

SCIENTIFIC REPORTS

OPEN

Identification of *NCAN* as a candidate gene for developmental dyslexia

Elisabet Einarsdottir^{1,2}, Myriam Peyrard-Janvid², Fahimeh Darki³, Jetro J. Tuulari⁴, Harri Merisaari^{4,5}, Linnea Karlsson^{4,6}, Noora M. Scheinin^{4,7}, Jani Saunavaara⁸, Riitta Parkkola⁹, Katri Kantojärvi^{10,11}, Antti-Jussi Ämmälä^{10,11}, Nancy Yiu-Lin Yu¹², Hans Matsson², Jaana Nopola-Hemmi¹², Hasse Karlsson^{4,7}, Tiina Paunio^{10,11}, Torkel Klingberg³, Eira Leinonen¹ & Juha Kere^{1,2,13}

A whole-genome linkage analysis in a Finnish pedigree of eight cases with developmental dyslexia (DD) revealed several regions shared by the affected individuals. Analysis of coding variants from two affected individuals identified rs146011974G > A (Ala1039Thr), a rare variant within the *NCAN* gene co-segregating with DD in the pedigree. This variant prompted us to consider this gene as a putative candidate for DD. The RNA expression pattern of the *NCAN* gene in human tissues was highly correlated ($R > 0.8$) with that of the previously suggested DD susceptibility genes *KIAA0319*, *CTNND2*, *CNTNAP2* and *GRIN2B*. We investigated the association of common variation in *NCAN* to brain structures in two data sets: young adults (Brainchild study, Sweden) and infants (FinnBrain study, Finland). In young adults, we found associations between a common genetic variant in *NCAN*, rs1064395, and white matter volume in the left and right temporoparietal as well as the left inferior frontal brain regions. In infants, this same variant was found to be associated with cingulate and prefrontal grey matter volumes. Our results suggest *NCAN* as a new candidate gene for DD and indicate that *NCAN* variants affect brain structure.

Developmental dyslexia (DD) is characterized by literacy difficulties despite access to education, and normal vision, hearing and IQ¹. Detection of chromosomal abnormalities such as balanced translocations in families has led to the identification of at least three DD susceptibility genes (*DYX1C1*, *ROBO1*, *CYP19A1*)^{2–4}. This approach has also been used to identify genes implicated in learning difficulties similar to DD (e.g. *CTNND2*)⁵.

Linkage studies have led to identification of a number of other DD candidate genes, including *C2ORF3-GCFC2/MRPL19*⁶, *DCDC2*⁷, *KIAA0319*⁸, *CNTNAP5*⁹, *GRIN2B*¹⁰, and *FOXP2*¹¹. Genome-wide association studies (GWAS), however, have been less fruitful. This may be due to low power, as DD is a heterogeneous phenotype. Despite large collaborations such as the NeuroDys study¹², no such studies have been large enough to overcome this limitation. Family-based studies thus continue to be a powerful method of choice in DD. For example, exome-sequencing in such a family recently allowed us to identify the *CEP63* gene as a novel DD candidate gene¹³. A recent review by Mascheretti *et al.*¹⁴ gives a summary of the DD genes robustly supported by multiple studies as well as a number of highly interesting and plausible, but less studied putative DD candidates.

¹Folkhälsan Institute of Genetics, and Molecular Neurology Research Program, University of Helsinki, Helsinki, Finland. ²Department of Biosciences and Nutrition, Karolinska Institutet, Huddinge, Sweden. ³Department of Neuroscience, Karolinska Institutet, Solna, Sweden. ⁴University of Turku, Institute of Clinical Medicine, Turku Brain and Mind Center, FinnBrain Birth Cohort Study, Turku, Finland. ⁵Turku PET Centre, University of Turku, Turku, Finland. ⁶University of Turku and Turku University Hospital, Department of Child Psychiatry, Turku, Finland. ⁷University of Turku and Turku University Hospital, Department of Psychiatry, Turku, Finland. ⁸Department of Medical Physics, Turku University Hospital, Turku, Finland. ⁹Department of Radiology, Turku University Hospital and Turku University, 20520, Turku, Finland. ¹⁰Genomics and Biomarkers Unit, National Institute for Health and Welfare, Helsinki, Finland. ¹¹Department of Psychiatry, University of Helsinki and Helsinki University Central Hospital, Helsinki, Finland. ¹²Department of Paediatric Neurology, Helsinki University Central Hospital, Children's Hospital, Helsinki, Finland. ¹³Department of Medical & Molecular Genetics, King's College London, Guy's Hospital, London, UK. Correspondence and requests for materials should be addressed to E.E. (email: elisabet.einarsdottir@ki.se)

Previously, several studies have been performed to elucidate the connection between cognitive traits and variation in brain anatomy, e.g.^{15–19}. As shown by Klingberg *et al.*²⁰, white matter integrity in certain brain regions, such as the left temporoparietal cortex, plays a critical role in reading ability. This may be due to variation in the microstructure of white matter tracts, ultimately affecting communication between the cortical areas involved in reading and thus the processing of auditory and visual signals²⁰.

Structural brain imaging has repeatedly highlighted certain regions in the left hemisphere correlated with DD, e.g.^{5, 18, 21–23}, and white matter changes related to allelic variation in the DD susceptibility genes *DYX1C1*, *DCDC2* and *KIAA0319* are a recurring observation. Furthermore, multiple studies have shown correlation of variation in certain brain regions with genetic variants in specific genes, e.g. in *CEP63*¹³, *SLC2A3*²⁴, and *NRSN1*²⁵. Dehaene *et al.*, in their recent review, provide an overview of what is known regarding the mechanism of language acquisition and its impact on the brain²⁶.

In the present study, we aimed to identify novel genes and variants co-segregating with DD in a large Finnish family. We combined initial prioritization of candidate genomic regions by genome-wide linkage analysis, with subsequent exome sequencing to identify putative risk variants. The results highlighted the *NCAN* gene as a plausible DD candidate. Subsequently, we explored the association of *NCAN* variants to brain volumes of infants and 6 to 25-year old typically developing participants.

Material and Methods

All methods were carried out in accordance with the relevant guidelines and regulations.

DD pedigree. The DD pedigree was originally recruited for the genetic study of dyslexia by Nopola-Hemmi *et al.*^{27, 28} from the Department of Paediatric Neurology at the Hospital for Children and Adolescents, University of Helsinki, Finland. The criteria for the probands (children under 16 years of age) included in the study were a severe reading impairment, an extended family history of dyslexia (at least four affected individuals) and a pedigree suggestive of autosomal dominant transmission. All individuals were Finnish, of North European/Caucasian origin and native Finnish speakers. Participants from the nuclear families and their first- and second degree relatives were interviewed by a pediatric neurologist or neuropsychologist and were sent a detailed questionnaire regarding their reading and spelling difficulties, school history, and remedial education.

All individuals included in the analyses (both affected and unaffected) were tested by a clinical neuropsychologist to verify the diagnosis of dyslexia. The diagnostic assessment included intelligence tests (WAIS-R, WISC-R), age- and grade-appropriate Finnish reading and spelling tests^{29, 30} and a neuropsychological test³¹. The criteria for dyslexia included a pronounced history of reading problems, remarkable deviation in age-related reading skills (depending on the age, at least two years) and a normal performance intelligence quotient (IQ > 85). The neurocognitive type of dyslexia segregating in the DD family consisted of deficits in phonological awareness, verbal short-term memory, and rapid naming. As the diagnoses in the family were performed more than 15 years ago, certain phenotypic information may have been lost or never collected. The protocol for diagnosing DD was, however, well according to the accepted approach at the time. Due to this, we here analyse only the binary DD affected/not affected phenotype in the current study.

Genomic DNA from whole blood was available from 14 family members in the current study; 8 of them had confirmed DD (Fig. 1). Samples from the family were collected and analyzed according to ethical permissions Dnro 53/2006 and Dnro 4U/2016 by the ethics committee of the Central Finland Health Care District. Informed consent was obtained from all participating family members.

All available samples were genotyped on the Illumina HumanCoreExome12v1-0 genotyping chip (Illumina, San Diego, USA) and non-parametric multipoint linkage (NPL) analysis was run using Merlin version 1.12³² and the-exp option. The genotyping and linkage analysis are described in detail in Supplementary methods.

