Polymorphisms in DCDC2 and S100B associate with developmental dyslexia

Matsson, Hans; Huss, Mikael; Persson, Helena; Einarsdottir, Elisabet; Tiraboschi, Ettore; Nopola-Hemmi, Jaana; Schumacher, Johannes; Neuhoff, Nina; Warnke, Andreas; Lyytinen, Heikki; Schulte-Körne, Gert; Nöthen, Markus M.; Leppänen, Paavo H.T.; Peyrard-Janvid, Myriam; Kere, Juha

Title: Polymorphisms in DCDC2 and S100B associate with developmental dyslexia

Year: 2015

Version:

Please cite the original version:
Matsson, H., Huss, M., Persson, H., Einarsdottir, E., Tiraboschi, E., Nopola-Hemmi, J., Schumacher, J., Neuhoff, N., Warnke, A., Lyytinen, H., Schulte-Körne, G., Nöthen, M. M., Leppänen, P. H., Peyrard-Janvid, M., & Kere, J. (2015). Polymorphisms in DCDC2 and S100B associate with developmental dyslexia. Journal of Human Genetics, 60(7), 399-401. https://doi.org/10.1038/jhg.2015.37

All material supplied via JYX is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.
SHORT COMMUNICATION

Polymorphisms in DCDC2 and S100B associate with developmental dyslexia

Hans Matsson1, Mikael Huss2,3, Helena Persson1, Elisabet Einarsdottir1, Ettore Tiraboschi1, Jaana Nopola-Hemmi4, Johannes Schumacher5, Nina Neuhoff6, Andreas Warnke7, Heikki Lytinen8, Gert Schulte-Körne9, Markus M Nöthen10, Myriam Peyrard-Janvid1 and Juha Kere1,2,11,12

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. Previous studies have demonstrated genetic components for DD with at least nine loci (DYXI-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.1–4 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.5 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 hypothesised 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/E09/04), 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 algorithm 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.
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); (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-control samples from Finland (including the discovery sample) and 100 families from Germany (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-control samples from Finland (including the discovery sample) and 100 families from Germany (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-control samples from Finland (including the discovery sample) and 100 families from Germany (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).
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. 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.

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, Hjärtfonden, 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 SNSS, 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-Korne, 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., Chalichouk, 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 Schuerer, M., Braun, D. A., Chandrasekar, G., Gee, H. Y., Klasson, T. D. & Halfvanger, 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., Czoernia, D., Scerri, T. S., Ramus, F., Coope, V. & Talcott, J. B. et al. Genetic analysis of dyslexia candidate genes in the European cross-linguistic Neuropysychiatric. Eur. J. Hum. Genet. 22, 675–680 (2014).
8 Poelmans, G., Engelen, J. J., Van Lent-Abbrecht, 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, 1318–1324 (2010).
10 Schumacher, J., Anthoni, H., Dahdouh, F., Konig, 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., Cheny, 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, E. V., Gunssovina, 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. & Carie, A. et al. A novel CNS gene required for neuronal migration and involved in X-linked subcortical lamination heterotopia and lissencephaly syndrome. Cell 92, 51–61 (1998).
14 Clop, A., Marcq, F., Takeda, H., Pirrotta, D., Tordoir, X. & Biile, 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-Engele, 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. & McPeek, 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., Zuccelli, M., Matsson, H., Muller-Mykhhok, B., Fransson, I. & Schumacher, J. et al. A locus on 2p12 containing the co-regulated MRPL19 and C2OORF3 genes is associated to dyslexia. Hum. Mol. Genet. 16, 667–677 (2007).
20 Matsson, H., Tamminen, K., Zuccelli, 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).