Am. J. Hum. Genet. 70:1287–1298, 2002

Evidence for Linkage and Association with Reading Disability, on 6p21.3-22

D. E. Kaplan,[1] J. Gayán,[2] J. Ahn,[1] T.-W. Won,[3] D. Pauls,[4] R. K. Olson,[5] J. C. DeFries,[5] F. Wood,[6] B. F. Pennington,[7] G. P. Page,[8] S. D. Smith,[9] and J. R. Gruen[1]

[1]Yale Child Health Research Center, Department of Pediatrics, Yale University School of Medicine, New Haven, CT; [2]University of Oxford, Oxford; [3]Biotechnology Department, Won International Patent and Law Firm, Seoul, Korea; [4]Psychiatric and Neurodevelopmental Genetics Unit, Harvard Medical School, Charlestown, MA; [5]University of Colorado, Boulder; [6]Wake Forest University School of Medicine, Winston-Salem, NC; [7]University of Denver, Denver; [8]Department of Biostatistics, University of Alabama at Birmingham, Birmingham; and [9]Munroe Meyer Institute, University of Nebraska Medical Center, Omaha

Reading disability (RD), or dyslexia, is a common heterogeneous syndrome with a large genetic component. Several studies have consistently found evidence for a quantitative-trait locus (QTL) within the 17 Mb (14.9 cM) that span D6S109 and D6S291 on chromosome 6p21.3-22. To characterize further linkage to the QTL, to define more accurately the location and the effect size, and to identify a peak of association, we performed Haseman-Elston and DeFries-Fulker linkage analyses, as well as transmission/disequilibrium, total-association, and variance-components analyses, on 11 quantitative reading and language phenotypes. One hundred four families with RD were genotyped with a new panel of 29 markers that spans 9 Mb of this region. Linkage results varied widely in degree of statistical significance for the different linkage tests, but multipoint analysis suggested a peak near D6S461. The average 6p QTL heritability for the 11 reading and language phenotypes was 0.27, with a maximum of 0.66 for orthographic choice. Consistent with the region of linkage described by these studies and others, there was a peak of transmission disequilibrium with a QTL centered at JA04 ($\chi^2 = 9.48$; empirical $P = .0033$; orthographic choice), and there was strong evidence for total association at this same marker ($\chi^2 = 11.49$; $P = .0007$; orthographic choice). Although the boundaries of the peak could not be precisely defined, the most likely location of the QTL is within a 4-Mb region surrounding JA04.

Introduction

Reading disability (RD), also known as "dyslexia," is defined as difficulty in learning to read despite adequate conventional instruction, intelligence, and sociocultural opportunity (Critchley 1970). It is the most common neurobehavioral disorder that affects children, with a prevalence rate of 10%–17.5% (Shaywitz et al. 1998), and it accounts for >80% of all learning disabilities (Lerner 1989). By use of a variety of well-defined cognitive reading tests, it is readily possible to identify RD. These tests measure a student's ability to perform the individual component processes—such as orthographic coding (OC), phonological decoding (PD), phoneme awareness (PA), and word recognition (WR)—that are thought to be involved in the whole reading process (Gayán et al. 1999). An expanding large body of evidence also indicates that RD is both familial and genetically based, with a reported heritability between 0.4 and 0.6 (Gayán and Olson 2001).

Linkage studies have identified potential RD loci on chromosomes 1 (Rabin et al. 1993), 2 (Fagerheim et al. 1999), 15 (Smith et al. 1983; Grigorenko et al. 1997; Nöthen et al. 1999; Morris et al. 2000), 18 (Fisher et al. 2002), and 6, with several consistent findings on 6p21.3-22 (Smith et al. 1991; Cardon et al. 1994, 1995; Grigorenko et al. 1997, 2000; Fisher et al. 1999; Gayán et al. 1999). In 1994, Cardon et al. (1994) published a peak of linkage surrounding D6S105. Other reported locus sizes were 13.4 cM (16.9 Mb) spanning D6S422 (pter) through D6S291 (Fisher et al. 1999), 5 cM (4.8 Mb) spanning D6S461 through D6S258 (Gayán et al. 1999), and 1.8 cM (7.9 Mb) spanning D6S299 through D6S273 (Grigorenko et al. 2000) (physical distances have been described by Ahn et al. [2001]). Overall, linkage mapping defined an RD locus that spans 17 Mb (14.9 cM) between D6S109 and D6S291. One study, with a more-specific phenotype and more-stringent ascertainment criteria, found no significant RD linkage on chromosome 6, with both qualitative and quantitative reading phenotypes (Field and Kaplan 1998; Petryshen et al. 2000). However, the overall consistency with

Received November 15, 2001; accepted for publication February 26, 2002; electronically published April 10, 2002.