Next generation sequencing. One hundred ng of genomic DNA from the affected individual 3935 was used for exome sequencing using an AmpliSeq library on the Ion Proton platform (Life Technologies, Carlsbad, CA) at the Uppsala Genome Center (Science for Life Laboratory, Uppsala University, Uppsala, Sweden). Variants identified using the Ion Torrent Suite software (Life Technologies) were filtered to retain only non-synonymous-, splice site- or stop codons variants. These were filtered further, to retain rare variants of < 1% frequency in the 1000 Genomes ALL dataset (all populations)³³ and the ESP6500 dataset (all populations) within the regions of linkage (NPL > 1). The exome sequencing is described in more detail in Supplementary methods.

Two µg of genomic DNA from a second affected family member (3821) were used for whole-genome sequencing at Science for Life Laboratory, Stockholm, Sweden using a TruSeq library run on Illumina HiSeqX and the GATK best-practice pipeline. The sequencing is described in more detail in the Supplementary methods. As we hypothesized a highly penetrant autosomal dominant inheritance of DD in the family and a variant of strong risk-effect, we opted to extract only exonic variants and focused on variants shared by the two sequenced individuals, 3821 and 3935. The filtering steps were as above.

WES and WGS sequence data are available in the DDBJ/EMBL/GenBank databases under the accession number PRJEB12695. We used the human genome reference hg19 for all genomic positions of variants described here. Individuals 3935 and 3821 were selected for sequencing based on being affected carriers of the putative risk haplotype and due to the availability of high quality DNA.

Gene expression correlation analysis. The Functional Annotation of the Mammalian Genome 5 (FANTOM5) human/mouse tissue and cell line promoter expression database³⁴ was used to examine the expression of *NCAN* and compare it to the global expression levels of a representative set of DD candidate genes (*MRPL19*, *KIAA0319*, *ROBO1*, *CTNND2*, *DYX1C1*, *CEP63*, *KIAA0319L*, *PCNT*, *GCFC2*, *CYP19A1*, *DCDC2*, *FOXP2*, *CNTNAP2* and *GRIN2B*) in 127 of the 152 tissue samples in the FANTOM5 dataset. We looked at the

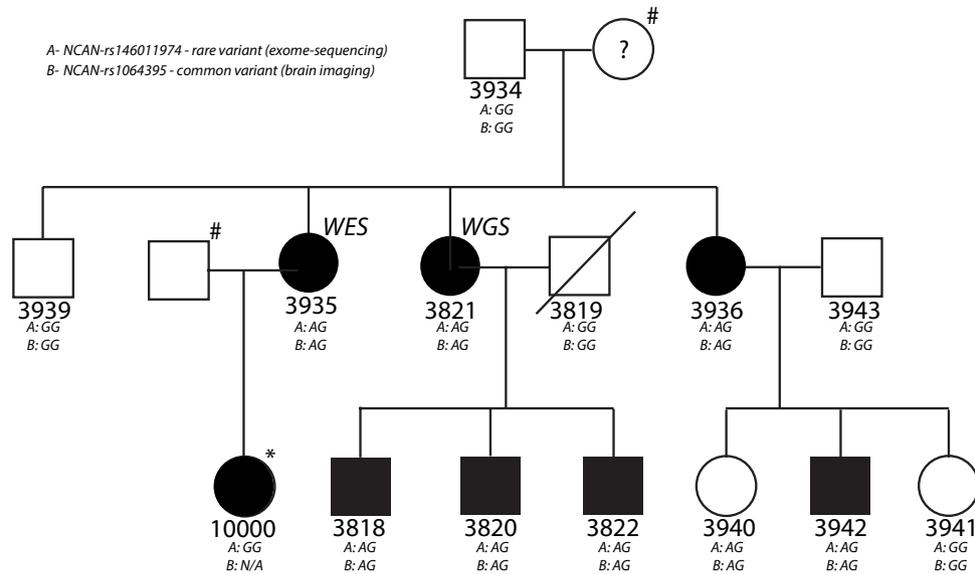


Figure 1. Pedigree of the presently studied family with developmental dyslexia (DD). Open boxes/circles indicate individuals without DD (confirmed no DD diagnosis), filled boxes/circles indicate individuals with confirmed DD. A question mark indicates an unknown phenotype. The exome-sequenced individual is marked with a “WES” and the whole-genome sequenced individual is marked with a “WGS”. The genotype of each individual for both *NCAN* markers is shown: (A) rs146011974 (a rare variant, identified through exome sequencing) and (B) rs1064395, (common variant, used in the brain imaging analysis). Individuals without DNA sample are indicated by the # symbol. Individual 10 000 (marked with a star) was not included in the linkage analysis. N/A = Genotype not available.

expression in all available human tissues (to get a general picture of the expression patterns), in brain tissue only (presumably yielding stronger correlation within the set of genes expressed primarily in neurological tissue), and in non-brain tissue (to further look into the expression patterns of the DD risk genes that are more ubiquitously expressed). Universal tissues, total RNA samples, and tissues with too shallow sequencing depths or low RNA quality were excluded from the analysis. The final dataset included 78 non-brain tissues and 49 brain tissues, listed in Supplementary Table 1. The ‘robust’ set of promoters was used for the gene-based analysis.

Spearman correlations (ρ) between the gene expressions were calculated for brain tissues, non-brain tissues, and all tissues (brain and non-brain tissues). As this non-parametric, ranking-based method gives R-values from -1 (complete negative correlation of ranks) to 1 (complete positive correlation of ranks), we here considered R-values of 0.8 or higher to indicate strong (positive) correlation. Hierarchical clustering was used to produce the correlation heatmaps.

Brain imaging. As common variants in a number of previously identified DD susceptibility genes have shown significant associations with certain brain structures, we analyzed the potential association of a common variant in *NCAN* with structural variation in brain MRI data. We chose the polymorphism rs1064395 (NC_00019.10:g.19250926 G > A) in *NCAN*, a common variant previously associated with cognitive performance in healthy individuals³⁵ and with grey matter volumes³⁶. The minor alleles of rs1064395 and the rare variant rs146011974 (introduced further in Results) are carried on the same rare haplotype (data not shown) and therefore likely tag the same rare haplotype segregating with DD in the present family.

A longitudinal dataset (Brainchild) of brain imaging in 76 healthy Swedish children and young adults (age 6 to 25 years, 41 males and 35 females) was used to test for correlation of the *NCAN* genetic variant to white- and grey matter volumes. These same individuals have been studied previously^{5, 13, 21, 22, 37}. Informed consent was obtained from all subjects (ethical approvals 2007/241–31/3 and 2012/116–32, Stockholm, Sweden); genomic DNA was analyzed using the Affymetrix Genome-wide SNP array 6.0 (Santa Clara, CA, USA). T1-weighted brain imaging was performed three times, each two years apart, by 3D magnetization prepared rapid gradient echo (MP-RAGE) sequence with TR = 2300 ms, TE = 2.92 ms, field of view of $256 \times 256 \text{ mm}^2$, 176 sagittal slices, and 1 mm^3 voxel size. Diffeomorphic Anatomical Registration Through Exponentiated Lie Algebra (DARTEL)³⁸ was performed to segment the brain into grey matter, white matter and CSF. An 8-mm Gaussian kernel was applied to the segmented white matter images. All modulated white and grey matter images were then analyzed using a flexible factorial design in Statistical Parametric Mapping (SPM) software (www.fil.ion.ucl.ac.uk/spm). The genetic variant rs1064395 was entered as a main factor, and the model was corrected for age, sex, handedness and total size of white matter volume. The gene interactions by age and sex were also added into the model. The main effect of the genetic variant on white matter structure was assessed by considering two thresholds ($p = 0.01$ and $p = 0.001$), and the clusters found significant at FDR corrected at the cluster level using nonstationary cluster extent correction³⁹. The genotype counts for rs1064395 (*NCAN*) were 56 (CC) and 20 (CT). The frequency of the minor allele (T) of rs1064395 was 13% in the Brainchild dataset. To assess the overlap between the regions of white matter we

Chromosome	max NPL	max NPL p-value	Locus	Size of locus (kb)
2	1.62	0.003	tel-rs6548285	963
	1.62	0.003	rs6706713-rs891881	5559
	1.62	0.003	rs1478644-rs12615297	855
3	1.62	0.003	rs11718068-rs6777084	3527
4	1.62	0.003	rs2353563-rs306364	16387
6	1.36	0.006	rs4896431-rs9376745	4246
8	2.63	0.0002	rs10088564-rs2527760	2879
18	1.36	0.006	rs695107-rs9950625	10608
19	1.36	0.006	rs11086080-rs1019937	12909
22	1.36	0.006	rs5750807-rs743377	2496

Table 1. Regions of NPL linkage > 1 in the DD pedigree. NPL = nonparametric lodscore.

overlaid the current NCAN findings at a threshold of $p = 0.01$ with the previous results from our imaging studies on DD/reading disability^{5, 13, 21, 22, 37} and susceptibility genes. Significant regions were overlaid on a human brain MRI template using MRICro software (www.mccauslandcenter.sc.edu/mricr/).

