

SHORT COMMUNICATION

Polymorphisms in *DCDC2* and *S100B* associate with developmental dyslexia

Hans Matsson¹, Mikael Huss^{2,3}, Helena Persson¹, Elisabet Einarsdottir¹, Ettore Tiraboschi¹, Jaana Nopola-Hemmi⁴, Johannes Schumacher⁵, Nina Neuhoﬀ⁶, Andreas Warnke⁷, Heikki Lyytinen⁸, Gert Schulte-Körne⁶, Markus M Nöthen⁹, Paavo HT Leppänen¹⁰, Myriam Peyrard-Janvid¹ and Juha Kere^{1,2,11,12}

Genetic studies of complex traits have become increasingly successful as progress is made in next-generation sequencing. We aimed at discovering single nucleotide variation present in known and new candidate genes for developmental dyslexia: *CYP19A1*, *DCDC2*, *DIP2A*, *DYX1C1*, *GCFC2* (also known as *C2orf3*), *KIAA0319*, *MRPL19*, *PCNT*, *PRMT2*, *ROBO1* and *S100B*. We used next-generation sequencing to identify single-nucleotide polymorphisms in the exons of these 11 genes in pools of 100 DNA samples of Finnish individuals with developmental dyslexia. Subsequent individual genotyping of those 100 individuals, and additional cases and controls from the Finnish and German populations, validated 92 out of 111 different single-nucleotide variants. A nonsynonymous polymorphism in *DCDC2* (corrected $P = 0.002$) and a noncoding variant in *S100B* (corrected $P = 0.016$) showed a significant association with spelling performance in families of German origin. No significant association was found for the variants neither in the Finnish case-control sample set nor in the Finnish family sample set. Our findings further strengthen the role of *DCDC2* and implicate *S100B*, in the biology of reading and spelling.

Journal of Human Genetics (2015) 60, 399–401; doi:10.1038/jhg.2015.37; published online 16 April 2015

INTRODUCTION

Developmental dyslexia (DD) is characterized by deficiencies in reading and writing with phonological difficulties persisting throughout life.¹ Previous studies have demonstrated genetic components for DD with at least nine loci (*DYX1-9*) identified, and perturbed gene expression or human loss of function phenotypes have revealed roles for candidate genes in neuronal migration, axon guidance and more recently in ciliary biogenesis and function.^{2–6} Recently, a study with over 900 cases analyzed the effect of common polymorphisms on DD, word reading and spelling using samples from eight European countries.⁷ Nominal association was found for variants in single populations; however, no variant reached a significant association in meta-analysis of all samples. These results suggest the need for larger sample sizes for association analyses of common variants. We hypothesized that susceptibility genes can harbour unreported single-nucleotide polymorphisms (SNPs) and novel variants in significant association with DD. Consequently, we embarked on an alternative approach using next-generation sequencing (NGS) of *CYP19A1*, *DCDC2*, *DIP2A*, *DYX1C1*, *GCFC2* (also known as *C2orf3*), *KIAA0319*, *MRPL19*, *PCNT*, *PRMT2*, *ROBO1* and *S100B* in 100 unrelated DD

cases from Finland. The first seven genes were selected because they harbour variants with replicated genetic association; *PCNT*, *DIP2A*, *S100B* and *PRMT2* were selected as they are contained within a deletion on chromosome 21q22.3 that co-segregates with DD in a Dutch family.⁸

MATERIALS AND METHODS

Ethical permissions were obtained from the regional committees in Helsinki, Finland (412/E9/04), Marburg and Würzburg, Germany (115/00), and Stockholm, Sweden (2008/153-32).

DNA from 100 DD cases (discovery sample) was divided into 10 pools á 10 individuals and the regions of interest (ROI) were captured using selector probes (Halo Genomics, Uppsala, Sweden). The ROI included all RefSeq (release 41) annotated exons as well as 3'-UTRs, 5'-UTRs and 50 base pair intronic sequence flanking exons of the 11 target genes. Captured sequences were sequenced on SOLID4 and 5500xl instruments (Life Technologies, Carlsbad, CA, USA). We chose the CRISP algorithm⁹ for the subsequent variant prediction and the parameters were set to allow detection of heterozygous variants from single individuals in each pool. For detailed material and method descriptions, see Supplementary Materials and Methods.

