall ALE procedure was identical to this of Phase 1. Analyses were run separately for coordinates where FA positively correlated with reading ability, 42 foci across 9 experiments; and where FA negatively correlated with reading ability, 2 foci across 2 experiments. In both analyses, a FDR threshold of 0.05 was applied, given independence of foci and typical conventions (Laird et al., 2005). Finally, the meta-analysis was repeated separately for children and adults. Five of the ten studies analyzed included adult participants, whereas the remaining five studies included children. Two studies also reported negative correlations, both of which included child participants. To safeguard against unwarranted generalizations, we ran an additional analysis excluding the study of Zhang et al. (2014), which involved participants reading Chinese text, and compared the findings with those of the overall analysis. All the analyses input and output files can be found in the online material.

## 5. Conflict of interests

All authors declare no conflict of interests with this manuscript. There is no financial or personal interest that could affect our objectivity.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <https://doi.org/10.1016/j.brainres.2018.01.014>.

## References

- Assaf, Y., Pasternak, O., 2008. Diffusion tensor imaging (dti)-based white matter mapping in brain research: a review. *J. Mol. Neurosci.* 34 (1), 51–61. <https://doi.org/10.1007/s12031-007-0029-0>.
- Beaulieu, C., Plewes, C., Paulson, L.A., Roy, D., Snook, L., Concha, L., Phillips, L., 2005. Imaging brain connectivity in children with diverse reading ability. *NeuroImage* 25 (4), 1266–1271. <https://doi.org/10.1016/j.neuroimage.2004.12.053>.
- Ben-Shachar, M., Dougherty, R.F., Wandell, B.A., 2007. White matter pathways in reading. *Curr. Opin. Neurobiol.* 17 (2), 258–270. <https://doi.org/10.1016/j.conb.2007.03.006>.
- Brunswick, N., McCrory, E., Price, C. J., Frith, C. D., Frith, U., 1999. Explicit and implicit processing of words and pseudowords by adult developmental dyslexics. A search for Wernicke's Wortschatz? *Brain*, 122(10), 1901–1917. <http://doi.org/10.1093/brain/122.10.1901>
- Carp, J., 2012. On the plurality of (methodological) Worlds: estimating the analytic flexibility of fMRI experiments. *Front. Neurosci.* 6, 149. <https://doi.org/10.3389/fnins.2012.00149>.
- Carter, J.C., Lanham, D.C., Cutting, L.E., Clements-Stephens, A.M., Chen, X., Hadzipasic, M., Kaufmann, W.E., 2009. A dual DTI approach to analyzing white matter in children with dyslexia. *Psychiatr. Res. Neuroimaging* 172 (3), 215–219. <https://doi.org/10.1016/j.psychres.2008.09.005>.
- Deutsch, G.K., Dougherty, R.F., Bammer, R., Siok, W.T., Gabrieli, J.D.E., Wandell, B., 2005. Children's reading performance is correlated with white matter structure measured by diffusion tensor imaging. *Cortex* 41 (3), 354–363. [https://doi.org/10.1016/S0010-9452\(08\)70272-7](https://doi.org/10.1016/S0010-9452(08)70272-7).
- Dougherty, R.F., Ben-Shachar, M., Deutsch, G.K., Hernandez, A., Fox, G.R., Wandell, B., a., 2007. Temporal-callosal pathway diffusivity predicts phonological skills in children. *PNAS* 104 (20), 8556–8561. <https://doi.org/10.1073/pnas.0608961104>.
- Eickhoff, S.B., Bzdok, D., Laird, A.R., Kurth, F., Fox, P.T., 2012. Activation likelihood estimation meta-analysis revisited. *NeuroImage* 59 (3), 2349–2361. <https://doi.org/10.1016/j.neuroimage.2011.09.017>.
- Eickhoff, S.B., Laird, A.R., Fox, P.M., Lancaster, J.L., Fox, P.T., 2017. Implementation errors in the GingerALE Software: description and recommendations. *Hum. Brain Mapp.* 38 (1), 7–11. <https://doi.org/10.1002/hbm.23342>.
- Eickhoff, S.B., Laird, A.R., Grefkes, C., Wang, L.E., Zilles, K., Fox, P.T., 2009. Coordinate-based activation likelihood estimation meta-analysis of neuroimaging data: a random-effects approach based on empirical estimates of spatial uncertainty. *Hum. Brain Mapp.* 30 (9), 2907–2926. <https://doi.org/10.1002/hbm.20718>.
