FURTHER EVIDENCE FOR \textit{DYX1C1} AS A SUSCEPTIBILITY FACTOR FOR DYSLEXIA

Faten Dahdouh\textsuperscript{1}, Heidi Anthoni\textsuperscript{2}, Isabel Tapia-Páez\textsuperscript{2}, Myriam Peyrard-Janvid\textsuperscript{2}, Gerd Schulte-Körne\textsuperscript{3}, Andreas Warnke\textsuperscript{4}, Helmut Remschmidt\textsuperscript{5}, Andreas Ziegler\textsuperscript{6}, Juha Kere\textsuperscript{2,7,8}, Bertram Müller-Myhsok\textsuperscript{9}, Markus M. Nöthen\textsuperscript{10}, Johannes Schumacher\textsuperscript{1,11}, Marco Zucchelli\textsuperscript{2}

\textsuperscript{1}Institute of Human Genetics, University of Bonn, Bonn, Germany
\textsuperscript{2}Department of Biosciences and Nutrition, Karolinska Institutet, Huddinge, Sweden
\textsuperscript{3}Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, University of Munich, Munich, Germany
\textsuperscript{4}Department of Child and Adolescent Psychiatry and Psychotherapy, University of Würzburg, Würzburg, Germany
\textsuperscript{5}Department of Child and Adolescent Psychiatry and Psychotherapy, University of Marburg, Marburg, Germany
\textsuperscript{6}Institute for Medical Biometry and Statistics, University of Lübeck, Lübeck, Germany
\textsuperscript{7}Department of Medical Genetics, University of Helsinki, Helsinki, Finland
\textsuperscript{8}Department of Clinical Research Center, Karolinska Institutet, Huddinge, Sweden
\textsuperscript{9}Max Planck Institute of Psychiatry, Munich, Germany
\textsuperscript{10}Department of Genomics, Life & Brain Center, University of Bonn, Bonn, Germany
\textsuperscript{11}Unit on the Genetic Basis of Mood and Anxiety Disorders, National Institute of Mental Health, National Institutes of Health, Bethesda, USA

Running Title
\textit{DYX1C1} in dyslexia

Corresponding author
Professor Juha Kere
Department of Biosciences and Nutrition, Karolinska Institutet, 14157 Huddinge, Sweden
Phone: +46 8 6089158, Fax: +46 8 7745 538
Email: Juha.Kere@biosci.ki.se

Keywords
Candidate gene, developmental genes, behavioral genetics, identification of disease genes
Abstract

DYX1C1 was the first gene associated to Dyslexia. Since the original report of 2003, eight replication attempts have been published reporting discordant results. As the dyslexia community still considers the role of DYX1C1 unsettled, we explored the contribution of this gene in a sample of 366 trios of German descent. To the common four markers used in previous studies, we added two new SNPs found by re-sequencing both the putative regulatory and coding region of the gene in randomly selected cases and controls. As linkage disequilibrium blocks of the region were not easy to define, we approached the association problem by running a transmission disequilibrium test over sliding windows of dimension 1 to 6 on consecutive markers. The significance of this test was calculated generating the empirical distribution of the global p-value by simulating the data. Since our study sample had a large female proband content, we also stratified our analysis by sex. We found statistically significant association with global corrected p-value of 0.036. The three markers haplotype G/G/G spanning rs3743205/rs3743204/rs600753 was the most associated with a p-value = 0.006 and OR 3.7 (95% CL = 1.4-9.6) in female probands. A detailed haplotype-phenotype analysis revealed that the dyslexia sub-phenotype short term memory (STM) contributed mainly to the observed findings. This is in accordance with a recent STM-DYX1C1 association in an independent sample of dyslexia. As significant association was proved in our sample, we could also conclude that denser maps, gender information and well defined sub-phenotypes are crucial to correctly determine the contribution of DYX1C1 to dyslexia.
INTRODUCTION
Dyslexia (MIM#12770) affects 5-12% of school-age children and is one of the most common neurobehavioral disorders [1]. In recent years, linkage studies have identified regions likely to harbor genes contributing to dyslexia [2]. In particular, nine chromosomal regions – DYX1 through DYX9 – have been reported in dyslexia and are listed by the Human Gene Nomenclature Committee (HGNC). Close to one of these loci, the gene dyslexia-susceptibility-1-candidate-1 (DYX1C1, MIM 608706) on chromosome 15q21 was identified as a susceptibility factor for dyslexia. A chromosomal translocation t(2;15)(q11;q21) causing a disruption of DYX1C1 cosegregated with dyslexia in a two-generation family, and two genetic variants at the DYX1C1 locus (rs3743205 and 1249G>T; submitted to dbSNP and designated rs61761345) showed significant association with dyslexia in an independent sample of Finnish origin [3]. Positive association findings at the DYX1C1 locus were subsequently reported by three independent groups. Wigg et al. (2004) studied both markers analyzed by Taipale et al. (2003) and found significant association in a dyslexia sample of European-Canadian descent. Marino et al. (2007) observed in an Italian dyslexia sample association at both variants using short term memory (STM) as phenotypic trait. It is well established that STM plays an important role in the development of dyslexia [4]. However, no significant association was observed in this sample using a categorical diagnosis of dyslexia [5]. The authors claimed that this discrepancy could be due to the fact that in their dyslexia study unaffected siblings were not included, reducing the overall power of the association test. Most recently, Bates et al. (2007) reported on significant association between DYX1C1 markers and dyslexia in a 789 families comprising twin series from Australia. However, the reported association evidence is complicated by the fact that different alleles and/or haplotypes were associated across studies. Along this line, although Scerri et al. (2004) found significant association at the two-marker haplotype rs3743205-rs61761345 in a UK dyslexia sample, they interpreted their results as negative replication. Other three studies failed to find any DYX1C1 association in their dyslexia samples which were mainly of European descent [6-8].