A second, independent dataset was used to explore the effect of rs1064395 on early structural characteristics of the brain. The FinnBrain Cohort is a Finnish general population-based pregnancy cohort where the main interest is to delineate the associations between maternal psychological well-being during pregnancy and the future development of the children (www.finnbrain.fi, Karlsson *et al.*, submitted). The study was approved by the Ethics Committee of the Hospital District of Southwest Finland (ETMK: 31/180/2011 nr. 210) and informed consent was obtained from all subjects. The FinnBrain sub-sample available for the current study included 26 healthy, term-born infants (14 girls and 12 boys), imaged at mean age 22.5 days (SD 7.2) from birth. All MRI scans were performed at the Turku University Hospital using Siemens Magnetom Verio 3 T scanner (Siemens Medical Solutions, Erlangen, Germany). 12-element Head Matrix coil allowed the use of the Generalized Autocalibrating Partially Parallel Acquisition (GRAPPA) technique to accelerate acquisitions (PAT factor of 2 was used). 2D Dual Echo TSE (Turbo Spin Echo) sequence was used to acquire anatomical PD- and T2-weighted images. Parameters were optimized so the “whisper” gradient mode could be used in order to reduce acoustic noise during the scan. Slice thickness was 1 mm in order to acquire isotropic $1.0 \times 1.0 \times 1.0$ mm voxels. TR time of 12070 ms and effective TE times of 13 ms and 102 ms were used to produce both PD- and T2-weighted images from the same acquisition. The total number of slices was 128. Only T2-weighted images were used in the subsequent analysis. Visual quality control (QC) was performed on the acquired volumes. Images were analyzed with iBEAT software¹³ that enables N3 bias correction, tissue segmentation and anatomical labeling (AAL atlas). iBEAT segmentation produces RAVENS maps that resemble modulated segmentation maps that are used in adult voxel-based morphometry (VBM) with more reliable metrics on highly abnormal brains (eg. infants). Subject to well-known uneven intensity distribution of the infant brain¹⁴, there were sub optimal results in segmenting the white matter, namely, portions of subcortical grey matter structures and myelinated central white matter appeared in grey matter segments. Correspondingly, white matter analysis in this data set was deemed unreliable and was not used in the analysis. Grey matter RAVENS maps were smoothed with 8 mm FWHM in SPM12 (<http://www.fil.ion.ucl.ac.uk/spm/software/spm12/>). Smoothed segmentations were entered into statistical modeling within the SPM. Similar to the Brainchild data set, we used a flexible factorial model, where the genetic variant rs1064395 was entered as a main factor and gestational age and sex as nuisance variables. The results were thresholded with two thresholds to provide an open view to the results: at $p < 0.001$ and with $p < 0.01$, FDR corrected ($p < 0.05$) for multiple comparison at the cluster level. The analysis was carried out in SPM and the MNI coordinates (in adult space) were used to identify the brain regions by overlaying the contrast on top of iBEAT AAL template with mricron.

In a complementary analysis, we used the individual tissue volumes that are produced in the iBEAT image processing pipelines. The volumes were calculated from the AAL template labeling and represent individual grey matter volumes (as contrasted to relative volumes/density of RAVENS maps).

The genotype counts for rs1064395 were 20 (CC) and 6 (CT). Genotyping was performed at the Estonian Genome Centre (Tartu, Estonia) on the Illumina Infinium PsychArray BeadChip. See Supplementary methods for further details on the analysis.

Results

Ten suggested regions of linkage to DD. We performed linkage analysis using single nucleotide variant genotype data. NPL plots and scores are shown in Supplementary Fig. S1 and Supplementary dataset S1, respectively. In the pedigree under study (Fig. 1), we identified ten genomic regions with NPL scores > 1 (Table 1) constituting the putative regions of interest in the subsequent exome data analysis. These regions did not overlap with the findings in Kaminen *et al.*⁴⁰, where only a subset of the current pedigree was included (data not shown).

Analysis of exome variants in the linkage regions. We extracted the coding variants in affected individuals 3935 and 3821 (Fig. 1). Table 2 shows the number of variants at each filtering step and in each region of linkage while Supplementary datasets S2 and S3 show the full list of these variants in 3935 and 3821, respectively. We identified only one rare non-synonymous variant within the regions of linkage that was shared by both individuals; it was located on chromosome 19. Sanger sequencing confirmed that all but one DD affected individual

		(WES)	(WGS)	
		3935	3821	
	total variants	48 833	4 058 706	
	PASS & depth > 9	48 015	3 578 878	
	exonic	20 484	24 646	
	exonic (ns/stop)	9 556	11 718	
	exonic (ns/stop) autosomal	9 389	11 501	
	<1% 1000Gall & ESP6500all autosomal	903	1304	
	in linkage regions (b37)	3935	3821	in common
2	tel-963334			
	50721511–56280428	1		
	70576997–71431885			
3	152631729–156158932			
4	115264835–131652041	1	1	
6	139411399–143657292			
8	2545224–5425042			
18	60597754–71205938	1	1	
19	17927041–30836248	14	2	1 (NCAN)
22	22600912–25097248	1		

Table 2. Exome variants identified in the linkage regions at each stage of filtering. Ns = non-synonymous. WES = whole-exome sequencing. WGS = whole-genome sequencing.

(10000) and one of the unaffected (3940), shared this variant (variant A in Fig. 1). This variant (rs146011974, NC_000019.10:g.19233884 G > A) is located within the *NCAN* gene (neurocan, also known as chondroitin sulfate proteoglycan 3, *CSPG3*), a gene expressed in multiple regions of the brain. It is predicted to result in an alanine to threonine change (NM_004386.2:c.3115 G > A (p.(Ala1039Thr)) and to be tolerated or probably benign by SIFT and Polyphen2, but possibly pathogenic by M-CAP⁴¹. Information on this variant has been submitted to the ClinVar database (www.ncbi.nlm.nih.gov/clinvar/), accession number SCV000330897. The alanine version of the protein is conserved in most mammals, including mouse, rat, chimp and cow. The minor allele (A) of rs146011974 has a frequency of 0.016% in the Genome Aggregation Database (gnomAD, gnomad.broadinstitute.org, accessed 7th July 2017) all-populations dataset, 0.04% in the Finnish gnomAD population, 0.03% in the Sequencing Initiative Suomi (SISU database, www.sisuproject.fi, accessed 7th July 2017), and it is not found at all in the 1000Genomes database. The rarity of this variant and the expression pattern of *NCAN* in human brain suggested that this variant might qualify as a candidate for DD susceptibility with high penetrance.

Gene expression pattern correlations with other dyslexia candidate genes. In order to better understand the role *NCAN* may play in DD, we studied the correlation of *NCAN* expression with the expression profiles of a small subset of previously described known or candidate DD susceptibility genes across a wide selection of human tissues/organs. The aim was not to look systematically at all known or suggested DD genes, merely to compare *NCAN* to a subset of these and test if *NCAN* might make a plausible addition to this group. Figure 2 shows heat maps of these correlations. When all the tissue samples from the FANTOM5 dataset (Fig. 2A) were included in the analysis, we noted that *KIAA0319*, *CTNND2*, *CNTNAP2*, *GRIN2B* and *NCAN* showed strong correlation in global RNA expression levels ($R > 0.80$, Supplementary Table S2). A second cluster of putative DD genes comprised *DYX1C1*, *ROBO1*, *MRPL19*, *KIAA0319L*, *CEP63*, and *PCNT*. A further putative cluster of DD genes included *GCFC2*, *DCDC2*, *FOXP2* and *CYP19A1*.

