A locus on 2p12 containing the co-regulated *MRPL19* and *C2ORF3* genes is associated to dyslexia

Heidi Anthoni[1], Marco Zucchelli[1], Hans Matsson[1], Bertram Müller-Myhsok[3], Ingegerd Fransson[1], Johannes Schumacher[4], Satu Massinen[6], Päivi Onkamo[7], Andreas Warnke[8], Heide Griesemann[8], Per Hoffmann[5], Jaana Nopola-Hemmi[9], Heikki Lyytinen[10], Gerd Schulte-Körne[11], Juha Kere[1,2,6,*], Markus M. Nöthen[5] and Myriam Peyrard-Janvid[1]

[1]Department of Biosciences and Nutrition and [2]Department of Clinical Research Center, Karolinska Institutet, 14157 Huddinge, Sweden, [3]Max Planck Institute of Psychiatry, 80804 Munich, Germany, [4]Institute of Human Genetics and [5]Department of Genomics, Life and Brain Center, University of Bonn, 53127 Bonn, Germany, [6]Department of Medical Genetics and [7]Department of Biological and Environmental Sciences, University of Helsinki, 00014 Helsinki, Finland, [8]Department of Child and Adolescent Psychiatry and Psychotherapy, University of Würzburg, 97080 Würzburg, Germany, [9]Department of Pediatric Neurology, Jorvi Hospital, 02740 Espoo, Finland, [10]Department of Psychology, University of Jyväskylä, 40014 Jyväskylä, Finland and [11]Department of Child and Adolescent Psychiatry and Psychotherapy, University of Marburg, 35039 Marburg, Germany

DYX3*, a locus for dyslexia, resides on chromosome 2p11-p15. We have refined its location on 2p12 to a 157 kb region in two rounds of linkage disequilibrium (LD) mapping in a set of Finnish families. The observed association was replicated in an independent set of 251 German families. Two overlapping risk haplotypes spanning 16 kb were identified in both sample sets separately as well as in a joint analysis. In the German sample set, the odds ratio for the most significantly associated haplotype increased with dyslexia severity from 2.2 to 5.2. The risk haplotypes are located in an intergenic region between *FLJ13391* and *MRPL19/C2ORF3*. As no novel genes could be cloned from this region, we hypothesized that the risk haplotypes might affect long-distance regulatory elements and characterized the three known genes. *MRPL19* and *C2ORF3* are in strong LD and were highly co-expressed across a panel of tissues from regions of adult human brain. The expression of *MRPL19* and *C2ORF3*, but not *FLJ13391*, were also correlated with the four dyslexia candidate genes identified so far (*DYX1C1*, *ROBO1*, *DCDC2* and *KIAA0319*). Although several non-synonymous changes were identified in *MRPL19* and *C2ORF3*, none of them significantly associated with dyslexia. However, heterozygous carriers of the risk haplotype showed significantly attenuated expression of both *MRPL19* and *C2ORF3*, as compared with non-carriers. Analysis of *C2ORF3* orthologues in four non-human primates suggested different evolutionary rates for primates when compared with the out-group. In conclusion, our data support *MRPL19* and *C2ORF3* as candidate susceptibility genes for *DYX3*.

*To whom correspondence should be addressed. Tel: +46 86089158; Fax: +46 87745538; Email: juha.kere@biosci.ki.se

INTRODUCTION

Developmental dyslexia is a specific disorder in learning to read and spell in spite of adequate educational resources, normal intelligence, no obvious sensory deficits and adequate sociocultural opportunity. Affecting ~5% of school-aged children, dyslexia is the most common learning disorder (1–3). Dyslexic individuals show impairments in several correlated cognitive processes, whereas the core deficit is most common in phonological processing (4). Neuroanatomical and functional studies have indicated several differences between dyslexic and normal readers, e.g. different brain activation patterns and processing pathways in response to auditory and visual perception tasks (5).

Dyslexia is strongly familial, and abundant evidence supports genetic factors in its etiology (6). Linkage and association studies have investigated dyslexia both as a categorical trait and as a composite condition, with several independent components analyzed contributing to the disorder (7). To date, nine (*DYX1–DYX9*) chromosomal regions have been confirmed (www.gene.ucl.ac.uk/nomenclature). Four candidate genes for the susceptibility of developing dyslexia have been suggested: *DYX1C1* for the *DYX1* locus (8), *DCDC2* (9,10) and *KIAA0319* (11,12) for the *DYX2* locus, and *ROBO1* for the *DYX5* locus (13).

Three independent genome-wide scans using different analytical approaches have shown linkage of dyslexia to 2p12-p16 (14–16). Fagerheim *et al.* (14) studied a single extended Norwegian pedigree, in which inheritance was consistent with an autosomal dominant transmission. Parametric linkage analysis found significant evidence of linkage (maximum LOD score 4.3) to 2p15-p16. Fisher *et al.* (15) analyzed two large sets of nuclear families from UK and USA using a quantitative non-parametric approach, and found significant single-point linkage results for orthographic coding ($P = 0.0007$ at 2p16) in the UK sample and phoneme awareness ($P = 0.0003$ at 2p13) in the US sample. Petryshen *et al.* (17) performed a linkage study in Canadian families by genotyping seven microsatellite markers spanning the region on 2p15-p16 reported by Fagerheim *et al.* Multipoint variance component linkage analysis of different reading-related measures yielded an LOD score of 3.82 for spelling. Francks *et al.* (18) performed a quantitative sib-pair association study by genotyping microsatellites in the 2p12-p21 region. Two loci at 2p21 and 2p12 yielded P-values <0.05 for a range of reading-related measures.

In our previous genome-wide scan on 11 Finnish pedigrees, we identified linkage to the broad Chromosome 2 *DYX3* locus using a categorical phenotype (16). Parametric linkage analysis peaked at marker D2S286, LOD score of 3.01 and non-parametric analysis at marker D2S2216, NPL score 2.55 ($P = 0.004$). We subsequently refined this candidate region to ~12 cM by linkage and association analysis using microsatellite markers (19). In the present study, we have further refined the 2p12 candidate region in two populations, Finnish and German, and report evidence supporting two genes, *MRPL19* (mitochondrial ribosomal protein 19) and *C2ORF3* (chromosome 2 open reading frame 3), as candidate susceptibility genes for *DYX3*.

RESULTS

Linkage disequilibrium mapping of the 2p12 dyslexia candidate region in Finnish families

A total of eight microsatellites and 43 single nucleotide polymorphisms (SNPs)/deletion–insertion polymorphisms (DIPs) were successfully genotyped in two rounds of linkage disequilibrium (LD) mapping in 11 Finnish families (Fig. 1A and B). Markers from the second stage were also genotyped in eight additional Finnish families (Fig. 1B). The genotype data were analyzed for single-marker and haplotype (two to four marker sliding window) associations using the transmission disequilibrium test (TDT).

In the first round of LD mapping, the most significant single-marker associations were observed for markers rs917235 and rs730148. Alleles G and C were over-transmitted to affected subjects (14 transmitted versus two non-transmitted, $P = 0.0027$ and 21 transmitted versus six non-transmitted, $P = 0.0039$, respectively). Haplotype analysis showed the most significant association for the four-marker haplotype rs1859708-rs1986238-rs2010599-rs730148 (CCAC, $P = 0.0039$, 11 transmitted versus 1 non-transmitted).

In the second stage, marker density was increased to one every 8 kb in a 157 kb region from rs718507 to rs3755477 (Fig. 1B). This region was chosen to cover only the three genes located in the area of the associated markers/haplotypes (Fig. 1B). Single-marker TDT gave the same results as in the previous stage for rs730148, while rs917235 showed 18 transmissions versus four non-transmissions of allele G ($P = 0.0028$), and the most significantly associated haplotypes were the two-marker haplotype rs917235-rs714039 (GG, $P = 0.0029$) and the three-marker haplotype rs-10000585-rs917235-rs714039 (GGG, $P = 0.0076$) (Table 1).