In the light of the cited studies, the role of DYX1C1 in dyslexia is thus considered unsettled [9]. In the present study, we aimed to explore the contribution of DYX1C1 to dyslexia in the German population and analyzed a large family-based sample of 366 trios. In a first step, we looked for association using the most significant associated SNPs reported by Taipale et al. (2003). Although we failed to detect association using these two markers we extended our analysis employing a combination of re-sequencing the entire susceptibility locus and further SNP genotyping for a better DYX1C1 marker coverage.

MATERIALS AND METHODS
Subjects
A total of 366 family trios of German descent (300 male and 66 female indices, see Supplementary Table I) were used for this study, which represent a part of a large ongoing national recruitment effort for families with dyslexia [10, 11]. All individuals, and in case of children younger than 14 years their parents, gave written informed consent to participation in the study. The families were recruited from the Departments of Child and Adolescent Psychiatry and Psychotherapy at the Universities of Marburg and Würzburg, and the study was approved by the local ethics committees.

The diagnostic inclusion criteria and phenotypic measures have been described in detail previously [10-14]. Briefly, the diagnosis of dyslexia was based on the spelling score using the T
distribution of the general population. To be diagnosed as dyslexic, the child had to meet the following discrepancy criterion: based on the correlation between IQ and spelling of 0.4 [12], an anticipated spelling score was calculated. The child was classified as dyslexic if the discrepancy between the anticipated and the observed spelling scores was at least one standard deviation (1 SD). Probands and all siblings fulfilling the inclusion criteria were assessed with several psychometric tests. These tests targeted different aspects of the dyslexia, i.e., word reading, phonological awareness, phonological decoding, rapid naming, STM and orthographic coding (see supplementary information for a more detailed description). Since it has been previously found that association findings might become stronger in samples of severely affected individuals [11], we also stratified our data by the severity of the phenotype (see supplementary information and Schumacher et al. (2006) for the definition of disease severity). As affected probands were defined by discrepancy to the observed spelling score of at least \( \geq 1 \) SD, two subgroups of families of more-severely affected children were identified by probands discrepancy of at least \( \geq 2 \) and \( \geq 2.5 \) SD, respectively.

**SNP genotyping**

SNP genotyping was carried out using matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry (Sequenom). Extension products were analyzed by a Mass ARRAY mass spectrometer (Bruker Daltonik) and peaks were identified using the SpectroTYPER software (Sequenom). All genotypes were scored independently by two individuals blind to the disease status, and tested for Mendelian inheritance using PedCheck. Genotype success ratio (GSR) was >0.80 for all analyzed SNPs. Detailed information on primer sequences, PCR amplification, genotyping procedure and genotype calling can be obtained on request.