For expression in brain only (Fig. 2B), we observed a strong correlation ($R = 0.84$) between *NCAN* and *GRIN2B*, and a somewhat weaker one between *NCAN* and *KIAA0319* ($R = 0.75$) (Supplementary Table S2). The correlations were weaker in non-brain tissues (Fig. 2C), the strongest being between *NCAN* and *CNTNAP2* ($R = 0.65$).

Genetic variation and brain structure. We have previously presented genotyping and neuroimaging data indicating that common genetic variation in several DD candidate genes significantly associated with white or grey matter volumes^{5, 13, 21, 22, 37}. In order to further investigate the role of *NCAN* in DD, we analyzed the association of a common genetic variant (rs1064395) in *NCAN* with structural white and grey matter variation in a longitudinal sample of 76 children and young adults (Brainchild dataset).

We found significant association of the minor allele of rs1064395 with increased white matter volume in the right temporoparietal and frontal regions ($p = 0.01$; p at cluster level: $p_{\text{FWE corrected}} = 1.56 \times 10^{-6}$) and the left temporoparietal, frontal and occipital regions ($p = 0.01$; p at cluster level $p_{\text{FWE corrected}} = 4.48 \times 10^{-6}$) (in red in Fig. 3A and Supplementary Table S3). We repeated the analysis with a lower p threshold to have fewer false positive voxels. The same clusters were significant at this threshold level (shown in orange in Fig. 3A and in Supplementary Table 4). The temporoparietal cluster partly overlapped with the superior longitudinal fasciculus, while the frontal and occipital clusters partly covered the inferior fronto-occipital fasciculus based on the John Hopkins University

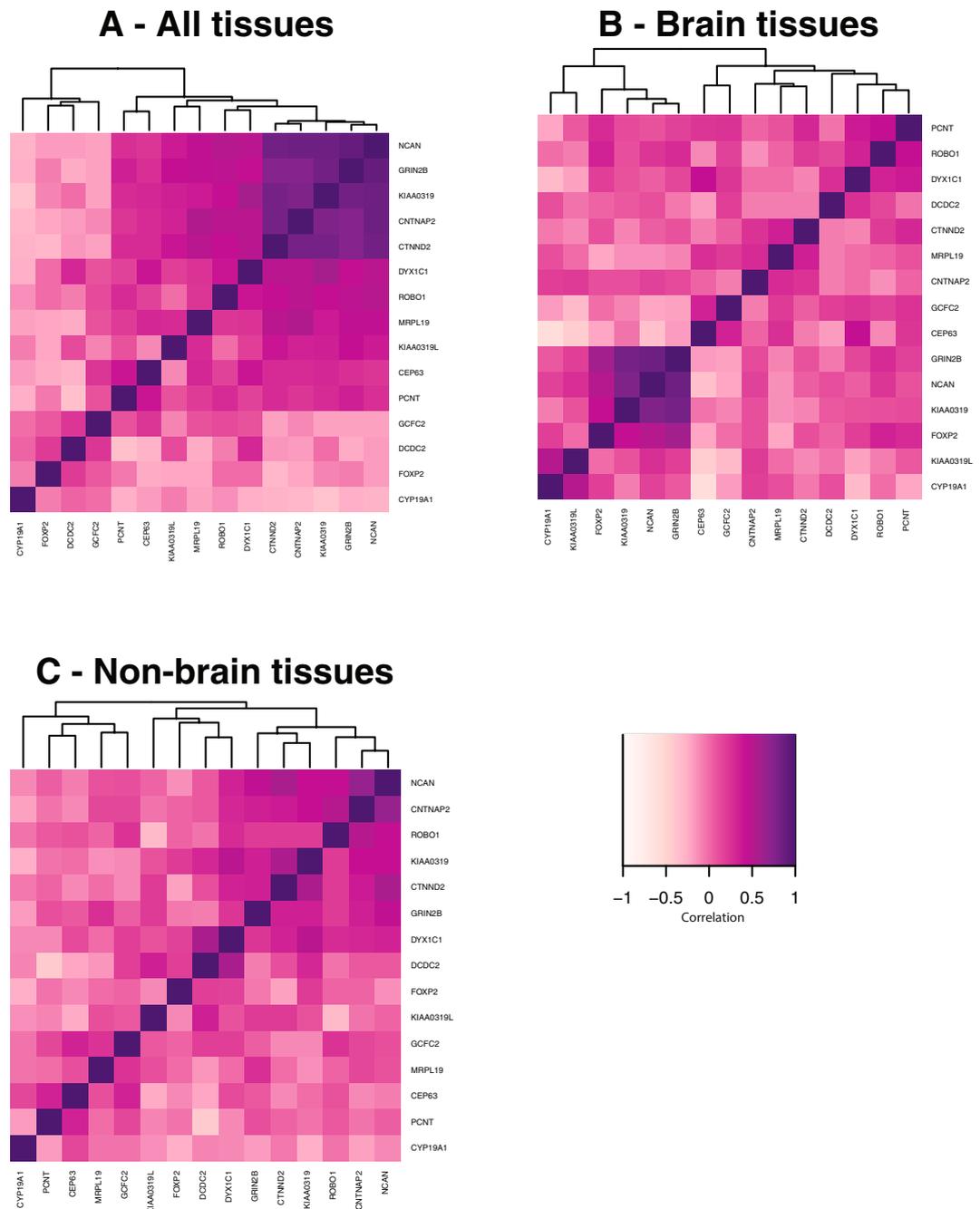


Figure 2. Correlation of mRNA expression of the dyslexia candidate genes from the FANTOM5 database. Spearman correlation of the global RNA expression of the previously described dyslexia candidate genes *KIAA0319*, *CTNND2*, *DYX1C1*, *ROBO1*, *MRPL19*, *KIAA0319L*, *CEP63*, *PCNT*, *GCFC2*, *DCDC2*, *CYP19A1*, *CNTNAP2*, *FOXP2* and *GRIN2B* as well as *NCAN*, the putative DD candidate gene identified in the present study. A dark purple color indicates strong positive correlation, while lighter color indicates negative correlation. (A) 127 tissue samples, (B) subset of 49 brain tissues and (C) subset of 78 non-brain tissues. Expression data was extracted from the FANTOM5 dataset, at <http://fantom.gsc.riken.jp/5/data/>.

(JHU) white matter atlas⁴². Moreover, we found an association of the same allele with greater grey matter volume in the right superior temporal cortex (cluster level $p_{\text{uncorrected}} = 0.009$, peak coordinate: 68, -7, -5). However, this region did not survive FDR or FWE correction at the cluster level (cluster level $p_{\text{FWE corrected}} = 0.22$).