Replication in a large independent sample set

Two rounds of genotyping were similarly performed in an independent set of 251 German families. Totally, 29 SNPs/DIPs were analyzed in the full sample set while four additional SNPs were only genotyped in a subset of 118 triads, due to DNA availability (Fig. 1A and B). In the first stage, a four-marker haplotype rs1859708-rs1986238-rs730148-rs721390 was over-transmitted to affected subjects (CCCC, $P = 0.0092$; 43 transmitted versus 22 non-transmitted). In the second round of LD mapping, the most significant association was seen for the three-marker haplotype rs917235-rs714939-rs6732511 (GGC, $P = 0.0036$). In a joint analysis of the two sample sets, two significant and overlapping three-marker haplotypes (P-values of 0.0049 and 0.0013, respectively), covering totally 16.6 kb, delineated the region of association in both populations (Table 1).

Correlation with the severity of phenotype

Because many studies of dyslexia have reported stronger positive associations with more severe phenotypes (10,11,20,21), we re-analyzed the most significantly associated haplotype in the German set (rs917235-rs714939-rs6732511, GGC) by stratifying for severity. Detailed phenotypic data were not
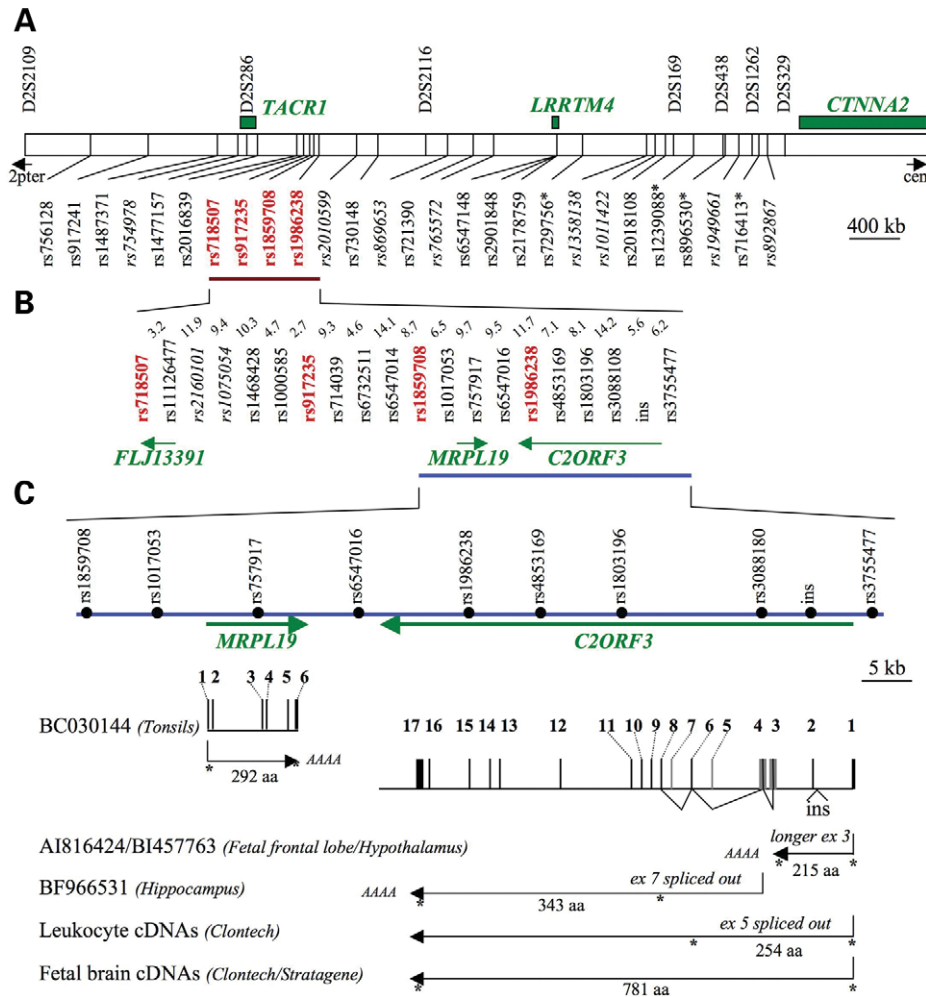
Figure 1. Physical map over the 2p12 dyslexia candidate region. (**A**) Microsatellite and SNP markers genotyped in the first round of mapping are indicated over and below, respectively, the horizontal bar representing an ~7.3 Mb long chromosomal region within 2p12. The physical distance between rs756128 and rs892867 is 5 Mb. Microsatellites and SNPs in italics were genotyped only in the Finnish sample set. Three of the newly genotyped microsatellites are located within this region: D2S2109, D2S438 and D2S1262. SNPs genotyped only in a subset of the German triads are marked with an asterisk. The red color (markers and the horizontal line) indicates a ~157 kb area showing a common association in the Finnish and German sample sets. Rectangles denote the genomic extent of the three functional candidate genes (*TACR1*, *LRRTM4* and *CTNNA2*) screened for mutations in the Finnish families. Note that D2S286 is included in the *TACR1* gene. (**B**) SNPs added in the second stage of LD mapping are indicated between the markers from the previous genotyping stage (in red). Markers rs2160101 and rs1075054 were only genotyped in the Finnish sample set (success rate <80% in the German sample set). 'Ins' denotes the position of the 19 bp insertion that was detected in two of the three individuals sequenced and further genotyped in all individuals. Intermarker distance is indicated in kilobases between the marker names. The genomic extent of only the three genes from the region, *FLJ13391*, *MRPL19* and *C2ORF3*, is symbolized by green arrows showing their chromosomal orientation. The blue bar indicates the genomic region sequenced in three individuals. (**C**) The genomic structure of the corresponding gene is drawn under each green arrow with exons numbered according to gene orientation. Detected alternative splicing events are indicated by triangles under the horizontal black bar representing the *C2ORF3* gene. Exons in gray were found to be either fully- (exons 5 and 7) or partially- (exons 3 and 4) spliced out. Fully sequenced I.M.A.G.E. clones and PCR-amplified cDNAs from four commercial cDNA libraries are drawn under each gene, with their tissue of origin; BC030144 for *MRPL19* and AI816424, BI457763 and BF966531 for *C2ORF3*. Asterisks denote the first methionin (ATG) and the stop codon for each mRNA species, with the size in amino acids of the corresponding protein product. '*AAAA*' denotes a polyadenylation tail in the corresponding mRNA.

available for the Finnish sample set. Probands were classified for severity with a discrepancy of 1, 1.5, 2 or 2.5 standard deviations (SD) between the observed and expected spelling scores. All 251 probands fulfilled the criteria of a difference of 1 SD, whereas 232 and 171 probands showed a 1.5 SD and 2 SD discrepancy, respectively, and 72 probands displayed the most severe spelling disorder (2.5 SD). The odds ratio (OR) for the risk haplotype increased from 2.2 (global $P = 0.006$) for all probands to 5.2 for the most severely affected cases (global $P = 0.00005$) (Table 1).