**DNA re-sequencing**

All PCR products were cleaned from unincorporated primers and dNTPs using shrimp alkaline phosphatase and exonuclease I, and further sequenced using a DYEnamic ET Dye terminator kit (Amersham Biosciences). Sequencing products were electrophoresed using a MegaBACE 1000 instrument and MegaBACE long read matrix; visualized using the Sequence Analyzer v3.0 software (Amersham Biosciences) and further aligned using the Pregap and Gap4 software (Staden package; staden.sourceforge.net). In addition, a separate viewer compared each FASTA output from sequencing results to corresponding genomic sequences (NT_010194) using Blast 2 sequences. Detailed information on PCR amplification, sequencing and analysis is available on request.

**Statistical analysis**

Association between dyslexia and SNPs and haplotypes was assessed using the software PDTPHASE [15]. Because there is evidence for a gender-specific genetic influence on dyslexia [16, 17], tests were performed for all families as well as separately for families with a female and a male index patient, respectively. Haplotype analysis was performed using a sliding window approach, with window sizes ranging from two to six. For each sliding window, differences in the haplotype distribution between affected individuals and parents (global p-values) as well as transmission rates of each individual haplotype were analyzed. To assess the significance of our findings the empirical distribution of the test statistics was generated by using 10,000 pedigrees/genotypes simulations under the null hypothesis using SIMPED [18]. To adjust the nominal p-values for the fact of having tested male and female probands separately as well as jointly, a Westfall-Young permutation-based minimum-P step-down procedure was done based...
on the 10,000 permutations performed [19]. Finally, our testing strategy provided a global empirical association p-value for the whole sample based on 21 x 3 tests (6 single SNPS and 15 haplotype tests on respectively males, females and the joint sample). The analysis of quantitative phenotypes was performed using QPDPHASE, which is a quantitative trait implementation of the PDT. We used the UNPHASED implementation as for the PDT [15].

RESULTS
Given the previous inconsistent association findings, we aimed to explore the contribution of the DIX1C1 locus to dyslexia in a large family sample of German descent (366 trios). In order to increase the DIX1C1 marker coverage and the resolution of the locus-specific haplotypic structure, we re-sequenced the entire coding DIX1C1 region (10 exons, including the flanking sequences) as well as the corresponding 5’ and 3’ UTR regions in 10 patients and 10 controls of German descent, all randomly selected. This allows the detection of SNPs with a minor allele frequency (MAF) of >0.10 with a power of ≥98% [20]. Comparison of the analyzed gene region to the public sequence (NT_010194) revealed three previously unstudied SNPs within the DIX1C1 promoter region – rs12899331, rs16787, rs8043269. Based on the re-sequencing results and on information from the public dbSNP database we then selected for further genotyping the three identified promoter variants, two additional SNP markers located within the coding gene region, rs3743204 and rs600753 and the two SNPs originally found associated in dyslexia, namely rs3743205 and rs61761345 (Fig. 1). Of these variants rs8043269 was discarded as it failed our assay design criteria.
Fig. 1. Genomic structure of the *DYX1C1* gene and relative positions of the genotyped SNPs.

The linkage disequilibrium structure of the gene is shown underneath.

By running our algorithm, we found significant association in our sample with global p-value = 0.036. This p-value is properly corrected for multiple testing and sex stratification.

A posteriori inspection of the individual sliding window tests provided the best association finding in the female’s group for the three-marker haplotype G/G/G at rs3743205/rs3743204/rs600753. This carried an odds ratio (OR) of 3.71 (95% CI 1.44-9.60, p-value = 0.006). The frequency of this common haplotype was estimated to be 0.49 in female patients (Table I). A replication of the whole analysis over the severity stratified sample did not provide any improvement of the association results.

Table 1. *DYX1C1* haplotypes associated to dyslexia. Global p-values for the best scoring SNP configurations and the relative best associating haplotype for all, male, and female probands, respectively. P-values are corrected for multiple testing. Asterisk denotes “any allele”.

<table>
<thead>
<tr>
<th>P-value</th>
<th>Haplotypes</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>Males</td>
</tr>
<tr>
<td>0.167</td>
<td>0.754</td>
</tr>
<tr>
<td>0.191</td>
<td>0.781</td>
</tr>
<tr>
<td>0.324</td>
<td>0.693</td>
</tr>
<tr>
<td>0.714</td>
<td>0.923</td>
</tr>
</tbody>
</table>
Further, we sought if the observed association was mainly driven by a particular sub-phenotype of dyslexia. The quantitative dyslexia sub-phenotypes spelling, word reading, phonological decoding, phonological awareness, orthographic processing, rapid naming, and short term memory were tested using QPDTPHASE. The same common three-marker haplotype G/G/G at rs3743205/rs3743204/rs600753, which was associated in our entire female sample, showed significant results by using STM as a phenotypic trait (global p-value of 0.011, haplotype-based p-value of 0.0046). All other sub-phenotypes showed no association to this particular haplotype in females and the haplotype-STM association was not significant in the entire or male samples (data not shown). This result is consistent with recently reported findings of association of dyslexia to STM [24].