Previous studies have shown that the minor allele variants in several known DD candidate genes associate with larger white matter volume in certain brain regions^{5, 13, 21, 22, 37}. We overlaid these brain regions with the *NCAN* regions from the present study. The right temporoparietal region associated with rs1064395 overlapped with a region previously associated with the DD susceptibility genes *KIAA0319*, *DYX1C1* and *MRPL19*²¹ as well as *CEP63*, a more recently published DD susceptibility gene¹³ (Fig. 3B and Supplementary Table S3). In addition, the left frontal region associated with rs1064395 overlapped with a region associated with variation in *CTNND2*,

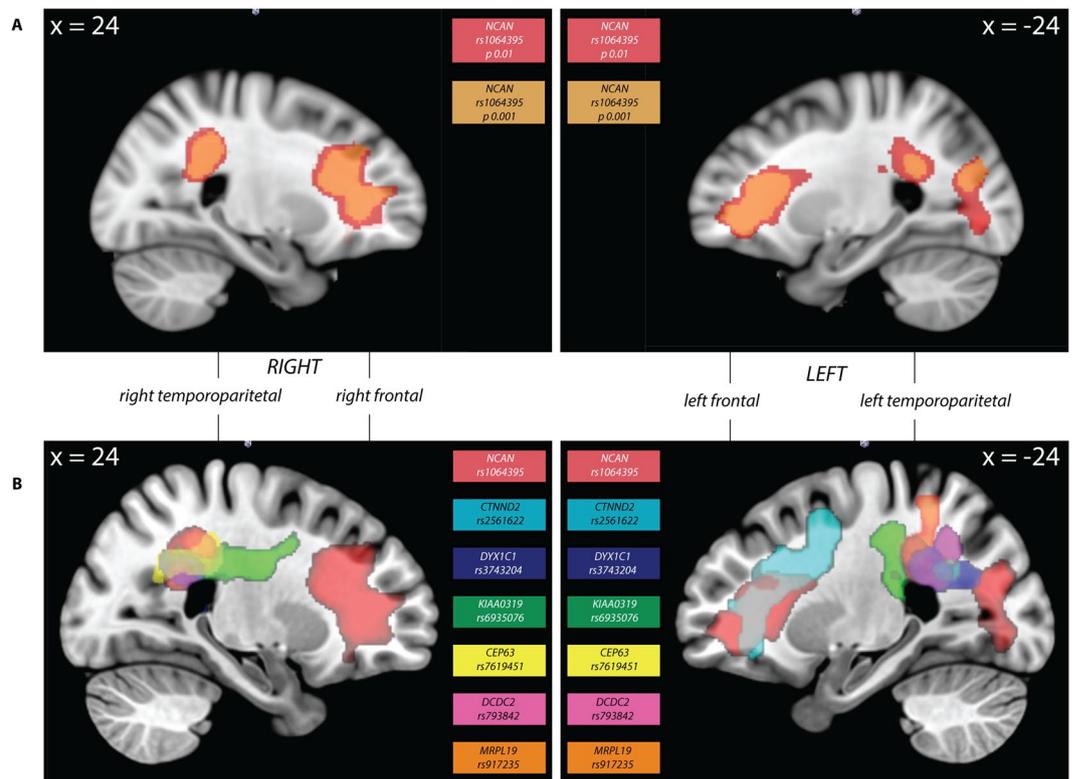


Figure 3. Common variation in DD susceptibility genes correlated with white matter volume. **(A)** Regions where white matter structure is significantly associated to common variation in *NCAN* (rs1064395; $p < 0.01$ shown in red and $p < 0.001$ shown in orange). **(B)** Summary of all the regions of white matter volume significantly associated to variation in *DYX1C1* (rs3743204, dark blue), *DCDC2* (rs793842, pink), *KIAA0319* (rs6935076, green), *MRPL19* (rs917235, purple), *CTNND2* (rs2561622, cyan), *CEP63* (rs7619451, yellow), and *NCAN* (rs1064395, red).

another recent candidate DD susceptibility gene⁵. White matter volume in a third region, the left temporoparietal region, was associated with variation in the *DYX1C1*, *DCDC2* and *KIAA0319* genes²¹ as well as in the *MRPL19* gene³⁷. In summary, these results showed that brain regions for which white matter volume associated with genetic variation in *NCAN* overlapped to a significant extent with those previously implicated for other DD susceptibility genes.

To further explore the effect of the *NCAN* rs1064395 variant on brain structure, we looked at its associations with grey matter volumes in an independent set of infants (the FinnBrain study). We found that the minor allele (CT) of rs1064395 was associated with increased volume in infant grey matter in the left inferior parietal lobule, the right precentral gyrus and the right middle frontal gyrus. With a more lenient thresholding ($p < 0.01$) larger grey matter volumes associated to the minor allele in the bilateral middle cingulate cortices and superior frontal cortices ((Fig. 4, Supplementary Table S5). Regression analysis on volumes yielded results that are in line with the VBM analysis (Supplementary Table S6).

While we were unable to identify any putative DD variants within the linkage region on chromosome 8, it contains the 5' and upstream regions of *CSMD1* (CUB and Sushi Multiple Domains 1) gene. Variants in *CSMD1* have been associated with disorders and traits affecting the central nervous system such as bipolar disorder disease⁴³ and schizophrenia^{44–46}; as well as general cognitive ability and executive function in healthy human subjects⁴⁷. This gene may thus be of interest also in future studies of DD and related cognitive traits.

Discussion

We used linkage analysis to prioritize potential candidate regions for autosomal dominant DD in a multiplex pedigree, followed up by evaluation of exome-sequencing data in those regions. This allowed us to highlight a new plausible candidate gene for DD, *NCAN*. Neither this gene nor genomic region have been previously suggested by linkage or GWA studies as a susceptibility locus or gene for dyslexia. While our average sequencing depth is satisfactory, we cannot exclude the existence of additional variants that would go unnoticed because of inexact base calling. Furthermore, even though the detected *NCAN* variant segregating with DD in the studied family leads to an amino-acid substitution, the true genetic risk variant(s) might conceivably be non-coding variation.

The rare *NCAN* variant identified in the current study is unlikely to explain a significant proportion of DD in the general population, and any studies of this specific variant are likely to be underpowered. This finding served,

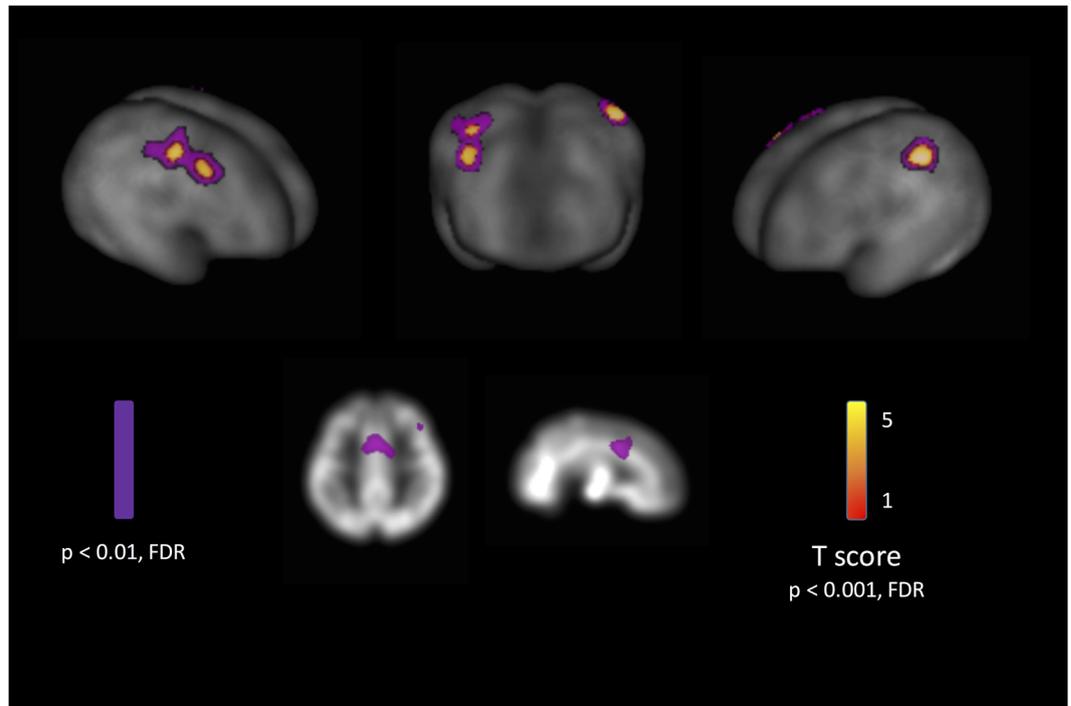


Figure 4. Association between the common *NCAN* genetic variant and infant grey matter structure (VBM). Upper panel - Cortical rendering of the results with thresholds $p < 0.001$ (red-to-yellow) and $p < 0.01$ (violet), FDR corrected for multiple comparisons. Lower panel - selected section from the same results, highlighting the cingulate associations.

however, to highlight the *NCAN* gene as a potential novel DD susceptibility gene, and our subsequent work thus focused on understanding the general involvement of this gene in DD.