## Attenuated expression of *MRPL19* and *C2ORF3* in heterozygous carriers of the risk haplotype

An extensive search for genes was performed in the 16.6 kb region of association and in up to 86 kb of its surroundings. However, the only genes present were the hypothetical *FLJ13391* and the verified *MRPL19* and *C2ORF3* genes. As the haplotype block structure of the region revealed a 62 kb block of strong LD containing *MRPL19* and *C2ORF3* (Fig. 2), we hypothesized that the risk haplotypes might lie

Table 1. Dyslexia risk haplotypes in the Finnish, German and the pooled data

Sample set	SD[a]	TDT		TDTHAP		Global P[a]	OR	95% CI	rs1000585	rs917235	rs714939	rs6732511
		T/U	P	T/U	P							
Finnish	–	17/3	0.003	17/7	0.06	0.08	2.3	1.2–5.2		G	G	
	–	14/3	0.008	16/7	0.09	0.07	2.2	1.1–4.9	G	G	G	
German	–	64/35	0.004	67/41	0.016	0.04	1.6	1.2–2.3		G	G	C
	1.0	44/20	0.003	46/21	0.003	0.006	2.2	1.4–3.4				
	1.5	42/18	0.002	43/19	0.003	0.006	2.2	1.4–3.6				
	2.0	30/12	0.005	31/13	0.009	0.005	2.3	2.3–4.1				
	2.5	18/3	0.001	18/3	0.002	0.00005	5.2	2.3–18.0				
Pooled	–	90/56	0.005	114/79	0.014	0.017	1.4	1.1–1.8	G	G	G	
	–	69/36	0.001	77/47	0.009	0.067	1.6	1.2–2.2		G	G	C

Severity data only available for the probands of the German sample set. T/U, transmitted/untransmitted chromosomes.
[a]Global P-values were obtained using TDTHAP.
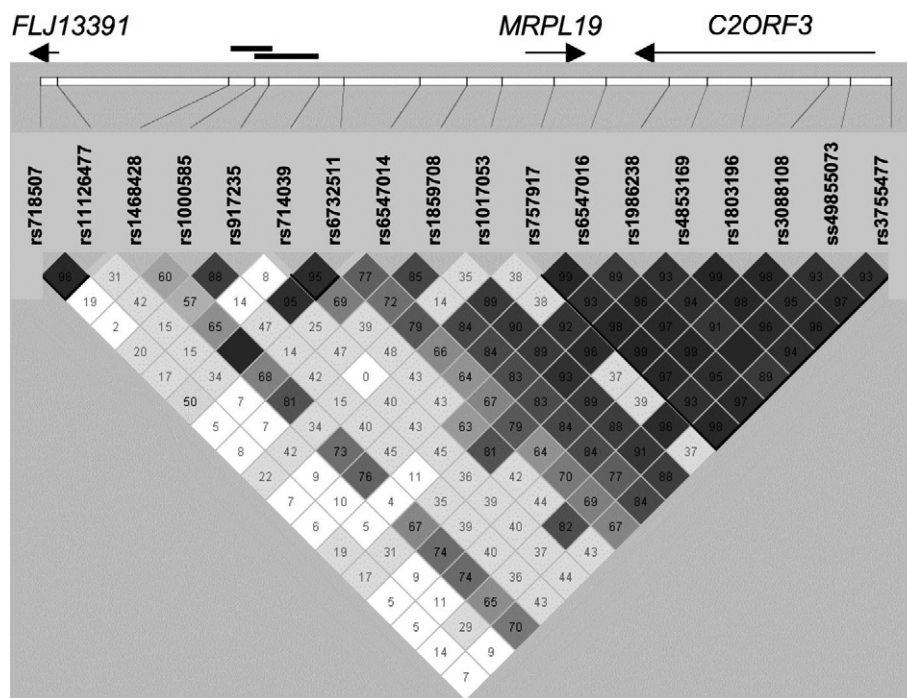
Figure 2. Linkage disequilibrium of the genotyped SNPs within the 157 kb region in the pooled set (Finnish and German samples). Solid black lines denote the two overlapping dyslexia risk haplotypes (GGG and GGC).

in a putative regulatory region of the two genes. We therefore evaluated the expression levels of *MRPL19* and *C2ORF3* in EBV-transformed lymphocyte cell lines of carriers and non-carriers of the risk haplotype. We measured whether both alleles were equally transcribed in affected and non-affected individuals heterozygous for synonymous variants of the two genes. Of the 15 samples analyzed, five dyslexic and four normal readers carried the risk haplotype at markers rs917235-rs714939 (GG) in heterozygous form. Three normal readers and three affected did not carry the haplotype. By comparing the peak height ratios in genomic DNA and cDNA, we observed a significant difference in the expression levels of *MRPL19* and *C2ORF3* for the two alleles of rs17689863 and rs1803196, respectively (Fig. 3). When associated with the risk haplotype, the more common allele

was significantly less transcribed for both genes. The cDNA to DNA ratio was <1 in all carriers.

Identification of new SNPs within the genomic region of *MRPL19* and *C2ORF3*

We hypothesized that there might be two separate mechanisms in this region for the susceptibility of developing dyslexia, i.e. the risk haplotype in the putative regulatory region and/or SNPs within the coding region of one of the two genes. Therefore, we sequenced all coding exons and the flanking sequences of *MRPL19* and *C2ORF3* in one affected individual from each of the 19 Finnish families. Five novel mutations, ss65713215 in exon 1 (Phe16Ala) and ss65713216 in exon 3 (Val93Ile) of *MRPL19*, and ss65713213 in exon 9
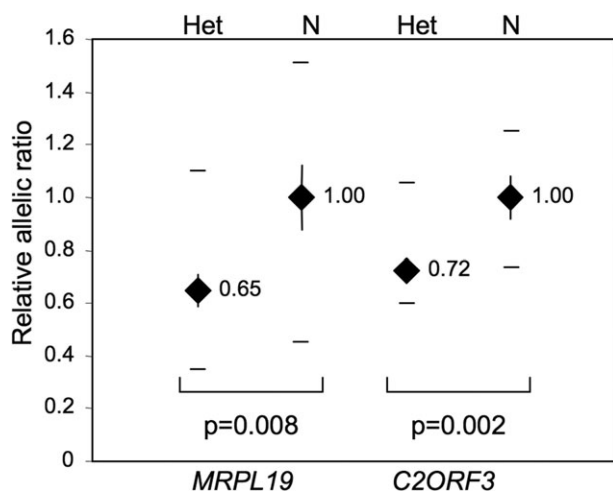
Figure 3. Attenuation of expression of the *MRPL19* and *C2ORF3* alleles associated with the dyslexia risk haplotype (GG). The results are expressed as the relative allelic ratio of mRNA in all heterozygous carriers (Het) when compared with the corresponding mRNA level in non-carriers (N). Data are shown as mean $\pm$ SEM. The small horizontal bars denote the range of ratios. Significance was assessed by a two-tailed *t* test.

(Gln425Glu) and ss65713214 in exon 11 (Ile553Asn) of *C2ORF3* were identified as heterozygous changes (Supplementary Material, Table S1). Each variation was seen in single individuals, except for ss65713215 and ss65713213, which were present in two and three unrelated individuals, respectively. We genotyped the identified novel coding SNPs (cSNPs), as well as all cSNPs reported in the dbSNP database, in the full sample set of Finnish and German families. No over-transmissions could be observed to affected individuals, and the allele frequencies in affected and unaffected were approximately equal (Supplementary Material, Table S1), suggesting that none of these variants was functionally relevant.

To fully explore the genetic variation at this locus, we performed genomic sequencing over a ~86 kb region (from 54677889 to 54764033 bp, according to the public-map contig NT_022184.14, build 36) encompassing *MRPL19* and *C2ORF3* (Fig. 1B). Two affected subjects (one of Finnish and one of German descent) homozygous for the four-marker risk haplotypes CCAC (495 kb) and CCCC (1.2 Mb), respectively, and a German affected individual homozygous for the opposite non-risk haplotype, were sequenced over all exonic, non-repetitive intronic and the putative promoter regions of *MRPL19* and *C2ORF3*. We could identify totally 121 SNPs and 10 DIPs, of which 27 and six where novel, respectively (submitted to dbSNP under accession numbers ss49855067-ss49855099, www.ncbi.nlm.nih.gov/SNP). No new cSNPs were found while six already known coding variants could be identified, one in *MRPL19* and five in *C2ORF3*.