- Emsell, L., Van Hecke, W., Tournier, J.D., 2015. Introduction to Diffusion Tensor Imaging. In: W. Van Hecke, L. Emsell, & S. Sunaert (Eds.), *Diffusion Tensor Imaging: A practical handbook* (pp. 7–22). Springer New York. <http://doi.org/10.1016/B978-0-12-397025-1.00291-8>
- Frye, R.E., Hasan, K., Xue, L., Strickland, D., Malmberg, B., Liederman, J., Papanicolaou, A., 2008. Splenium microstructure is related to two dimensions of reading skill. *Neuroreport* 19 (16), 1627–1631. <https://doi.org/10.1097/WNR.0b013e328314b8ee>.
- Habib, M., Giraud, K., 2013. Chapter 23 – Dyslexia. In: *Handbook of Clinical Neurology* (Vol. 111, pp. 229–235). <http://doi.org/10.1016/B978-0-444-52891-9.00023-3>
- Hoefl, F., Meyler, A., Hernandez, A., Juel, C., Taylor-Hill, H., Martindale, J.L., Gabrieli, J.D.E. (2007). Functional and morphometric brain dissociation between dyslexia and reading ability. In: *Proceedings of the National Academy of Sciences of the United States of America*, 104(10), 4234–4239. <http://doi.org/DOI 10.1073/pnas.0609399104>
- Horwitz, B., Rumsey, J.M., Donohue, B.C., 1998. Functional connectivity of the angular gyrus in normal reading and dyslexia. *Proc. Natl. Acad. Sci.* 95 (15), 8939–8944. <https://doi.org/10.1073/pnas.95.15.8939>.
- Keller, T.A., Just, M.A., 2009. Altering cortical connectivity: remediation-induced changes in the white matter of poor readers. *Neuron* 64 (5), 624–631. <https://doi.org/10.1016/j.neuron.2009.10.018>.
- Klingberg, T., Hedehus, M., Temple, E., Salz, T., Gabrieli, J.D., Moseley, M.E., Poldrack, R.A., 2000. Microstructure of temporoparietal white matter as a basis for reading ability: evidence from diffusion tensor magnetic resonance imaging. *Neuron* 25 (2), 493–500. [https://doi.org/10.1016/S0896-6273\(00\)80911-3](https://doi.org/10.1016/S0896-6273(00)80911-3).
- Koerte, I.K., Willems, A., Muehlmann, M., Moll, K., Corneli, S., Pixner, S., Schulte-Körne, G., 2016. Mathematical abilities in dyslexic children: a diffusion tensor imaging study. *Brain Imag. Behav.* 10 (3), 781–791. <https://doi.org/10.1007/s11682-015-9436-y>.
- Laird, A.R., Fox, P.M., Price, C.J., Glahn, D.C., Uecker, A.M., Lancaster, J.L., Fox, P.T., 2005. ALE meta-analysis: Controlling the false discovery rate and performing statistical contrasts. *Hum. Brain Mapp.* 25 (1), 155–164. <https://doi.org/10.1002/hbm.20136>.
- Lancaster, J.L., Tordesillas-Gutiérrez, D., Martínez, M., Salinas, F., Evans, A., Zilles, K., Fox, P.T., 2007. Bias between MNI and Talairach coordinates analyzed using the ICBM-152 brain template. *Hum. Brain Mapp.* 28 (11), 1194–1205. <https://doi.org/10.1002/hbm.20345>.

- Lebel, C., Shaywitz, B., Holahan, J., Shaywitz, S., Marchione, K., Beaulieu, C., 2013. Diffusion tensor imaging correlates of reading ability in dysfluent and non-impaired readers. *Brain Lang.* 125 (2), 215–222. <https://doi.org/10.1016/j.bandl.2012.10.009>.
- Lindgren, S.D., De Renzi, E., Richman, L.C., 1985. Cross-national comparisons of developmental dyslexia in Italy and the United States. *Child Dev.* 56 (6), 1404–1417.
- Lyon, G.R., Shaywitz, S.E., Shaywitz, B.A., 2003. A definition of dyslexia. *Annals Dyslexia* 53 (1), 1–14. <https://doi.org/10.1007/s11881-003-0001-9>.