¹Department of Biosciences and Nutrition, and Center for Innovative Medicine, Karolinska Institutet, Huddinge, Sweden; ²Science for Life Laboratory, Stockholm, Sweden; ³Department of Biochemistry and Biophysics, Stockholm University, Stockholm, Sweden; ⁴Division of Child Neurology, Department of Gynecology and Pediatrics, Jorvi Hospital, University Central Hospital, Espoo, Finland; ⁵Institute of Human Genetics and Department of Genomics, Life and Brain Center, University of Bonn, Bonn, Germany; ⁶Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, University of Munich, Munich, Germany; ⁷Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, University of Würzburg, Würzburg, Germany; ⁸Department of Psychology and Child Research Center, University of Jyväskylä, Jyväskylä, Finland; ⁹Department of Genomics, Life & Brain Center, University of Bonn, Bonn, Germany; ¹⁰Department of Psychology, University of Jyväskylä, Jyväskylä, Finland; ¹¹Molecular Neurology Research Program, Research Programs Unit, University of Helsinki, Helsinki, Finland and ¹²Folkhälsan Institute of Genetics, Helsinki, Finland
Correspondence: Dr H Matsson, Department of Biosciences and Nutrition, and Center for Innovative Medicine, Karolinska Institutet, Hälsovägen 7, SE-141 83 Huddinge, Sweden. E-mail: Hans.Matsson@ki.se

RESULTS AND DISCUSSION

We identified 282 SNPs and novel variants in the ROI by SOLiD4 sequencing (Table 1). We also sequenced available DNA pools (9 out of 10) on a 5500xl platform to increase the total sequencing coverage and depth, and 291 SNPs and novel variants were detected (Table 1). A total of 236 variants were detected in common by both sequencing methods (Table 1). For validation of individual variants by genotyping, we selected all (i) coding variants; (ii) novel variants neither present in dbSNP130 nor the 1000 Genomes database (Pilots 1,2,3 release March 2010); (iii) 3' and 5'-UTR SNPs and (iv) intronic SNPs predicted by both CRISP and FreeBayes algorithms, making a total of

111 variants. Out of those, we validated 92 and 89 variants in case-controls and families from Finland (including the discovery sample) as well as families from Germany, respectively (Supplementary Tables 1 and 2). The extended case-control sample set (169 affected and 194 unaffected) included Finnish case-controls combined with unrelated affected and unaffected individuals from families (Table 2). The German family sample (419 affected and 118 unaffected individuals) was stratified based on observed spelling ability (Table 2).

As we sequenced selected genes, we corrected transmission disequilibrium test (TDT) and allelic association *P*-values for the number of independent markers for each gene (Supplementary Materials and Methods). Neither TDT in Finnish families nor allelic association test in the extended case-control set revealed variants in significant association with DD. TDT in the German families yielded significant association (*P*-value = 0.0003; corrected *P* = 0.002, 7 independent markers) with spelling for the A allele of a non-synonymous variant (rs2274305, p.Ser221Gly) in *DCDC2* (Table 3) using the group with most severe spelling deficiency (≥ 2.5 s.d. below expected spelling score, 72 affected). This confirms and strengthens previous results.¹⁰ Power analyses¹¹ ($\alpha = 0.05$) show 98% power to detect association with the rs2274305 SNP in the total German sample and 59% in the 72 trios belonging to the ≥ 2.5 s.d. group. Rs2274305 is predicted to be possibly damaging (Polyphen2)¹² and is located within the doublecortin domain, a structure important for the neuronal migration function of *DCDC2*.¹³ Furthermore, we found a 3'-UTR variant (rs9722) in *S100B* in significant association with spelling for the groups defined as 2 s.d. (T allele, corrected *P* = 0.016, 1 marker, power = 69% at $\alpha = 0.05$) and 1.5 s.d. (T allele, corrected *P* = 0.016, 1 marker, power = 73% at $\alpha = 0.05$) below anticipated spelling score (Table 3). The association disappeared in the ≥ 2.5 s.d. group. This is likely an effect of reduced power because of the limited number of cases ($n = 72$). Sequence variants in the 3'-UTR may affect transcript stability, for example, by creating or disrupting target sites for microRNAs (miRNAs).^{14–16} We therefore searched for putative miRNA target sites overlapping rs9722 in TargetScan predictions.¹⁷ Indeed, rs9722 was located in, or adjacent to, multiple predicted miRNA target sites (Supplementary Table 2), a result that warrant further analyses. None of the five validated novel variants, distributed in *PCNT* and *DIP2A*, showed association with DD and their low number makes tests for excess of rare variants in cases versus controls, for example gene burden tests, unsuitable.

Tests combining case-controls and related individuals can increase power. To explore this option, we used genotypes of the 92 validated variants for the Finnish family (156 affected, 258 unaffected) and case-

Table 1 Total number of predicted SNVs after a targeted sequencing

Platform	Samples (pools)	Overlapping SNVs	Variants in ROI ^a	Validated SNVs (Finnish/German sample)
SOLiD4	10		282	92/89
5500xl	9 ^b		291	
SOLiD4 +5500xl	10+9 ^b	236	310	85 ^c

Abbreviations: ROI, regions of interest; SNV, single-nucleotide variants.