Correlation of expression in different regions of adult human brain

We studied the expression levels of *MRPL19*, *C2ORF3* and *FLJ13391* in nine different regions of the human brain as well as in whole adult and fetal brain using quantitative real-time RT–PCR. *MRPL19* and *C2ORF3* were highly expressed in all areas of adult brain, and average threshold cycle (Ct) values of 26 and 31 were obtained, respectively. The expression in fetal whole brain was at the same level as the expression in adult. The average expression level for *FLJ13391* was considerably lower (Ct 36). When normalized to adult whole brain expression, *MRPL19* and *C2ORF3* mRNA levels were all above, or similar to, this baseline and their pattern of expression over the different parts of the brain tested was correlated ($R^2 = 0.48$) (Fig. 4A). In contrast, *FLJ13391* displayed a very different pattern of expression across the various samples. The expression of the already identified dyslexia candidate genes *DYX1C1, ROBO1, DCDC2* and *KIAA0319* was studied as a comparison. Overall, their average expression levels were high across all the studied brain regions (Ct 29, 28, 31 and 27, respectively). The expression of *C2ORF3* was correlated across the different brain regions with *DYX1C1, ROBO1* and *DCDC2* ($R^2 = 0.54$, 0.69 and 0.76, respectively) (Fig. 4B). *KIAA0319* showed a very even expression across the studied brain regions, and hence a different pattern when compared with *C2ORF3* ($R^2 = 0.15$). Interestingly, for *MRPL19* the correlation was strongest for *KIAA0319* ($R^2 = 0.47$), and weaker for *DYX1C1, ROBO1* and *DCDC2* ($R^2 = 0.35$, 0.43 and 0.20, respectively) (data not shown).

Transcript characterization of the three genes in the region

We verified the gene structure and the exon–intron borders for *FLJ13391*, *MRPL19* and *C2ORF3* by PCR on human fetal brain and lymphocyte cDNA libraries and by fully sequencing I.M.A.G.E. clones. For *FLJ13391*, clone BC063016 consisted of a 1516 bp mRNA containing the full coding region of 456 bp encoding the 152 amino acids (aa) of the protein. However, the first untranslated exon according to the 1817 bp NM_032181 was missing. For *MRPL19*, only one 1347 bp transcript was identified which was in agreement with the public database sequence, NM_014763 (Fig. 1C). For *C2ORF3*, we could detect the main long mRNA consisting of 17 exons in cDNA libraries from both fetal brain and leukocytes. However, the transcript was shortened in both the 5′ and 3′ untranslated region (UTR), yielding a 4366 bp transcript (predicting a 781 aa protein) instead of the 5185 bp NM_003203. The other already known transcript (AAH00853) containing exons 1–3 only, was detected in two of the I.M.A.G.E. clones, however with a longer form of exon 3 (EF158467) (Fig. 1C). Exon 5 was found spliced out in ~50% of the cDNAs from leukocytes, which has not been reported previously (EF158468). When exon 5 was excluded (116 bp), 15 new aa were introduced before a premature stop codon in exon 6, leading to a 254 aa protein. In BF966531 (3650 bp), exon 4 started 87 bp upstream of the consensus sequence and exon 7 was spliced out (EF158469) (Fig. 1C). Five-prime RACE experiments performed on *MRPL19* and *C2ORF3* did not reveal any additional coding base pairs that have not previously been reported and verified.
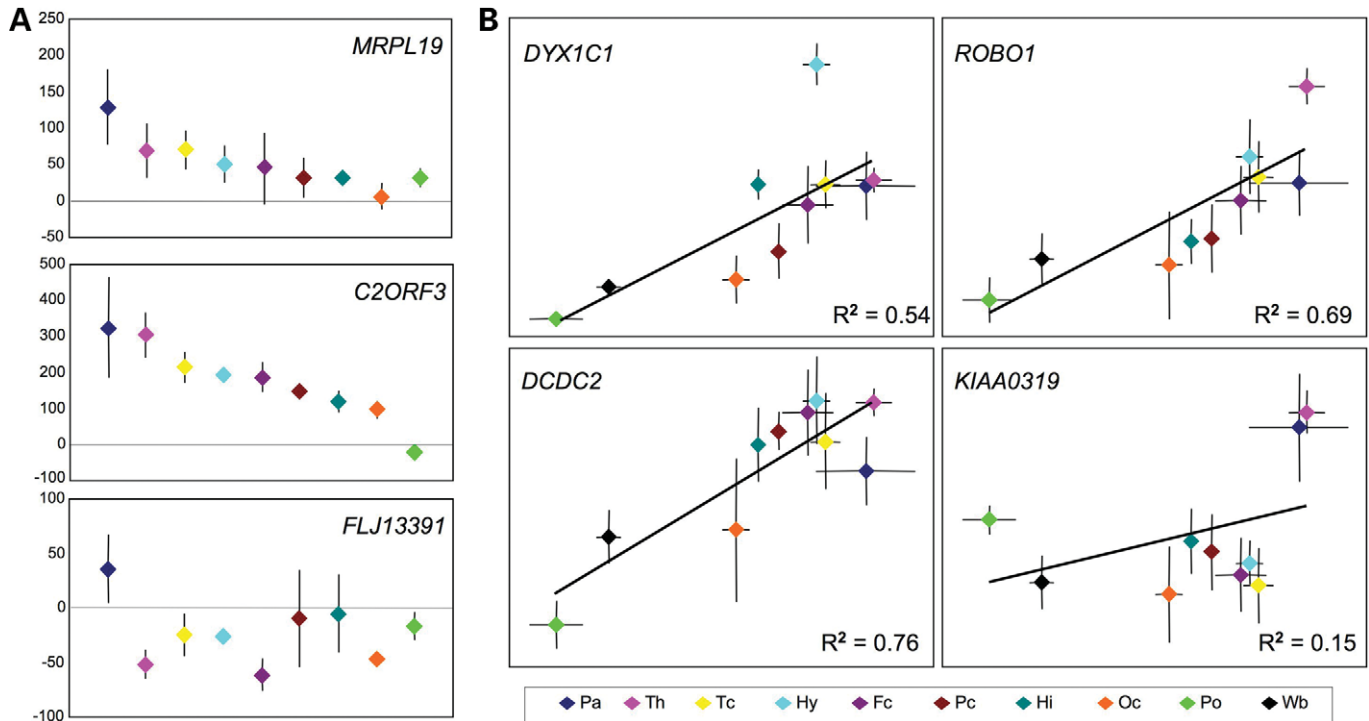
Figure 4. Expression studies of the *MRPL19*, *C2ORF3* and *FLJ133091* genes in nine regions of adult human brain as well as adult whole brain. (**A**) Quantitative real-time PCR results for *MRPL19*, *C2ORF3* and *FLJ133091*. The level of expression is shown in percentage on *y*-axis after normalization to adult whole brain expression. (**B**) Correlation of *C2ORF3* (*x*-axis) expression with *MRPL19* and the dyslexia candidate genes *DYX1C1*, *DCDC2*, *ROBO1* and *KIAA0319* (*y*-axis, respectively) in the different brain regions. *x*- and *y*-axes are in $\log_2$ scale in arbitrary units. For clarity, the scales are not shown. Pa, paracentral gyrus; Th, thalamus; Tc, temporal cortex; Hy, hypothalamus; Fc, frontal cortex; Pc, parietal cortex; Hi, hippocampus; Oc, occipital cortex; Po, post-central gyrus; Wb, whole brain.

Mutation screening of two additional positional candidate genes

In addition to the three studied genes in the region of association (*FLJ13391*, *MRPL19* and *C2ORF3*), *CTNNA2* (catenin alpha-2) and *LRRTM4* (leucine-rich repeat transmembrane neuronal 4) are the only known genes, besides a cluster of pancreatic-specific genes, within a 5 Mb genomic region from *TACR1* to *CTNNA2* (Fig. 1A). As these two genes are highly expressed in the human brain and represented functional candidate genes for dyslexia, they were screened during the mapping process for mutations/variations in affected subjects of Finnish origin. However, no coding variants were detected in the coding exons or splice sites of either of them. Furthermore, TDT did not reveal any signs of association in the *LRRTM4* gene in the Finnish or the German sample set.