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G., The PRISMA Group, 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med* 6 (7), e1000097. <https://doi.org/10.1371/journal.pmed1000097>.
- Moreau, D., Kirk, I.J., Waldie, K.E., 2016. Seven pervasive statistical flaws in cognitive training interventions. *Front. Hum. Neurosci.* 10 (153).
- Nagy, Z., Westerberg, H., Klingberg, T., 2004. Maturation of white matter is associated with the development of cognitive functions during childhood 16 (7), 1227–1233. <https://doi.org/10.1162/0898929041920441>.
- Niogi, S.N., McCandliss, B.D., 2006. Left lateralized white matter microstructure accounts for individual differences in reading ability and disability. *Neuropsychologia* 44 (11), 2178–2188. <https://doi.org/10.1016/j.neuropsychologia.2006.01.011>.
- Norton, E.S., Beach, S.D., DE Gabrieli, J., 2015. Neurobiology of dyslexia. *Curr. Opin. Neurobiol.* 30, 73–78. <https://doi.org/10.1016/j.conb.2014.09.007>.
- Odegard, T.N., Farris, E.A., Ring, J., McColl, R., Black, J., 2009. Brain connectivity in non-reading impaired children and children diagnosed with developmental dyslexia. *Neuropsychologia* 47 (8), 1972–1977. <https://doi.org/10.1016/j.neuropsychologia.2009.03.009>.
- Paulesu, E., Demonet, J.-F., Fazio, F., McCrory, E., Chanoine, V., Brunswick, N., Frith, U., 2001. Dyslexia: cultural diversity and biological unity. *Science* 291 (5511), 2165–2167.
- Ramus, F., Altarelli, I., Jednoróg, K., Zhao, J., Scotto di Covella, L., 2018, January 1. Neuroanatomy of developmental dyslexia: Pitfalls and promise. *Neurosci. Biobehav. Rev.*, pp. 434–452. Pergamon. <http://doi.org/10.1016/j.neubiorev.2017.08.001>
- Richards, T., Stevenson, J., Crouch, J., Johnson, L.C., Maravilla, K., Stock, P., Berninger, V., 2008. Tract-based spatial statistics of diffusion tensor imaging in adults with dyslexia. *Am. J. Neuroradiol.* 29 (6), 1134–1139. <https://doi.org/10.3174/ajnr.A1007>.
- Richlan, F., 2014. Functional neuroanatomy of developmental dyslexia: the role of orthographic depth. *Front. Hum. Neurosci.* 8, 347. <https://doi.org/10.3389/fnhum.2014.00347>.
- Richlan, F., Kronbichler, M., Wimmer, H., 2009. Functional abnormalities in the dyslexic brain: a quantitative meta-analysis of neuroimaging studies. *Hum. Brain Mapp.* 30 (10), 3299–3308. <https://doi.org/10.1002/hbm.20752>.
- Rimrodt, S.L., Peterson, D.J., Denckla, M.B., Kaufmann, W.E., Cutting, L.E., 2010. White matter microstructural differences linked to left perisylvian language network in children with dyslexia. *Cortex; J. Devoted Study Nerv. Syst. Behav.* 46 (6), 739–749. <https://doi.org/10.1016/j.cortex.2009.07.008>.
- Rumsey, J.M., Andreason, P., Zametkin, A.J., Aquino, T., King, A.C., Hamburger, S.D., Cohen, R.M., 1992. Failure to activate the left temporoparietal cortex in dyslexia. *Arch. Neurol.* 49 (5), 527. <https://doi.org/10.1001/archneur.1992.00530290115020>.
- Shaywitz, B.A., Shaywitz, S.E., Pugh, K.R., Mencl, W.E., Fulbright, R.K., Skudlarski, P., Gore, J.C., 2002. Disruption of posterior brain systems for reading in children with developmental dyslexia. *Biol. Psychiatry* 52 (2), 101–110. [https://doi.org/10.1016/S0006-3223\(02\)01365-3](https://doi.org/10.1016/S0006-3223(02)01365-3).
- Shaywitz, B.A., Skudlarski, P., Holahan, J.M., Marchione, K.E., Constable, R.T., Fulbright, R.K., Shaywitz, S.E., 2007. Age-related changes in reading systems of dyslexic children. *Ann. Neurol.* 61 (4), 363–370. <https://doi.org/10.1002/ana.21093>.