^aVariants predicted by the CRISP algorithm located in the captured ROI.

^b90% of the libraries were sequenced by using the 5500xl platform.

^cOverlapping validated SNVs from both sequencing runs (SOLiD4+5500xl) by using CRISP variant prediction.

Table 2 Sample sets used in the study

Sample name	Severity ^a	Affected	Unaffected	Unknown	Total
Discovery sample (Finland)		100 (1)	0	NA	100
Case-control (Finland)		92 (NA)	67 (NA)	NA	159
Extended case-control (Finland)		169 (1.2)	194 (0.9)	NA	363
Finnish families		156 (1.2)	258 (1)	0	414
German families total		419 (2.6)	118 (0.7)	511 (1)	1048
	2.5 s.d.	72 (5.5)	0	NA	72
	2.0 s.d.	171 (5.1)	0	NA	171
	1.5 s.d.	177 (5)	0	NA	177

Abbreviation: NA, not applicable.

Male:female ratio is presented in parenthesis.

^aSeverity cutoff for spelling performance in s.d. below the expected population average.

Table 3 Top ranking TDT results by using validated SNPs in the German family sample set

Severity/cases ^a	Chr	SNP	Position (bp)	A1	A2	T	U	OR	<i>P</i> ^b	Corrected <i>P</i> ^c	Position ^d	Gene
≥ 2.5 s.d./72	6	rs2274305	24399182	A	G	41	14	2.929	0.0003	0.002	p.Ser221Gly	<i>DCDC2</i>
	15	rs600753	53546485	T	C	24	42	0.571	0.0267	0.107	p.Glu191Gly	<i>DYX1C1</i>
2 s.d./171	21	rs9722	46843667	T	C	37	19	1.947	0.0162	0.016	3'-UTR	<i>S100B</i>
	6	rs2274305	24399182	A	G	85	59	1.441	0.0303	0.240	p.Ser221Gly	<i>DCDC2</i>
1.5 s.d./227	21	rs9722	46843667	T	C	46	25	1.84	0.0127	0.013	3'-UTR	<i>S100B</i>
	6	rs2274305	24399182	A	G	109	80	1.363	0.0349	0.280	p.Ser221Gly	<i>DCDC2</i>
Total/419	6	rs6937665	24280905	G	A	102	137	0.745	0.0236	0.189	3'-UTR	<i>DCDC2</i>
	15	rs2289105	49294800	A	G	195	155	1.258	0.0325	0.195	Intronic	<i>CYP19A1</i>

Abbreviations: Chr, chromosome; SNPs, single-nucleotide polypeptide; T, number of transmitted minor allele (A1); TDT, transmission disequilibrium test; U, number of untransmitted minor allele (A1); OR, odds ratio.

^aSeverity of spelling disability/number of cases

^bUncorrected *P*-value.

^c*P*-values are corrected for multiple testing (Bonferroni) for each gene separately and for the number of independent SNPs per gene.

^dPosition within the genomic structure of the gene or position of the concerned amino-acid.

control samples (92 affected, 67 unaffected) in MQLS (a more powerful quasi-likelihood score test) capable of case-control association testing with the related individuals.¹⁸ None of the variants showed a significant association with DD. It is plausible that the association in the families, at least in part, is driven by polymorphisms in other genes.^{19,20}

We conclude that sample sets based on related individuals with severe phenotypes can increase the likelihood to find genetic association. Our results strengthen a previous association signal in *DCDC2* as well as suggesting *S100B* as a new candidate gene for DD.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

This study was supported by grants from the Karolinska Institutet, Hjärnfonden, the Swedish Research Council, the Knut and Alice Wallenberg Foundation, the Bank of Sweden Tercentenary Foundation and the Deutsche Forschungsgemeinschaft (DFG). The authors acknowledge support from the Science for Life Laboratory, the national infrastructure SNISS, Uppmax for providing assistance in massive parallel sequencing and computational infrastructure as well as the Mutation Analysis Facility (MAF) at Karolinska Institutet, for service and support on genotyping. The aligned sequence reads from the next-generation sequencing are available in the DDBJ/EMBL/GenBank databases under the study accession number PRJEB4175. Individual genotype counts for all validated variants in all sample sets used were submitted to dbSNP.