Analysis of selection pressure during the evolution of *MRPL19* and *C2ORF3*

We looked for signs of recent selection in the *MRPL19* and *C2ORF3* genes since the divergence from the orangutan and gorilla branches, by sequencing the coding regions in four non-human primate species. For *MRPL19*, only one non-synonymous substitution was identified in chimpanzee, as well as one non-synonymous and one silent substitution in pigmy chimpanzee and gorilla, respectively (Supplementary

Material, Table S2). In orangutan, 18 different variants were discovered. On the contrary, several variants were identified in *C2ORF3* in all primates analyzed with a total of 18 SNPs in pigmy chimpanzee, 24 in chimpanzee and 29 in gorilla (Supplementary Material, Table S3). The predicted C2ORF3 proteins for pigmy chimpanzee, chimpanzee and gorilla differ in 8, 12 and 15 amino acids (1.0, 1.5 and 1.9% of residues), respectively, when compared with the human homologue. An over-representation of nucleotide substitutions was found in exon 1 for all primates. The orangutan exons could not be amplified with the human-specific primers used, suggesting a very low sequence identity in the flanking intronic sequence (50–100 bp).

We calculated the rate of synonymous (dS) and non-synonymous substitutions (dN) in all species studied. We applied likelihood ratio tests (LRTs) to analyze the selection pressure $\omega = \mathrm{d}N/\mathrm{d}S$ for *MRPL19* and *C2ORF3* during primate evolution using the CODEML software included in the paml3.15 package (22) (Supplementary Material, Table S4). For *MRPL19*, the low number of sequence alterations drastically reduces power in the LRTs and the estimates will not be reliable. For *C2ORF3*, a model specifying independent ω ratios for all branches (free ratios) showed a significantly better fit to our data when compared with a model assuming a single ω for all lineages (Table 2 and Fig. 5). To predict selection pressure changes during primate evolution, we constructed a two-ratio model specifying an identical ω for the primates and a different ω for the out-group. The

Table 2. Analysis of the dyslexia candidate C2ORF3 protein evolution by LRTs

Models tested
(A) One ratio = Single ω for all branches
(B) Free ratios = Independent variable ω for whole tree
(C) Two ratios = Separate ω for in-group and out-group

Test	Model	χ^2 value[a]	P	Conclusion
Homogeneity for whole tree	A versus B	23.23	0.002	Freely variable ω more likely than single ω in all lineages
In-group = Out-group	A versus C	21.63	<0.001	ω out-group > ω in-group
Homogeneity for in-group	C versus B	1.60	0.90	No evidence of heterogeneity in in-group

ω, dN/dS ratio.
[a]The χ^2 value is two times the difference in log-likelihood values.
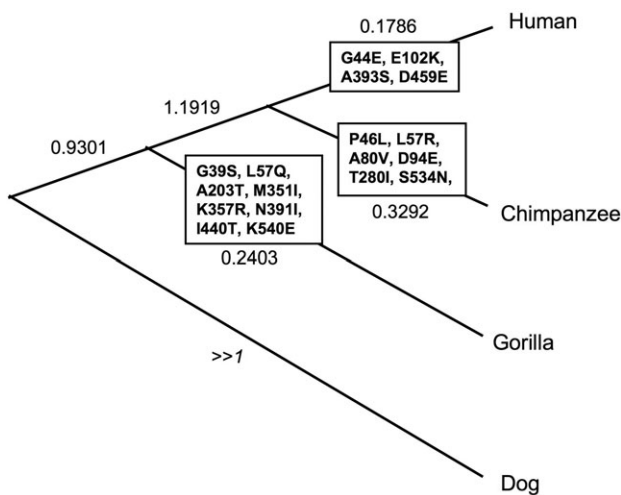
Figure 5. Phylogenetic tree of the dyslexia candidate C2ORF3 protein evolution in primates. Identified non-synonymous changes are shown in the boxes. dN/dS ratios of the branches were calculated with the CODEML program of the paml3.15 package, assuming a free variable dN/dS ratio. Dog was used as the out-group in sequence comparisons.

two-ratio model was more likely than the free ratios model indicating evidence of change in the selection pressure from dog to primates owing to a significant sequence diversity in protein coding regions of *C2ORF3*. However, there was no significant heterogeneity among the primates, indicating no change in selection pressure during the evolution of *C2ORF3* from non-human primates to the human lineage (Table 2). The same conclusion was drawn from an LRT without the dog *C2ORF3* sequence, using gorilla as the out-group (data not shown).

DISCUSSION

We have previously confirmed the presence of a dyslexia locus on 2p12 using both linkage and association analysis in Finnish families (16,19). Here, we refine its location from 12 Mb to a 157 kb region and identify two overlapping risk haplotypes of 16 kb segregating with dyslexia in both Finnish and German sample sets. Moreover, the OR for the most common susceptibility haplotype increased significantly in more severe cases of dyslexia (severity measures available for the German probands). The 157 kb candidate region harbors two co-regulated genes, *MRPL19* and *C2ORF3*. Despite much effort, we have not obtained evidence for additional transcripts in this genomic region. We propose these genes as candidates for the susceptibility of developing dyslexia at the *DYX3* locus.

Because the two independent sample sets supported the association findings, there is strong evidence for the involvement of this locus in dyslexia. Other studies have found support for this region as well, although the linked/associated loci have been widely spread over the short arm of chromosome 2. This suggests the possibility of two dyslexia loci, one on 2p15 and our locus on 2p12, both supported by Fisher *et al*. (15). Alternatively, there is only one locus with an inaccurate definition of its position by linkage. As we used a categorical diagnosis of dyslexia both in the initial genome-wide scan (16) and in further fine mapping studies (19), this locus seems to have a general effect on dyslexia, i.e. word reading and spelling. This conclusion is further supported by the observation of a stronger effect in the more severe cases from Germany. Even though Fisher *et al*. (15). and Francks *et al*. (18). studied different quantitative processes of dyslexia, they found evidence of linkage and association at approximately the same locus as we report here.

The associated haplotypes that we identified in the two sample sets are located in the intergenic region between *FLJ13391* and the *MRPL19–C2ORF3* genes. Because our extensive search for possible novel genes throughout the whole ~80 kb region between *FLJ13391* and *MRPL19* yielded no results, it is unlikely that the associated region would harbor an, yet, unidentified susceptibility gene. Instead, the associated SNPs might be non-coding variants in a regulatory region for *MRPL19* and *C2ORF3*. In support of this hypothesis, our expression data showed that these genes are co-regulated in different brain areas. This observation is further supported by publicly available data from pooled microarray experiments (http://microarray.cpmc. columbia.edu/tmm). Moreover, these two genes are in strong LD belonging to a single haplotype block.

The suggested regulatory effect of the associated haplotype was further supported by allele-specific expression analysis.

We assessed the allelic balance in mRNA for the two genes in two groups: individuals heterozygous for the risk haplotype, and those heterozygous for non-risk haplotypes. We found that the level of expression assayed by a synonymous SNP in each of the two genes was significantly decreased for the alleles associated with the risk haplotype. There are several reported examples of haploinsufficiency associations to susceptibility, such as for the dyslexia candidate genes *ROBO1* (13) and *KIAA0319* (11,12). Furthermore, despite an effort to identify new cSNPs that might provide a simple functional explanation for the susceptibility at this locus, none of the observed coding changes in *MRPL19* and *C2ORF3* were associated with dyslexia (Supplementary Material, Table S1).