- Shaywitz, S.E., Morris, R., Shaywitz, B.A., 2008. The education of dyslexic children from childhood to young adulthood. *Annu. Rev. Psychol.* 59 (1), 451–475. <https://doi.org/10.1146/annurev.psych.59.103006.093633>.
- Shaywitz, S.E., Shaywitz, B.A., Fulbright, R.K., Skudlarski, P., Mencl, W.E., Constable, R.T., Gore, J.C., 2003. Neural systems for compensation and persistence: young adult outcome of childhood reading disability. *Biol. Psychiatry* 54 (1), 25–33.
- Shaywitz, S.E., Shaywitz, B.A., Pugh, K.R., Fulbright, R.K., Constable, R.T., Mencl, W.E., Gore, J.C., 1998. Functional disruption in the organization of the brain for reading in dyslexia. *PNAS* 95 (5), 2636–2641.
- Simos, P.G., Breier, J.L., Fletcher, J.M., Bergman, E., Papanicolaou, A.C., 2000. Cerebral mechanisms involved in word reading in dyslexic children: a magnetic source imaging approach. *Cereb. Cortex* 10 (8), 809–816. <https://doi.org/10.1093/cercor/10.8.809>.
- Siok, W.T., Niu, Z., Jin, Z., Perfetti, C.A., Tan, L.H., 2008. A structural-functional basis for dyslexia in the cortex of Chinese readers. *PNAS* 105 (14), 5561–5566. <https://doi.org/10.1073/pnas.0801750105>.
- Soares, J. M., Marques, P., Alves, V., Sousa, N., 2013. A hitchhiker's guide to diffusion tensor imaging. *Front. Neurosci.* 7(7 MAR), 1–14. <http://doi.org/10.3389/fnins.2013.00031>
- Steinbrink, C., Vogt, K., Kastrup, A., Müller, H.-P., Juengling, F. D., Kassubek, J., Riecker, A., 2008. The contribution of white and gray matter differences to developmental dyslexia: insights from DTI and VBM at 3.0T. *Neuropsychologia* 46(13), 3170–3178. <http://doi.org/10.1016/j.neuropsychologia.2008.07.015>
- Turkeltaub, P.E., Eickhoff, S.B., Laird, A.R., Fox, M., Wiener, M., Fox, P., 2012. Minimizing within-experiment and within-group effects in activation likelihood estimation meta-analyses. *Hum. Brain Mapp.* 33 (1), 1–13. <https://doi.org/10.1002/hbm.21186>.
- Turner, R.M., Bird, S.M., Higgins, J.P.T., Lathlean, T., Babidge, W., 2013. The impact of study size on meta-analyses: examination of underpowered studies in cochrane reviews. *PLoS ONE* 8 (3), e59202. <https://doi.org/10.1371/journal.pone.0059202>.
- Van Hecke, W., Emsell, L., 2015. Strategies and challenges in DTI analysis. In: W. Van Hecke, L. Emsell, & S. Sunaert (Eds.), *Diffusion Tensor Imaging: A practical handbook* (pp. 153–174). Springer New York. <http://doi.org/10.1016/B978-0-12-397025-1.00291-8>
- Vanderauwera, J., Vandermosten, M., Dell'Acqua, F., Wouters, J., Ghesquière, P., 2015. Disentangling the relation between left temporoparietal white matter and reading: a spherical deconvolution tractography study. *Hum. Brain Mapp.* 36 (8), 3273–3287. <https://doi.org/10.1002/hbm.22848>.
- Vandermosten, M., Boets, B., Wouters, J., Ghesquière, P., 2012. A qualitative and quantitative review of diffusion tensor imaging studies in reading and dyslexia. *Neurosci. Biobehav. Rev.* 36 (6), 1532–1552. <https://doi.org/10.1016/j.neubiorev.2012.04.002>.
- Zhang, M., Chen, C., Xue, G., Lu, Z., Mei, L., Xue, H., Dong, Q., 2014. Language-general and -specific white matter microstructural bases for reading. *NeuroImage* 98, 435–441. <https://doi.org/10.1016/j.neuroimage.2014.04.080>.
- Zhao, J., Thiebaut de Schotten, M., Altarelli, I., Dubois, J., Ramus, F., 2016. Altered hemispheric lateralization of white matter pathways in developmental dyslexia: evidence from spherical deconvolution tractography. *Cortex* 76, 51–62. <https://doi.org/10.1016/j.cortex.2015.12.004>.