Address for correspondence and reprints: Dr. Jeffrey R. Gruen, Yale Child Health Research Center, Department of Pediatrics, Yale University School of Medicine, 464 Congress Avenue, New Haven, CT 06520-8081. E-mail: jeffrey.gruen@yale.edu

0002-9297/2002/7005-0019$15.00

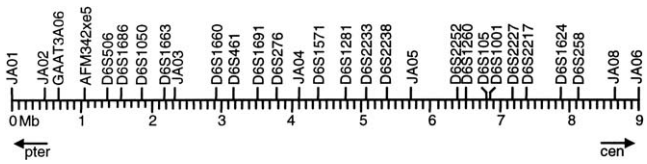

Figure 1 Location and distribution of the 29 STR markers used for linkage and association studies as described elsewhere (J. Ahn, T.-W. Won, D. E. Kaplan, E. R. Londin, P. Kuzmic, J. Gelernter, and J. R. Gruen, unpublished data).

which evidence for a putative chromosome 6 locus has been replicated is remarkable and is highly unusual for complex behavioral traits, attesting to the trait's high heritability, particularly at chromosome 6p, and to the quality of phenotypic assessments.

After genetic-linkage mapping, case-control association analysis has been especially useful for the refinement of the location of candidate genes for complex inherited disorders similar to RD. Notable examples include the gene for angiotensin-converting enzyme (*ACE*), in cardiovascular disease (Cambien 1994; Schachter et al. 1994; Yoshida et al. 2000); the gene for angiotensinogen (*AGT*), in hypertension (Hata et al. 1994); and the insulin gene (*INS*), in insulin-dependent diabetes mellitus (Julier et al. 1994). Classical association-analysis tests for differences in marker-allele frequencies in cases compared to controls that have been matched for ethnicity, race, and sex. When the alleles of a particular marker occur more frequently than would be expected by random association, the marker locus may be in linkage disequilibrium with the disease trait. Linkage disequilibrium decays with time in proportion to the recombination fraction between loci, so that disequilibrium between unlinked loci that is due to population admixture, selection, or drift decays very rapidly, whereas disequilibrium between closely linked loci decays much more slowly. Therefore, linkage disequilibrium between a marker allele and a trait locus can lead to the identification of a disease susceptibility gene very close to the marker.

Schork et al. (2001) have reported that case-control association analysis is vulnerable to overt or hidden population substructure that can increase false-positive findings. Family-based control-association designs such as haplotype relative risk (HRR) (Falk and Rubinstein 1987; Terwilliger and Ott 1992) and transmission/disequilibrium testing (TDT) (Spielman et al. 1993) have been shown to avoid these confounding effects. TDT was useful in the identification of the non-obese diabetes gene (*NOD2*) that has recently been described for Crohn disease (Hugot et al. 2001; Ogura et al. 2001). Also, using HRR and extended TDT, Morris et al. (2000) showed linkage disequilibrium between marker loci on 15q and selected RD phenotypes.

The goal of this project was to perform a very dense fine-scale linkage, transmission disequilibrium, and association study of a region on 6p21 where, previously, several groups have reported only linkage. We genotyped 104 families, with a new panel of 29 ordered STR markers distributed along the 9 Mb of the RD linkage region. Genetic linkage, heritability, effect size, and general location were assessed by Haseman-Elston and DeFries-Fulker analyses. Transmission disequilibrium (via TDT) and total association were assessed. We also characterized intermarker linkage disequilibrium across the entire marker panel independently of the phenotype data. Finally, we discuss the implications of the linkage and transmission-disequilibrium analyses and propose a likely location for the QTL.