NCAN shares similar expression profiles with *CNTNAP2*, *CTNND2*, *KIAA0319* and *GRIN2B*, with higher expression in several brain regions and lower expressions in the rest of the body. This was reflected in the strong correlations seen in the expression correlation heatmaps. The specific functions of these genes and how their dysfunction might contribute to DD susceptibility are only partially understood. It is, however, tempting to speculate that these genes may represent a group of genes involved in shared mechanisms or pathways that are critical for the optimal development of the brain and its function in literacy skills. DD genes with different patterns of expression, e.g. the *CEP63* and *DCDC2* genes that are expressed in most of the tissues in the human body, are likely to contribute to DD risk through other, potentially less brain-specific, mechanisms.

NCAN is important for cell adhesion and neuronal cell migration and is a negative regulator of neurite outgrowth⁴⁸. The rare variant rs146011974 lies within a highly conserved region in an EGF-like domain near the C-terminus of the protein. While it remains unclear how an amino acid change in this domain might affect DD risk, slight variation in signaling through this domain could plausibly influence communication of neurons with each other and their shared environment.

Carriers of the rare *NCAN* variant were unfortunately not available for imaging analysis. Instead, we looked further at how common variation in *NCAN* might affect white and grey matter volumes in two independent imaging datasets. The rationale behind this approach is that if a rare high-risk variant in *NCAN* drives DD susceptibility in the pedigree under study, common variants in the same gene might modulate brain structure in the general population. The common variant rs1064395 within the *NCAN* 3' UTR here correlates with variation in white and grey matter structures and has previously been implicated in a number of psychiatric phenotypes such as bipolar disease⁴⁹, schizophrenia⁵⁰, cortical folding in schizophrenia⁵¹ and mania⁵². This is not the first time the same genetic variants have been connected to both dyslexia and other neurological disorders^{53,54}.

The white matter variability linked to *NCAN* in the Brainchild dataset was found in the left and right temporoparietal, occipital and frontal regions. Of these regions, the left inferior frontal area, with an extension to the region functionally defined as the Broca's area, is known to be of importance for language processing, and overlaps with the region previously associated with *CTNND2*⁵. The other cluster in the left temporoparietal region is also associated with several previously reported DD candidate brain regions^{22,37}. Disrupted brain activation patterns have also been reported in the left temporoparietal region when comparing poor readers to normal controls^{55–57}. Moreover, both grey matter and white matter deviations in this region have been related to dyslexia and impaired reading^{17,20,58}. A recent study³⁵ showed association of rs1064395 with brain activation in the temporal lobe during a semantic verbal fluency task in healthy subjects. Previous studies have also reported associations of the same genetic variant with grey matter volumes of the hippocampus and amygdala³⁶ as well as with volumes of the occipital region and prefrontal cortex⁵¹.

We went on to look further at this common *NCAN* variant in an independent dataset based on infant brain imaging. We were particularly interested in testing if the putative novel DD susceptibility gene might affect brain structure already at such a young age. We detected associations of grey matter volumes with *NCAN* variation in the right superior frontal cortex, inferior parietal and bilateral cingulate cortices in the infants. Interestingly, the direction of the effect was the same; the minor allele of rs1064395 was correlated with increased white matter volumes in the older imaging population and increased grey matter volumes in the infant dataset.

This suggests that *NCAN* may be important for efficient neuronal development, with the minor and putative DD risk allele of rs1064395 correlating with less efficient neuronal guidance. Such a defect has been suggested previously also with *DYX1C1*, *DCDC2*, *ROBO1* and *KIAA0319* (reviewed in ref. 8). Although the brain regions implicated are not language areas *per se*, the variation in them may modulate later cognitive performance in general¹, and thus have intricate connections to reading, which requires a coordinated brain network function that directs attention⁵⁹. It is possible that the white matter associations relate to the implicated initial grey matter characteristics, as the maturation of the white and grey matter in frontoparietal networks is strongly co-modulated⁶⁰, but this remains to be studied with additional imaging techniques such as DTI. Of additional importance is the fact that the FinnBrain imaging study was performed soon after birth, minimizing the influence of postnatal environmental factors and thus yielding a possible insight into brain structural development as early as possible to *in utero*.

Although we were able to assess the genetic associations to brain morphology in a wide spectrum of ages (from infants to young adults), direct comparison of the findings from the two datasets is problematic and larger longitudinal datasets would be needed. Growing interest in low frequency variants also means that future studies need to be larger in order to attain an acceptable size for each genotype group.

The small size of both imaging datasets, the inherent limitations in studying only a common variant, and the questions arising recently about the validity of certain imaging studies⁶¹ and their power⁶², call for caution in drawing conclusions. However, with these datasets, we can lend robust support for the idea that *NCAN* is related to brain developmental processes and structural development at the more general level. How these findings directly relate to development of DD needs to be studied further.

In conclusion, we report here a family with multiple cases of DD individuals, and highlight *NCAN* as a putative novel DD susceptibility gene, which shares similar RNA expression profiles throughout multiple tissues within the human body with *KIAA0319*, *CTNND2*, *CNTNAP2* and *GRIN2B*. The brain imaging data reported in the present study support that *NCAN* variants have effects on brain structure from infancy to early adulthood, but the possible associations between genotype, neuroimaging and phenotype remain to be addressed in future studies.

References

- Habib, M. & Giraud, K. Dyslexia. *Handbook of clinical neurology* **111**, 229–235, doi:10.1016/B978-0-444-52891-9.00023-3 (2013).
- Anthoni, H., Sucheston, L. E. & Lewis, B. A. *et al.* The aromatase gene CYP19A1: several genetic and functional lines of evidence supporting a role in reading, speech and language. *Behavior genetics* **42**, 509–527, doi:10.1007/s10519-012-9532-3 (2012).
- Hannula-Jouppi, K., Kaminen-Ahola, N. & Taipale, M. *et al.* The axon guidance receptor gene ROBO1 is a candidate gene for developmental dyslexia. *PLoS genetics* **1**, e50, doi:10.1371/journal.pgen.0010050 (2005).
- Taipale, M., Kaminen, N. & Nopola-Hemmi, J. *et al.* A candidate gene for developmental dyslexia encodes a nuclear tetratricopeptide repeat domain protein dynamically regulated in brain. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 11553–11558, doi:10.1073/pnas.1833911100 (2003).
- Hofmeister, W., Nilsson, D. & Topa, A. *et al.* CTNND2—a candidate gene for reading problems and mild intellectual disability. *Journal of medical genetics* **52**, 111–122, doi:10.1136/jmedgenet-2014-102757 (2015).
- Anthoni, H., Zucchelli, M. & Matsson, H. *et al.* A locus on 2p12 containing the co-regulated MRPL19 and C2ORF3 genes is associated to dyslexia. *Human molecular genetics* **16**, 667–677, doi:10.1093/hmg/ddm009 (2007).
- Schumacher, J., Anthoni, H. & Dahdouh, F. *et al.* Strong genetic evidence of DCDC2 as a susceptibility gene for dyslexia. *American journal of human genetics* **78**, 52–62, doi:10.1086/498992 (2006).
- Kere, J. The molecular genetics and neurobiology of developmental dyslexia as model of a complex phenotype. *Biochemical and biophysical research communications* **452**, 236–243, doi:10.1016/j.bbrc.2014.07.102 (2014).
- Pagnamenta, A. T., Bacchelli, E. & de Jonge, M. V. *et al.* Characterization of a family with rare deletions in CNTNAP5 and DOCK4 suggests novel risk loci for autism and dyslexia. *Biological psychiatry* **68**, 320–328, doi:10.1016/j.biopsych.2010.02.002 (2010).
- Konig, I. R., Schumacher, J. & Hoffmann, P. *et al.* Mapping for dyslexia and related cognitive trait loci provides strong evidence for further risk genes on chromosome 6p21. *Am J Med Genet B Neuropsychiatr Genet* **156B**, 36–43, doi:10.1002/ajmg.b.31135 (2011).
- MacDermot, K. D., Bonora, E. & Sykes, N. *et al.* Identification of FOXP2 truncation as a novel cause of developmental speech and language deficits. *American journal of human genetics* **76**, 1074–1080, doi:10.1086/430841 (2005).
- Becker, J., Czamara, D. & Scerri, T. S. *et al.* Genetic analysis of dyslexia candidate genes in the European cross-linguistic NeuroDys cohort. *European journal of human genetics* **22**, 675–680, doi:10.1038/ejhg.2013.199 (2014).
- Einarsdottir, E., Svensson, I. & Darki, F. *et al.* Mutation in CEP63 co-segregating with developmental dyslexia in a Swedish family. *Human genetics* **134**, 1239–1248, doi:10.1007/s00439-015-1602-1 (2015).
- Mascheretti, S., De Luca, A. & Trezzi, V. *et al.* Neurogenetics of developmental dyslexia: from genes to behavior through brain neuroimaging and cognitive and sensorial mechanisms. *Translational psychiatry* **7**, e987, doi:10.1038/tp.2016.240 (2017).
- Beaulieu, C., Plewes, C. & Paulson, L. A. *et al.* Imaging brain connectivity in children with diverse reading ability. *NeuroImage* **25**, 1266–1271, doi:10.1016/j.neuroimage.2004.12.053 (2005).
- Deutsch, G. K. *et al.* Children's reading performance is correlated with white matter structure measured by diffusion tensor imaging. *Cortex; a journal devoted to the study of the nervous system and behavior* **41**, 354–363 (2005).
- Niogi, S. N. & McCandliss, B. D. Left lateralized white matter microstructure accounts for individual differences in reading ability and disability. *Neuropsychologia* **44**, 2178–2188, doi:10.1016/j.neuropsychologia.2006.01.011 (2006).
- Eicher, J. D., Powers, N. R. & Miller, L. L. *et al.* Genome-wide association study of shared components of reading disability and language impairment. *Genes Brain Behav* **12**, 792–801, doi:10.1111/gbb.12085 (2013).
- Filley, C. M. & Fields, R. D. White matter and cognition: making the connection. *Journal of neurophysiology* **116**, 2093–2104, doi:10.1152/jn.00221.2016 (2016).
- Klingberg, T., Hedehus, M. & Temple, E. *et al.* Microstructure of temporo-parietal white matter as a basis for reading ability: evidence from diffusion tensor magnetic resonance imaging. *Neuron* **25**, 493–500 (2000).