Further support for the involvement of either *MRPL19* or *C2ORF3* or both in dyslexia was obtained by correlating their expression with the previously proposed dyslexia candidate genes *DYX1C1* (8), *DCDC2* (9,10), *KIAA0319* (11,12) and *ROBO1* (13). Our quantitative expression analysis across the different brain regions showed high expression of *MRPL19* and *C2ORF3* in all brain areas tested, and abundant expression in regions implicated in reading by functional and imaging methods such as the inferior frontal and temporal occipital area; the superior temporal, parietal temporal and middle temporal–middle occipital gyri (5,23). The expression of both *C2ORF3* and *MRPL19* correlated strongly with the other dyslexia candidate genes. In contrast, *FLJ13391* showed a very different pattern of expression than any of the other genes studied, and therefore is considered as a much less likely candidate for dyslexia susceptibility at the *DYX3* locus.

Finally, an evolutionary analysis revealed high levels of variation in *C2ORF3* in primate and non-primate species. An accelerated rate of protein evolution in primates, especially in the human lineage, has been shown for a number of genes important for nervous system development and function (24,25). Positive selection during recent human evolution was suggested for *FOXP2* (26), and the selection pressure was also found to be different for *ROBO1* between the human, chimpanzee and gorilla branches when compared with the orangutan (13) although *ROBO1* has been proposed to be a slowly evolving gene due to the large excess of silent changes in each primate lineage (6). A test for heterogeneity among the primate species revealed no evidence of change in the selection pressure during primate evolution of *MRPL19* and *C2ORF3*. The relatively low dN/dS ratios estimated for *C2ORF3* in this study are consistent with previous reports of low dN/dS ratios for nervous system genes (27). However, the stringent definition of adaptive evolution, $\omega > 1$, in estimations of selection pressure may be misleading for many genes expressed in brain as low ω-values may mask signs of adaptive evolution. Furthermore, we report a nearly equal proportion of synonymous to non-synonymous substitutions in primate *C2ORF3* (50% and 45% synonymous changes in chimpanzee and gorilla, respectively) and a 98% identity relative to the human orthologue at the protein level. This finding together with the fact that the non-primate lineage show comparatively higher dN/dS ratios may indicate that C2ORF3 is under functional constraint owing to an important function in the brain acquired during primate evolution.

In contrast, *MRPL19* is a highly conserved gene with only a few nucleotide changes. Therefore, the maximum likelihood estimates using dN/dS ratios for *MRPL19* were inconclusive. This gene may have a central role in ribosome biogenesis and mitochondrial protein synthesis. Nevertheless, minor changes in the protein, leading to marginally impaired energy metabolism may have developmental consequences in critical tissues. Many of the mitochondrial ribosomal proteins encoded in the nucleus have been associated with several neurological disorders, such as deafness (28), in accordance to the fact that energy production is critical in the active brain.

In conclusion, our data support the involvement of the 2p12 locus in the development of dyslexia and the role of either or both genes, *MRPL19* and *C2ORF3*. MRPL19 protein may participate in mitochondrial energy metabolism, whereas the cellular function of C2ORF3 is unknown (it was initially falsely thought to be a transcription factor due to a chimeric cDNA clone), and needs to be addressed in future studies. Several lines of evidence support either or both of these genes as relevant candidates for the *DYX3* locus.

MATERIALS AND METHODS

Subjects and genomic DNA preparation

Eleven Finnish three-generation pedigrees consisting of 83 subjects (34 affected, 41 healthy, 8 phenotype unknown) were genotyped in the first round of fine mapping. Kindreds have been partly described previously (16), but because of sample and/or phenotype availability, 13 more subjects were included. In the second round of fine mapping, eight additional families (47 individuals, 16 affected, 22 healthy, nine unknown) were added to the analysis. This expanded the Finnish sample set to 130 individuals (50 affected, 63 healthy, 17 unknown). All phenotypes were ascertained as previously described (29).

The replication set contained 251 German families. The sample set consisted of altogether 1050 individuals, with 429 affected (251 probands and 178 sibs), 119 healthy sibs and 502 of unknown phenotype (all parents). The samples were recruited from the Departments of Child and Adolescent Psychiatry and Psychotherapy at the Universities of Marburg (59%) and Würzburg (41%). The diagnostic criteria and phenotypic measures have been described in detail previously (30).

Genomic DNA for the Finnish and German samples was extracted from blood using standard methods (31,32). Genetic studies have been approved by the appropriate ethical committees in Finland, Sweden and Germany.

Genotyping

In the first stage of fine mapping, eight microsatellite markers (D2S2109, D2S438, D2S1262, D2S253, D2S289, D2S2162, D2S435, D2S394) and 24 SNPs with an average spacing of 225 kb (range 42–556 kb) were genotyped in the Finnish kindreds. Twenty of these SNPs were genotyped in the German samples. In the second stage of fine mapping, 15 additional SNPs were selected over the 157 kb candidate region and genotyped in the full Finnish and German sample sets. We

also tagged the *LRRTM4* gene using three additional SNPs (rs654148, rs2901848, rs2178759) and moreover, 11 cSNPs with minor allele frequency >5% were genotyped in the full sample set of Finnish and German families. Altogether, nine SNPs with low success rates were removed from analysis.

All SNPs were genotyped using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF, Sequenom) as described previously (19), by sequencing, or by PCR amplification and visualization in agarose gels (19 bp DIP ss49855073). The Sequenom assays were designed using the SpectroDESIGNER software and are available upon request. A genotyping success rate of 80% was required for inclusion in analyses. All genotypes were independently confirmed by two investigators. Data were checked for Mendelian consistency using Pedcheck (33), and unresolved inconsistencies were assigned as missing genotypes.

PCR and sequencing reactions

PCRs (all primer sequences available upon request) were carried out in $10-25$ µl reactions containing $0.5-1$ ng/µl of genomic DNA, $1.5-3$ mM $MgCl_2$, 0.4 mM of each dNTP, 0.8 µM of each primer and 0.03 U/µl of HotStarTaq DNA polymerase (Qiagen). We used a touch-down protocol with 42 cycles of amplification with 1°C of decrease in annealing temperature at each round; two cycles at 63°C and at 62°C, respectively, three cycles at 61°C and 56°C, respectively and 10 cycles at 55°C and 54°C, respectively. PCR cycles had an initial denaturation at 95°C for 15 min, 30 s at each annealing temperature and 30 s to 1 min 30 s elongation at 72°C with a final extension of 10 min at 72°C. Primate DNA PCR was carried out following the touch-down protocol but ending at 55°C for 25 cycles.

PCR products were dephosphorylated by 0.4 U/µl shrimp alkaline phosphatase (Amersham Biosciences/GE) and 2 U/µl exonuclease I (New England BioLabs) and were further sequenced using DYEnamic ET Dye terminator kit (Amersham Biosciences/GE) following the manufacturer's instructions. Each fragment was sequenced in both directions using the amplification primers. Purified sequencing products were resolved using a MegaBACE 1000 instrument and MegaBACE long-read matrix (Amersham Biosciences/GE), visualized using the Sequence Analyzer v3.0 software (Amersham Biosciences/GE), and assembled and analyzed using the Pregap and Gap4 software (www.cbi.pku.edu.cn/tools/staden), comparing to the sequence NT_022184.14, build 35 (www.ncbi.nih.gov). Sequences were verified visually by two independent readers.

Genomic sequencing was performed for two affected individuals homozygous for the susceptibility haplotype, one of German and one of Finnish descent, and one affected individual (German) homozygous for the opposite non-risk haplotype. The ~86 kb genomic sequence was first masked for repeats (woody.embl-heidelberg.de/repeatmask) and the unique segments (~46 kb) were sequenced. In total, 109 fragments ($200-1000$ bp, with $100-200$ bp overlaps) were amplified by PCR and sequenced as described above.