Subjects and Methods

Sample with RD

The sample used in the current study consisted of nuclear families collected by the Colorado Learning Disabilities Research Center (CLDRC) (DeFries et al. 1997). Subjects included members of MZ twin pairs (in which case, only one member of the MZ twin pair was used), DZ twin pairs, and nontwin siblings who have previously been reported in two papers that showed the original evidence for linkage in this region. The DZ twin sample studied by Cardon et al. (1994) showed linkage with a reading composite score, and the twin and sibling samples described by Gayán et al. (1999) showed linkage with several reading and language phenotypes. This current sample consisted of 127 families. However, a number of families were uninformative because of missing phenotypes and/or genotypes. Therefore, the sample included in the analyses reported here included 104 families and 392 individuals (parents and siblings), of whom

Table 1

***P* Values for Single-Point Linkage Analyses**

	P Value[a] by			
Phenotype	HE	DFB	NHE	DFA
OC	.002	.0028		
OCH	.00097	.00049		
HCH	.0055	.0047		.03
PD	.05	.02	.0008	.0076
PA	.03	.012	.0016	.02
PTP	.002	.007	.017	.04
PDL	.02			
WR	.018	.024		.024
TWR	.0068		.0056	.0011
PWR	.059		.0045	.0074
DISC	.025		.014	.012

[a] The most significant single-point *P* values ($P < .05$ or close) in the 8.8-cM region for each phenotype are reported.

Table 2

***P* Values for Multipoint Linkage Analyses**

	P Value[a] by			
Phenotype	HE	DFB	NHE	DFA
OC	.019	.008		
OCH	.006	.016		
HCH	.004	.015		
PD		.016		
PA		.067		
PTP		.032		
PDL				
WR	.073			.039
TWR	.028		.026	.0025
PWR				.067
DISC			.047	.041

[a] The most significant multipoint *P* values (i.e., $P < .05$ or close) in the 8.8-cM region for each phenotype are reported.

221 were siblings. There were 8 families with one offspring, 79 families with two offspring, 13 families with three offspring, and 4 families with four offspring. These offspring comprised a total of 142 sib pairs who were informative for linkage analyses, of whom 117 were independent sib pairs, computed as $n - 1$ per family of n offspring.

Families were ascertained so that at least one sibling in each family had a school history of reading problems. Families were predominantly white middle-class families ascertained from school districts in the state of Colorado. Subjects for whom English was a second language were not included in the initial sample. Subjects with evidence of serious neurological, emotional, or uncorrected sensory deficits were excluded from the present analyses. The average age of the 221 siblings analyzed was 11.55 years, ranging from 8.02 to 18.53 years. This study was approved by the Human Research Committee of the University of Colorado at Boulder.

Phenotyping

Subjects were brought to the laboratory at the University of Colorado for an extensive battery of psychometric tests, which consisted of many cognitive, language, and reading tasks, including the intelligence quotient (Wechsler 1974, 1981) and the Peabody individual achievement test (PIAT) (Dunn and Markwardt 1970). Quantitative-trait data were provided for the following 11 phenotypes: OC (1) is the ability to recognize words' specific orthographic patterns and was measured here with our experimental tests for orthographic choice (OCH [2]) and homonym choice (HCH [3]); a composite score for both tests (i.e., OC composite) was created by averaging the *z* scores for both tasks. PD (4) is the oral reading of nonwords, which have straightforward pronunciations that are based on their spelling. PA (5) is the ability to isolate and manipulate abstract subsyllabic sounds in speech; for the present analyses, it was measured with our experimental phoneme-transposition (PTP [6]) and phoneme-deletion (PDL [7]) tasks, as well as with a composite score for both tests. WR (8) was measured with our experimental timed-word-recognition (TWR [9]) task and the untimed standardized PIAT-word-recognition (PWR [10]) task, which required subjects to read words aloud; a composite score for both tests was also created. Finally, the discriminant score (DISC [11]) for reading was a weighted composite of the reading recognition, reading comprehension, and spelling subtests of the PIAT. These psychometric tasks have been described in detail elsewhere (DeFries and Fulker 1985; Olson et al. 1989, 1994; DeFries et al. 1997; Gayán et al. 1999). The population average was estimated from the large twin database available at the CLDRC. After age regression and standardization, the phenotypic data for each of the reading tasks formed a continuous distribution of quantitative *z* scores, which were used in the analyses.