DISCUSSION
In the present study we aimed to explore the contribution of \( \text{DYX1C1} \) to dyslexia in the German population. We re-sequenced the coding and putative regulatory gene region in randomly selected patients and controls and carried out association studies using genetic variants covering the \( \text{DYX1C1} \) locus in a large family-based sample. For the following reasons we believe that our results might be of interest for the dyslexia research community: Firstly, we identified a three-marker \( \text{DYX1C1} \) haplotype which was significantly associated to female dyslexia patients. This finding still remains significant after correction for multiple testing and was associated with an OR of 3.71. Although we are aware by the fact that our female trios represent only a small fraction of our entire dyslexia sample (66 females of 366 cases), our results indicate that consideration of gender can influence the outcome of association findings in dyslexia. To our knowledge this study represents the first molecular genetic study of dyslexia applying a gender-separated analysis, although differences in prevalence rates between females and males have been reported in epidemiological dyslexia studies [16], and gender-specific differences have been already reported for a variety of other complex traits in humans [16, 21]. Moreover, our findings may shed some light on the contradictory association results reported so far for \( \text{DYX1C1} \). As female to male probands ratios may vary largely across studies, the power of the association tests may also vary consistently.

Secondly, in our sample of female patients we observed a significant association of the dyslexia sub-phenotype STM by applying an exploratory haplotype-phenotype analysis. Although this finding needs further replication by independent studies, our results are in accordance with the observed STM-\( \text{DYX1C1} \) association in the dyslexia sample of Marino et al. (2007). Both studies point to STM as a dyslexia sub-phenotype which might be more directly influenced by the genetic variation at the \( \text{DYX1C1} \) locus than the disorder itself. Thirdly, our results together with the study of Bates et al. (2007) indicate that genotyping the SNP markers which showed association in the Finnish population [3], is insufficient in capturing the dyslexia risk haplotype in Central European populations. In both studies the detection of association was only possible by increasing the \( \text{DYX1C1} \) marker coverage. Fourthly, our results and the findings of other studies [22, 23] indicate that the putative \( \text{DYX1C1} \)-causing mutation in Central Europeans is located on the common haplotype – G/G – at rs3743205-rs61761345. This might point to a common ancestor in Central Europeans, whereas the results in the Finnish and the Italian populations point to independent \( \text{DYX1C1} \) mutation events [3, 24].
We are aware of the fact that the present study is hampered by several limitations, e.g., moderate number of female patients, exploratory analysis of dyslexia sub-phenotypes, and moderate $DYXIC1$ marker coverage. However, our findings are in line with five independent reports indicating a role for $DYXIC1$ in the development of dyslexia or dyslexia-related phenotypes [3, 22-25]. Based on this association evidence it seems premature to reject $DYXIC1$ from the list of potential candidate genes in dyslexia. Instead, it seems necessary to carry out more detailed $DYXIC1$-dyslexia association studies in future. Those analyses should include large dyslexia samples from different populations, the consideration of gender- and detailed sub-phenotype information as well as maximum marker coverage across the gene locus. The results of the forthcoming genome-wide association studies (e.g., www.neurodys.com) may allow the field to determine the relevance of $DYXIC1$ to dyslexia.
ACKNOWLEDGMENTS

We are grateful to all investigated individuals for their cooperation in this study. This study was supported by the Deutsche Forschungsgemeinschaft (DFG), MMN received support for this work from the Alfred Krupp von Bohlen und Halbach-Stiftung, JS is a Research Fellow of the NIH/DFG Research Career Transition Awards Program, MPJ is a recipient of a research position from Swedish Research Council, ITP is supported by a grant from Swedish Brain Foundation, HA is supported by the Centennial Foundation of Helsingin Sanomat, MZ is partly supported by the Bioinformatics and Expression Analysis Core Facility at Karolinska Institutet, and JK is supported by grants from Swedish Research Council, Academy of Finland, Sigrid Jusélius Foundation, and Päivikki and Sakari Sohlberg Foundation.

REFERENCES