Four candidate genes, *LRRTM4*, *CTNNA2*, *MRPL19* and *C2ORF3* were screened for polymorphisms by direct sequencing of all their coding exons and exon–intron junctions. Nineteen affected individuals were sequenced (11 for *CTNNA2*). The human primers were used to sequence *MRPL19* and *C2ORF3* in Gorilla (*Gorilla gorilla*), chimpanzee (*Pan troglodytes*), pigmy chimpanzee (*Pan paniscus*) and orangutan (*Pongo pygmaeus*) DNA samples (Primate panel PRP00001 IPBIR, Camden, New Jersey, USA).

Gene characterization and expression analysis of *FLJ13391*, *MRPL19* and *C2ORF3*

The gene structures were verified and improved by fully sequencing I.M.A.G.E. clones BC030144 (primary B-cells from tonsils) for *MRPL19*; BF665321 (primitive neuroectoderm), AI816424 (fetal frontal lobe), BI457763 (hypothalamus) and BF966531 (hippocampus) for *C2ORF3*; and BC063016 (pooled pancreas and spleen) for *FLJ13391*. To confirm the 5′ end of *MRPL19* and *C2ORF3* genes, we performed 5′ RACE experiments using Marathon-Ready cDNA from fetal brain tissue (cat. no. 639302; Clontech) following the manufacturer's instructions.

The expression of the three genes was studied by PCR on human cDNA libraries from fetal brain (human fetal brain large-insert cDNA library; cat. no. HL5504u, Clontech; human fetal brain Uni-ZAP XR library, cat. no. 052001b; Stratagene) and from leukocytes (human leukocyte large-insert cDNA library, cat. no. HL5509u, Clontech; human leukocyte 5′ STRETCH PLUS cDNA library, cat. no. HL5019t, Clontech). PCR products were visualized by agarose gel electrophoresis and further sequenced.

Putative new genes/exons from the 86 kb sequence between *FLJ13391* and *MRPL19* were predicted *in silico* with Genscan (genes.mit.edu/GENSCAN.html and vega.sanger.ac.uk/Homo_sapiens) and GrailEXP (grail.lsd.ornl.gov/grailexp). The expression of all 27 predicted genes/exons was then tested by PCR on the four human brain and leukocyte cDNA libraries. One gene prediction (GENSCAN59094, vega.sanger.ac.uk/Homo_sapiens) was thoroughly tested by screening >1 000 000 clones from each of two human fetal brain cDNA libraries because of its overlap with the risk haplotype at rs1000585-rs917235-rs714939.

Ready-made TaqMan gene expression assays for *FLJ13391* (Hs00259924_m1), *MRPL19* (Hs00608519_m1), *C2ORF3* (Hs00162632_m1), *DYX1C1* (Hs00370049_m1), *DCDC2* (Hs00393203_m1), *KIAA0319* (Hs00207788_m1), *ROBO1* (Hs00268049_m1), *GAPDH* (4310884E) and *18S rRNA* (4319413E) were purchased from Applied Biosystems. We assayed expression levels for these genes in total RNA from nine different areas of adult human brain: thalamus, hypothalamus, frontal-, occipital-, parietal-, temporal cortex (cat. nos 6762, 6864, 6810, 6812, 6814, 6816; Ambion), hippocampus, paracentral- and post-central gyrus (cat. nos 636565, 636574, 636573; Clontech), and from whole adult and fetal brain (cat. nos 636530, 636526; Clontech). For each tissue, three independent cDNA syntheses (500 ng total RNA per reaction) were performed using the SuperScript III first-strand synthesis kit (cat. no. 18080-051; Invitrogen). From each cDNA synthesis, quantitative real-time PCR was performed in quadruplets, using 5 ng of RNA per gene assay and run on ABI PRISM 7700 Sequence Detection PCR System (Applied Biosystems). All assays were performed in

10 μl reactions according to the manufacturer's instructions. Relative standard expression curves were drawn for 18S rRNA and all tested genes. Relative quantification of the data was performed using the comparative Ct (threshold cycle) method (Sequence Detection System bulletin 2, Applied Biosystems). Ct values were adjusted to 18S rRNA and thereafter normalized to the whole brain sample.

To quantify mRNA expression levels from each allele of *MRPL19* and *C2ORF3*, we analyzed individuals heterozygous for rs17689863 (Ser277Ser) in *MRPL19* and rs1803196 (Val536Val) in *C2ORF3*. Five Finnish dyslexic and six normal readers, and three German dyslexic and one normal reader were studied. Total RNA was extracted from EBV-transformed lymphocyte cell lines using the RNeasy purification kit (cat. no. 74004; Qiagen) and cDNA synthesis (500 ng of total RNA per reaction) was performed using the SuperScript III first-strand synthesis kit (cat. nos 18080-051 and 12371-019; Invitrogen; for the Finnish and the German samples, respectively). Both cDNA and genomic DNA from each individual were sequenced in six independent reactions, originating from at least two separate PCR amplifications. Peak heights were compared and an allelic ratio was calculated for each sequence. The cDNA ratio values (unknown proportions) were normalized by dividing with the genomic values (1:1 proportion by definition). Data were pooled by genotype (risk haplotype heterozygotes versus non-risk haplotype heterozygotes) to evaluate (by two-tailed *t*-test) whether the normalized value differed from equal expression.

Statistical analyses

TDT (34) was used to test for single marker as well as for haplotype (two to four markers) associations. Phased haplotypes and global *P*-values were obtained using TDTHAP (35). To assess global *P*-values, 50 000 permutations were run.

Intermarker LD was visualized and haplotype blocks were constructed using the Haploview3.2 software (36).

Evolutionary analysis of the *MRPL19* and *C2ORF3* genes was performed with an LRT using the CODEML program of the paml3.15 package (22). Mouse sequence (ENSMUSP-00000032124) was used as out-group for *MRPL19*, and dog (XP_540209) for *C2ORF3*.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at HMG Online.

ACKNOWLEDGEMENTS

Conflict of Interest statement. None declared.