21. Darki, F., Peyrard-Janvid, M., Matsson, H., Kere, J. & Klingberg, T. Three dyslexia susceptibility genes, DYX1C1, DCDC2, and KIAA0319, affect temporo-parietal white matter structure. *Biological psychiatry* **72**, 671–676, doi:10.1016/j.biopsych.2012.05.008 (2012).
22. Darki, F., Peyrard-Janvid, M., Matsson, H., Kere, J. & Klingberg, T. DCDC2 polymorphism is associated with left temporoparietal gray and white matter structures during development. *The Journal of neuroscience: the official journal of the Society for Neuroscience* **34**, 14455–14462, doi:10.1523/JNEUROSCI.1216-14.2014 (2014).
23. Meda, S. A., Gelernter, J. & Gruen, J. R. *et al.* Polymorphism of DCDC2 Reveals Differences in Cortical Morphology of Healthy Individuals-A Preliminary Voxel Based Morphometry Study. *Brain imaging and behavior* **2**, 21–26, doi:10.1007/s11682-007-9012-1 (2008).
24. Skeide, M. A., Kirsten, H. & Kraft, I. *et al.* Genetic dyslexia risk variant is related to neural connectivity patterns underlying phonological awareness in children. *NeuroImage* **118**, 414–421, doi:10.1016/j.neuroimage.2015.06.024 (2015).
25. Skeide, M. A., Kraft, I. & Muller, B. *et al.* NRSN1 associated grey matter volume of the visual word form area reveals dyslexia before school. *Brain: a journal of neurology* **139**, 2792–2803, doi:10.1093/brain/aww153 (2016).
26. Dehaene, S., Cohen, L., Morais, J. & Kolinsky, R. Illiterate to literate: behavioural and cerebral changes induced by reading acquisition. *Nature reviews. Neuroscience* **16**, 234–244, doi:10.1038/nrn3924 (2015).
27. Nopola-Hemmi, J., Myllyluoma, B. & Haltia, T. *et al.* A dominant gene for developmental dyslexia on chromosome 3. *Journal of medical genetics* **38**, 658–664 (2001).
28. Nopola-Hemmi, J. *et al.* Familial dyslexia: neurocognitive and genetic correlation in a large Finnish family. *Dev Med Child Neurol* **44**, 580–586 (2002).
29. Häyrynen T., S.-S. S., Korkman M. Reading and writing test designed for and normated in Finnish elementary school. In Finnish. (Psykologien kustannus Oy, 1999).
30. Leinonen, S., Leppänen, M. K., Aro, P., Ahonen, M. & Lyytinen, T. H. Heterogeneity in adult dyslexic readers: relating processing skills to the speed and accuracy of oral text reading. *Read. Writ. Interdisc. J.* **14**, 265–296 (2001).
31. Korkman, M. NEPS-U. Lasten neuropsykologinen tutkimus. Uudistettu versio., (Psykologien Kustannus Oy., 1988).
32. Abecasis, G. R., Cherny, S. S., Cookson, W. O. & Cardon, L. R. Merlin—rapid analysis of dense genetic maps using sparse gene flow trees. *Nature genetics* **30**, 97–101, doi:10.1038/ng786 (2002).
33. Genomes Project, C., Abecasis, G. R. & Auton, A. *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65, doi:10.1038/nature11632 (2012).
34. Andersson, R., Gebhard, C. & Miguel-Escalada, I. *et al.* An atlas of active enhancers across human cell types and tissues. *Nature* **507**, 455–461, doi:10.1038/nature12787 (2014).
35. Raum, H., Dietsche, B. & Nagels, A. *et al.* A genome-wide supported psychiatric risk variant in NCAN influences brain function and cognitive performance in healthy subjects. *Human brain mapping* **36**, 378–390, doi:10.1002/hbm.22635 (2015).
36. Dannlowski, U., Kugel, H. & Grotegerd, D. *et al.* NCAN Cross-Disorder Risk Variant Is Associated With Limbic Gray Matter Deficits in Healthy Subjects and Major Depression. *Neuropsychopharmacology: official publication of the American College of Neuropsychopharmacology* **40**, 2510–2516, doi:10.1038/npp.2015.86 (2015).
37. Scerri, T. S., Darki, F. & Newbury, D. F. *et al.* The dyslexia candidate locus on 2p12 is associated with general cognitive ability and white matter structure. *PLoS one* **7**, e50321, doi:10.1371/journal.pone.0050321 (2012).
38. Ashburner, J. A fast diffeomorphic image registration algorithm. *NeuroImage* **38**, 95–113, doi:10.1016/j.neuroimage.2007.07.007 (2007).
39. Hayasaka, S., Phan, K. L., Liberzon, I., Worsley, K. J. & Nichols, T. E. Nonstationary cluster-size inference with random field and permutation methods. *NeuroImage* **22**, 676–687, doi:10.1016/j.neuroimage.2004.01.041 (2004).
40. Kaminen, N., Hannula-Jouppi, K. & Kestila, M. *et al.* A genome scan for developmental dyslexia confirms linkage to chromosome 2p11 and suggests a new locus on 7q32. *Journal of medical genetics* **40**, 340–345 (2003).
41. Jagadeesh, K. A., Wenger, A. M. & Berger, M. J. *et al.* M-CAP eliminates a majority of variants of uncertain significance in clinical exomes at high sensitivity. *Nature genetics* **48**, 1581–1586, doi:10.1038/ng.3703 (2016).
42. Mori, S., Oishi, K. & Faria, A. V. White matter atlases based on diffusion tensor imaging. *Current opinion in neurology* **22**, 362–369, doi:10.1097/WCO.0b013e32832d954b (2009).
43. Xu, W., Cohen-Woods, S. & Chen, Q. *et al.* Genome-wide association study of bipolar disorder in Canadian and UK populations corroborates disease loci including SYNE1 and CSMD1. *BMC medical genetics* **15**, 2, doi:10.1186/1471-2350-15-2 (2014).
44. Havik, B., Le Hellard, S. & Rietschel, M. *et al.* The complement control-related genes CSMD1 and CSMD2 associate to schizophrenia. *Biological psychiatry* **70**, 35–42, doi:10.1016/j.biopsych.2011.01.030 (2011).
45. Donohoe, G., Walters, J. & Hargreaves, A. *et al.* Neuropsychological effects of the CSMD1 genome-wide associated schizophrenia risk variant rs10503253. *Genes Brain Behav* **12**, 203–209, doi:10.1111/gbb.12016 (2013).
46. Rose, E. J., Morris, D. W. & Hargreaves, A. *et al.* Neural effects of the CSMD1 genome-wide associated schizophrenia risk variant rs10503253. *Am J Med Genet B Neuropsychiatr Genet* **162B**, 530–537, doi:10.1002/ajmg.b.32182 (2013).
47. Koiliari, E., Roussos, P. & Pasparkis, E. *et al.* The CSMD1 genome-wide associated schizophrenia risk variant rs10503253 affects general cognitive ability and executive function in healthy males. *Schizophrenia research* **154**, 42–47, doi:10.1016/j.schres.2014.02.017 (2014).
48. Inatani, M., Honjo, M. & Otori, Y. *et al.* Inhibitory effects of neurocan and phosphacan on neurite outgrowth from retinal ganglion cells in culture. *Investigative ophthalmology & visual science* **42**, 1930–1938 (2001).
49. Cichon, S., Muhleisen, T. W. & Degenhardt, F. A. *et al.* Genome-wide association study identifies genetic variation in neurocan as a susceptibility factor for bipolar disorder. *American journal of human genetics* **88**, 372–381, doi:10.1016/j.ajhg.2011.01.017 (2011).
50. Muhleisen, T. W., Mattheisen, M. & Strohmaier, J. *et al.* Association between schizophrenia and common variation in neurocan (NCAN), a genetic risk factor for bipolar disorder. *Schizophrenia research* **138**, 69–73, doi:10.1016/j.schres.2012.03.007 (2012).
51. Schultz, C. C., Muhleisen, T. W. & Nenadic, I. *et al.* Common variation in NCAN, a risk factor for bipolar disorder and schizophrenia, influences local cortical folding in schizophrenia. *Psychological medicine* **44**, 811–820, doi:10.1017/S0033291713001414 (2014).
52. Miro, X., Meier, S. & Dreisow, M. L. *et al.* Studies in humans and mice implicate neurocan in the etiology of mania. *The American journal of psychiatry* **169**, 982–990, doi:10.1176/appi.ajp.2012.11101585 (2012).
53. Stefansson, H., Meyer-Lindenberg, A. & Steinberg, S. *et al.* CNVs conferring risk of autism or schizophrenia affect cognition in controls. *Nature* **505**, 361–366, doi:10.1038/nature12818 (2014).
54. Ulfarsson, M. O., Walters, G. B. & Gustafsson, O. *et al.* 15q11.2 CNV affects cognitive, structural and functional correlates of dyslexia and dyscalculia. *Translational psychiatry* **7**, e1109, doi:10.1038/tp.2017.77 (2017).
55. Richlan, F., Kronbichler, M. & Wimmer, H. Meta-analyzing brain dysfunctions in dyslexic children and adults. *NeuroImage* **56**, 1735–1742, doi:10.1016/j.neuroimage.2011.02.040 (2011).
56. Shaywitz, B. A., Shaywitz, S. E. & Blachman, B. A. *et al.* Development of left occipitotemporal systems for skilled reading in children after a phonologically-based intervention. *Biological psychiatry* **55**, 926–933, doi:10.1016/j.biopsych.2003.12.019 (2004).
57. Shaywitz, B. A., Shaywitz, S. E. & Pugh, K. R. *et al.* Disruption of posterior brain systems for reading in children with developmental dyslexia. *Biological psychiatry* **52**, 101–110 (2002).
58. Silani, G., Frith, U. & Demonet, J. F. *et al.* Brain abnormalities underlying altered activation in dyslexia: a voxel based morphometry study. *Brain: a journal of neurology* **128**, 2453–2461, doi:10.1093/brain/awh579 (2005).