6p21.3-22-RD-Locus Marker Panel

For the genotyping, a new panel of 29 STR markers was used. These markers were distributed relatively uniformly through the 9 Mb (8.8 cM) of the 6p21.3-22 RD region (fig. 1). The average intermarker distance was 300 kb, with a range of 80–680 kb. The panel was comprised of 24 (CA)n and 5 (NNNN)n repeats, with an average heterozygosity of 0.73 (range 0.5–0.8) as determined in 21 whites from the Coriell Cell Repository. Optimization and multiplexing were as described elsewhere (J. Ahn, T.-W. Won, D. E. Kaplan, E. R. Londin, P. Kuzmic, J. Gelernter, and J. R. Gruen, unpublished data). The entire panel was resolved in four ABI377 gel lanes (ABI/Perkin-Elmer). Four of the markers—D6S461, D6S276, D6S105, and D6S258—were previously used to genotype a subset of these families, but those genotype data were not included in this study (Cardon et al. 1994; Gayán et al. 1999).

Genotyping

DNA was extracted from blood and buccal samples that were obtained from parents and offspring. Technicians were blinded through the assignment of random tracking numbers to the clinical samples and through predetermined PCR plate, pooling plate, and gel-lane assignments. Although all members of a single family were analyzed on the same gel, no two relatives were resolved in consecutive lanes. To avoid errors due to overflow, we staggered loading between consecutive lanes by 100 scans. In each gel, two external controls—CEPH1331-1 and CEPH1331-2—and two of nine unrelated internal controls chosen from the sample with RD were included.

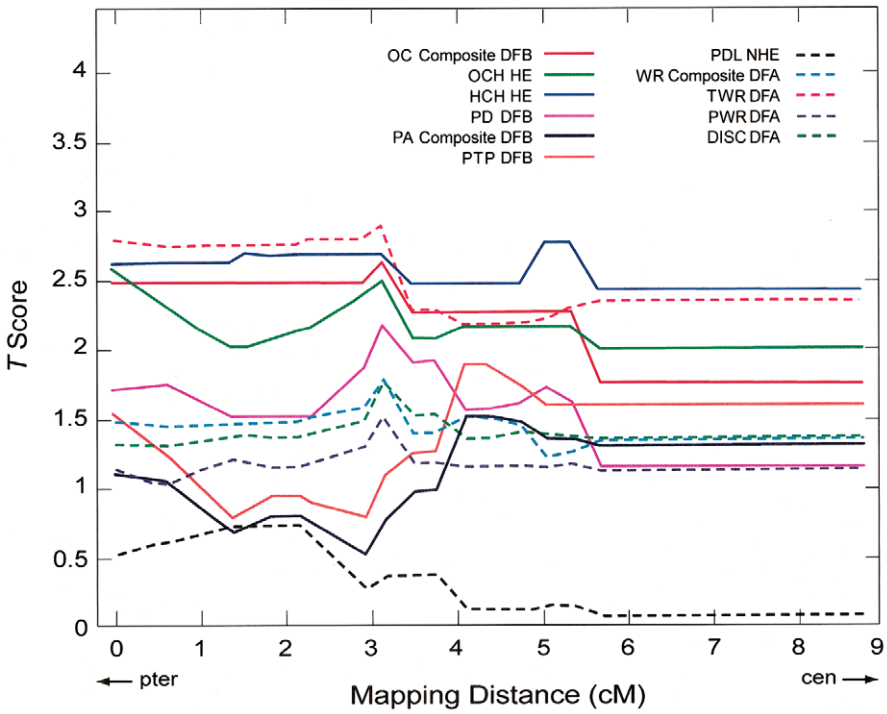

Figure 2 *T* scores across the chromosomal region for the most significant test of linkage for each phenotype. Chromosomal location (in cM) is expressed proximally from JA01.

GeneScan and GenoTyper (ABI/Perkin-Elmer) were used to track and convert ABI377-fluorescent-chromatogram data to base-pair assignments. Two technicians independently scored the output of every gel, compared the consistency of base-pair assignments by use of a Microsoft Excel macro (Allele Comparison), and resolved conflicts or flagged alleles for regenotyping. The final base-pair assignments were then ported to Genetic Analysis System software package, version 2.0 (A. Young, Oxford University, 1993–95), for allele binning and for the identification of allele-inheritance inconsistencies within pedigrees.