REFERENCES

1. Shaywitz, S.E., Shaywitz, B.A., Fletcher, J.M. and Escobar, M.D. (1990) Prevalence of reading disability in boys and girls. Results of the Connecticut Longitudinal Study. *JAMA*, **264**, 998–1002.
2. Schulte-Korne, G. (2001) Annotation: genetics of reading and spelling disorder. *J. Child Psychol. Psychiatry*, **42**, 985–997.
3. Francks, C., MacPhie, I.L. and Monaco, A.P. (2002) The genetic basis of dyslexia. *Lancet Neurol.*, **1**, 483–490.
4. Vellutino, F.R., Fletcher, J.M., Snowling, M.J. and Scanlon, D.M. (2004) Specific reading disability (dyslexia): what have we learned in the past four decades? *J. Child Psychol. Psychiatry*, **45**, 2–40.
5. Demonet, J.F., Taylor, M.J. and Chaix, Y. (2004) Developmental dyslexia. *Lancet*, **363**, 1451–1460.
6. Fisher, S.E. and Francks, C. (2006) Genes, cognition and dyslexia: learning to read the genome. *Trends Cogn. Sci.*, **10**, 250–257.
7. Schulte-Korne, G., Ziegler, A., Deimel, W., Schumacher, J., Plume, E., Bachmann, C., Kleensang, A., Propping, P., Nothen, M.M., Warnke, A. *et al.* (2006) Interrelationship and familiarity of dyslexia related quantitative measures. *Ann. Hum. Genet.*, **70**, 1–16.
8. Taipale, M., Kaminen, N., Nopola-Hemmi, J., Haltia, T., Myllyluoma, B., Lyytinen, H., Muller, K., Kaaranen, M., Lindsberg, P.J., Hannula-Jouppi, K. *et al.* (2003) A candidate gene for developmental dyslexia encodes a nuclear tetratricopeptide repeat domain protein dynamically regulated in brain. *Proc. Natl Acad. Sci. USA*, **100**, 11553–11558.
9. Meng, H., Smith, S.D., Hager, K., Held, M., Liu, J., Olson, R.K., Pennington, B.F., DeFries, J.C., Gelernter, J., O'Reilly-Pol, T. *et al.* (2005) DCDC2 is associated with reading disability and modulates neuronal development in the brain. *Proc. Natl Acad. Sci. USA*, **102**, 17053–17058.
10. Schumacher, J., Anthoni, H., Dahdouh, F., Konig, I.R., Hillmer, A.M., Kluck, N., Manthey, M., Plume, E., Warnke, A., Remschmidt, H. *et al.* (2006) Strong genetic evidence of DCDC2 as a susceptibility gene for dyslexia. *Am. J. Hum. Genet.*, **78**, 52–62.
11. Cope, N., Harold, D., Hill, G., Moskvina, V., Stevenson, J., Holmans, P., Owen, M.J., O'Donovan, M.C. and Williams, J. (2005) Strong evidence that KIAA0319 on chromosome 6p is a susceptibility gene for developmental dyslexia. *Am. J. Hum. Genet.*, **76**, 581–591.
12. Paracchini, S., Thomas, A., Castro, S., Lai, C., Paramasivam, M., Wang, Y., Keating, B.J., Taylor, J.M., Hacking, D.F., Scerri, T. *et al.* (2006) The chromosome 6p22 haplotype associated with dyslexia reduces the expression of KIAA0319, a novel gene involved in neuronal migration. *Hum. Mol. Genet.*, **15**, 1659–1666.
13. Hannula-Jouppi, K., Kaminen-Ahola, N., Taipale, M., Eklund, R., Nopola-Hemmi, J., Kaariainen, H. and Kere, J. (2005) The axon guidance receptor gene *ROBO1* is a candidate gene for developmental dyslexia. *PLoS Genet.*, **1**, e50.
14. Fagerheim, T., Raeymaekers, P., Tonnessen, F.E., Pedersen, M., Tranebjaerg, L. and Lubs, H.A. (1999) A new gene (*DYX3*) for dyslexia is located on chromosome 2. *J. Med. Genet.*, **36**, 664–669.
15. Fisher, S.E., Francks, C., Marlow, A.J., MacPhie, I.L., Newbury, D.F., Cardon, L.R., Ishikawa-Brush, Y., Richardson, A.J., Talcott, J.B., Gayan, J. *et al.* (2002) Independent genome-wide scans identify a chromosome 18 quantitative-trait locus influencing dyslexia. *Nat. Genet.*, **30**, 86–91.
16. Kaminen, N., Hannula-Jouppi, K., Kestila, M., Lahermo, P., Muller, K., Kaaranen, M., Myllyluoma, B., Voutilainen, A., Lyytinen, H., Nopola-Hemmi, J. *et al.* (2003) A genome scan for developmental dyslexia confirms linkage to chromosome 2p11 and suggests a new locus on 7q32. *J. Med. Genet.*, **40**, 340–345.
17. Petryshen, T.L., Kaplan, B.J., Hughes, M.L., Tzenova, J. and Field, L.L. (2002) Supportive evidence for the *DYX3* dyslexia susceptibility gene in Canadian families. *J. Med. Genet.*, **39**, 125–126.
18. Francks, C., Fisher, S.E., Olson, R.K., Pennington, B.F., Smith, S.D., DeFries, J.C. and Monaco, A.P. (2002) Fine mapping of the chromosome 2p12-16 dyslexia susceptibility locus: quantitative association analysis and positional candidate genes *SEMA4F* and *OTX1*. *Psychiatr. Genet.*, **12**, 35–41.
19. Peyrard-Janvid, M., Anthoni, H., Onkamo, P., Lahermo, P., Zucchelli, M., Kaminen, N., Hannula-Jouppi, K., Nopola-Hemmi, J., Voutilainen, A., Lyytinen, H. *et al.* (2004) Fine mapping of the 2p11 dyslexia locus and exclusion of *TACR1* as a candidate gene. *Hum. Genet.*, **114**, 510–516.

20. Deffenbacher, K.E., Kenyon, J.B., Hoover, D.M., Olson, R.K., Pennington, B.F., DeFries, J.C. and Smith, S.D. (2004) Refinement of the 6p21.3 quantitative trait locus influencing dyslexia: linkage and association analyses. *Hum. Genet.*, **115**, 128–138.

21. Francks, C., Paracchini, S., Smith, S.D., Richardson, A.J., Scerri, T.S., Cardon, L.R., Marlow, A.J., MacPhie, I.L., Walter, J., Pennington, B.F. *et al.* (2004) A 77-kilobase region of chromosome 6p22.2 is associated with dyslexia in families from the United Kingdom and from the United States. *Am. J. Hum. Genet.*, **75**, 1046–1058.

22. Yang, Z. (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.*, **13**, 555–556.

23. Shaywitz, S.E. and Shaywitz, B.A. (2005) Dyslexia (specific reading disability). *Biol. Psychiatry*, **57**, 1301–1309.

24. Dorus, S., Vallender, E.J., Evans, P.D., Anderson, J.R., Gilbert, S.L., Mahowald, M., Wyckoff, G.J., Malcom, C.M. and Lahn, B.T. (2004) Accelerated evolution of nervous system genes in the origin of *Homo sapiens*. *Cell*, **119**, 1027–1040.

25. Khaitovich, P., Hellmann, I., Enard, W., Nowick, K., Leinweber, M., Franz, H., Weiss, G., Lachmann, M. and Paabo, S. (2005) Parallel patterns of evolution in the genomes and transcriptomes of humans and chimpanzees. *Science*, **309**, 1850–1854.

26. Enard, W., Przeworski, M., Fisher, S.E., Lai, C.S., Wiebe, V., Kitano, T., Monaco, A.P. and Paabo, S. (2002) Molecular evolution of *FOXP2*, a gene involved in speech and language. *Nature*, **418**, 869–872.

27. Duret, L. and Mouchiroud, D. (2000) Determinants of substitution rates in mammalian genes: expression pattern affects selection intensity but not mutation rate. *Mol. Biol. Evol.*, **17**, 68–74.

28. O'Brien, T.W., O'Brien, B.J. and Norman, R.A. (2005) Nuclear MRP genes and mitochondrial disease. *Gene*, **354**, 147–151.

29. Nopola-Hemmi, J., Myllyluoma, B., Haltia, T., Taipale, M., Ollikainen, V., Ahonen, T., Voutilainen, A., Kere, J. and Widen, E. (2001) A dominant gene for developmental dyslexia on chromosome 3. *J. Med. Genet.*, **38**, 658–664.

30. Ziegler, A., Konig, I.R., Deimel, W., Plume, E., Nothen, M.M., Propping, P., Kleensang, A., Muller-Myhsok, B., Warnke, A., Remschmidt, H. *et al.* (2005) Developmental dyslexia-recurrence risk estimates from a German bi-center study using the single proband sib pair design. *Hum. Hered.*, **59**, 136–143.

31. Lahiri, D.K. and Nurnberger, J.I., Jr. (1991) A rapid non-enzymatic method for the preparation of HMW DNA from blood for RFLP studies. *Nucleic Acids Res.*, **19**, 5444.

32. Miller, S.A., Dykes, D.D. and Polesky, H.F. (1988) A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res.*, **16**, 1215.

33. O'Connell, J.R. and Weeks, D.E. (1998) PedCheck: a program for identification of genotype incompatibilities in linkage analysis. *Am. J. Hum. Genet.*, **63**, 259–266.

34. Spielman, R.S., McGinnis, R.E. and Ewens, W.J. (1993) Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am. J. Hum. Genet.*, **52**, 506–516.

35. Clayton, D. and Jones, H. (1999) Transmission/disequilibrium tests for extended marker haplotypes. *Am. J. Hum. Genet.*, **65**, 1161–1169.

36. Barrett, J.C., Fry, B., Maller, J. and Daly, M.J. (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*, **21**, 263–265.