- 1 Svensson, I. & Jacobson, C. How persistent are phonological difficulties? A longitudinal study of reading retarded children. *Dyslexia* **12**, 3–20 (2006).
- 2 Scerri, T. S. & Schulte-Körne, G. Genetics of developmental dyslexia. *Eur. Child Adolesc. Psychiatry* **19**, 179–197 (2010).
- 3 Skiba, T., Landi, N., Wagner, R. & Grigorenko, E. L. In search of the perfect phenotype: an analysis of linkage and association studies of reading and reading-related processes. *Behav. Genet.* **41**, 6–30 (2011).
- 4 Kere, J. The molecular genetics and neurobiology of developmental dyslexia as model of a complex phenotype. *Biochem. Biophys. Res. Commun.* **452**, 236–243 (2014).
- 5 Grati, M., Chakchouk, I., Ma, Q., Bensaid, M., Desmidt, A. & Turki, N. *et al.* A missense mutation in *DCDC2* causes human recessive deafness DFNB66, likely by interfering with sensory hair cell and supporting cell cilia length regulation. *Hum. Mol. Genet.* **24**, 2482–2491 (2015).
- 6 Schueler, M., Braun, D. A., Chandrasekar, G., Gee, H. Y., Klasson, T. D. & Halbritter, J. *et al.* *DCDC2* mutations cause a renal-hepatic ciliopathy by disrupting Wnt signaling. *Am. J. Hum. Genet.* **96**, 81–92 (2015).

- 7 Becker, J., Czamara, D., Scerri, T. S., Ramus, F., Csepe, V. & Talcott, J. B. *et al.* Genetic analysis of dyslexia candidate genes in the European cross-linguistic NeuroDys cohort. *Eur. J. Hum. Genet.* **22**, 675–680 (2014).
- 8 Poelmans, G., Engelen, J. J., Van Lent-Albrechts, J., Smeets, H. J., Schoenmakers, E. & Franke, B. *et al.* Identification of novel dyslexia candidate genes through the analysis of a chromosomal deletion. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **150B**, 140–147 (2009).
- 9 Bansal, V. A statistical method for the detection of variants from next-generation resequencing of DNA pools. *Bioinformatics* **26**, i318–i324 (2010).
- 10 Schumacher, J., Anthoni, H., Dahdouh, F., König, I. R., Hillmer, A. M. & Kluck, N. *et al.* Strong genetic evidence of *DCDC2* as a susceptibility gene for dyslexia. *Am. J. Hum. Genet.* **78**, 52–62 (2006).
- 11 Purcell, S., Cherny, S. S. & Sham, P. C. Genetic power calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics* **19**, 149–150 (2003).
- 12 Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A. & Bork, P. *et al.* A method and server for predicting damaging missense mutations. *Nat. Methods* **7**, 248–249 (2010).
- 13 des Portes, V., Pinard, J. M., Billuart, P., Vinet, M. C., Koulakoff, A. & Carrie, A. *et al.* A novel CNS gene required for neuronal migration and involved in X-linked subcortical laminar heterotopia and lissencephaly syndrome. *Cell* **92**, 51–61 (1998).
- 14 Clop, A., Marq, F., Takeda, H., Pirotin, D., Tordoir, X. & Bibe, B. *et al.* A mutation creating a potential illegitimate microRNA target site in the myostatin gene affects muscularity in sheep. *Nat. Genet.* **38**, 813–818 (2006).
- 15 Saunders, M. A., Liang, H. & Li, W. H. Human polymorphism at microRNAs and microRNA target sites. *Proc. Natl Acad. Sci. USA* **104**, 3300–3305 (2007).
- 16 Chin, L. J., Ratner, E., Leng, S., Zhai, R., Nallur, S. & Babar, I. *et al.* A SNP in a let-7 microRNA complementary site in the *KRAS* 3' untranslated region increases non-small cell lung cancer risk. *Cancer Res.* **68**, 8535–8540 (2008).
- 17 Grimson, A., Farh, K. K., Johnston, W. K., Garrett-Engle, P., Lim, L. P. & Bartel, D. P. MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol. Cell.* **27**, 91–105 (2007).
- 18 Thornton, T. & McPeck, M. S. Case-control association testing with related individuals: a more powerful quasi-likelihood score test. *Am. J. Hum. Genet.* **81**, 321–337 (2007).
- 19 Anthoni, H., Zucchelli, M., Matsson, H., Müller-Myhök, B., Fransson, I. & Schumacher, J. *et al.* A locus on 2p12 containing the co-regulated *MRPL19* and *C2ORF3* genes is associated to dyslexia. *Hum. Mol. Genet.* **16**, 667–677 (2007).
- 20 Matsson, H., Tammimies, K., Zucchelli, M., Anthoni, H., Onkamo, P. & Nopola-Hemmi, J. *et al.* SNP variations in the 7q33 region containing *DGKI* are associated with dyslexia in the Finnish and German populations. *Behav. Genet.* **41**, 134–140 (2011).



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>

Supplementary Information accompanies the paper on Journal of Human Genetics website (<http://www.nature.com/jhg>)