59. Houde, O., Rossi, S., Lubin, A. & Joliot, M. Mapping numerical processing, reading, and executive functions in the developing brain: an fMRI meta-analysis of 52 studies including 842 children. *Developmental science* **13**, 876–885, doi:10.1111/j.1467-7687.2009.00938.x (2010).
60. Nagy, Z., Westerberg, H. & Klingberg, T. Maturation of white matter is associated with the development of cognitive functions during childhood. *Journal of cognitive neuroscience* **16**, 1227–1233, doi:10.1162/0898929041920441 (2004).
61. Eklund, A., Nichols, T. E. & Knutsson, H. Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. *Proceedings of the National Academy of Sciences of the United States of America* **113**, 7900–7905, doi:10.1073/pnas.1602413113 (2016).
62. Button, L., Peter, B., Stoel-Gammon, C. & Raskind, W. H. Associations among measures of sequential processing in motor and linguistics tasks in adults with and without a family history of childhood apraxia of speech: a replication study. *Clin Linguist Phon* **27**, 192–212, doi:10.3109/02699206.2012.744097 (2013).

Acknowledgements

The authors thank Jingwen Wang for help in bioinformatic analysis of sequence data, Ingegerd Fransson and Auli Saarinen for expert technical assistance, and Ripley Högnadóttir for critical comments. Genotyping of the DD pedigree was performed by the SNP&SEQ Technology Platform in Uppsala. The platform is part of Science for Life Laboratory at Uppsala University and supported as a national infrastructure by the Swedish Research Council. Tomas Axelsson is especially thanked for expert project management. Genotyping within the FinnBrain study was performed at the Estonian Genome Center, Tartu, Estonia. Prof. Andres Metspalu and Dr. Lili Milani are thanked for their help with the genotyping, and Auli Toivola for her project management. The authors would like to acknowledge support of the National Genomics Infrastructure (NGI) hosted by SciLifeLab Stockholm/Uppsala Genome Center (especially Adam Ameur) and the Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX) for providing assistance in massive parallel sequencing and computational infrastructure (project b2014023). Work performed at NGI/Uppsala Genome Center has been funded by RFI/VR, the Knut and Alice Wallenberg Foundation and Science for Life Laboratory, Sweden. This study was supported by grants from the Swedish Research Council, Swedish Brain Foundation (“Hjärnfonden”), the Knut and Alice Wallenberg Foundation, the Bank of Sweden Tercentenary Foundation (“Riksbankens Jubileumsfond”), Academy of Finland, and the Sigrid Jusélius Foundation. The FinnBrain study was supported by the Academy of Finland (grants number 134950, 253270 and 264363 to HK and 135094 and 253346 to TP, www.aka.fi/skidi-kids), the NARSAD Brain and Behavior Foundation Young Investigator Grant (#19056) to LK and ERVA state grants of the Hospital District of Southwestern Finland to JT and NS as well as a Gyllenberg Foundation grant to NS. JK is a recipient of The Royal Society Wolfson Research Excellence Award. WES and WGS sequence data are available in the DDBJ/EMBL/GenBank databases under the accession number PRJEB12695.

Author Contributions

E.E. Planned the study, performed the genetic analysis and wrote the manuscript. M.P.-J. contributed to the genetic analysis and also to writing of the manuscript. H.Matsson. contributed to the genetic analysis and edited the manuscript. F.D. and T.K. performed the Brainchild imaging analysis. J.J.T. performed the FinnBrain imaging analysis and contributed to writing the manuscript. L.K. and H.K. supervised the FinnBrain imaging and study design. N.M.S., J.S., R.P. and H.Merisaari contributed to the FinnBrain imaging and analyses. K.K. and A.-J.Ä. performed the genetic analysis for the FinnBrain study. N.Y.-L.Y. Performed gene expression correlation analysis J.N.-H. Diagnosed the family members. T.P. Supervised the genetic analysis for the FinnBrain study. E.L. Coordinated the study and organized samples for the genetic analysis. J.K. Planned and coordinated the study, supervised the genetic analyses, and edited the manuscript.

Additional Information

Supplementary information accompanies this paper at doi:10.1038/s41598-017-10175-7

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017