All the genotyping, including repeat analyses for failed initial PCRs and resolution of all disputed allele calls, was completed prior to the receipt of any phenotype data. Statistical analyses were completed in four sequential steps: (1) single-point and multipoint analyses by use of Haseman-Elston and DeFries-Fulker tests of genetic linkage, (2) TDT analyses by use of multiple models of association through quantitative transmission/disequilibrium tests (QTDTs), (3) analyses of heritability and variance components, and (4) determination of intermarker linkage disequilibrium.

Linkage Studies

By use of parental and sibling genotypic data, sib-pair identity-by-descent (IBD) status was estimated for the linkage studies performed by use of Merlin (Abecasis et al. 2002). IBD was estimated from the full set of markers in a multipoint fashion. In addition, because of the reported loss of power in the presence of genotypic or marker-map errors (Douglas et al. 2000), IBD was estimated in a single-point fashion, one marker at a time. Sib-pair phenotypic and IBD data were managed and model-free linkage analyses were run by use of a modification of the QTL macro for SAS software (Lessem and Cherny 2001). This macro performs the DeFries-Fulker and Haseman-Elston linkage analyses (Haseman and Elston 1972; Fulker et al. 1991; Elston et al. 2000). Thus, the following four linkage tests were performed: DeFries-Fulker basic (DFB), DeFries-Fulker augmented (DFA), original Haseman-Elston (HE), and new Haseman-Elston (NHE). We report empirical *P* values, since no correction was made for multiple tests.

For each phenotype analyzed, a subset of families was selected in which at least one member scored two or more SDs below the estimated population mean. This selection scheme yielded different numbers of sib pairs for each phenotype, as follows: for OC, 22 sib pairs; for OCH, 49 sib pairs; for HCH, 31 sib pairs; for PD, 68 sib pairs; for PA, 50 sib pairs; for PTP, 44 sib pairs; for PDL, 58 sib pairs; for WR, 85 sib pairs; for TWR, 75 sib pairs; for PWR, 85 sib pairs; and, for DISC, 86 sib pairs.

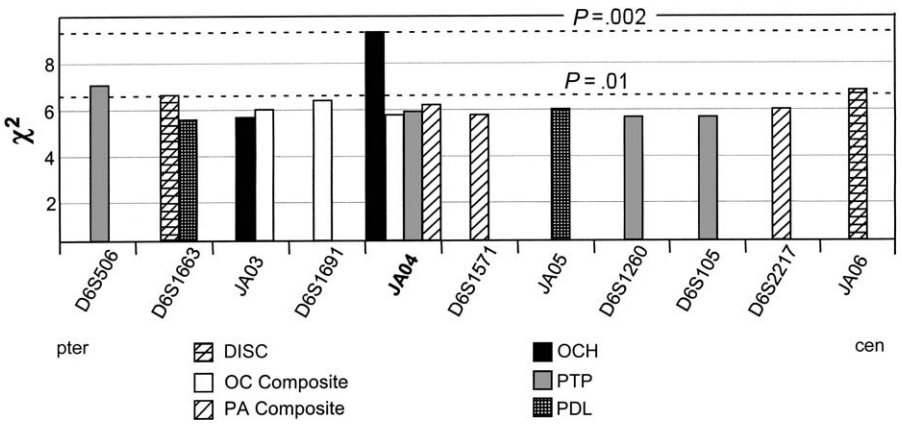

Figure 3 Significant results from QTDT output for the orthogonal model of association with environmental, polygenic, and additive major locus variance-components modeling. All results with $P < .02$ are graphed corresponding to table 3. Significance levels are shown on the basis of the nominal P values from the QTDT output.

The DFB linkage test also provides an estimate of the heritability of the QTL in a selected sample. When the data have been transformed through the expression of each score as a deviation from the mean of the unselected population and the division of each score by the difference between the affected group mean and the population mean, the regression coefficient for P provides a direct estimate of the QTL's heritability, $h_a^2(QTL)$.

Association Studies

Three association models were used to perform TDT with quantitative traits: the Allison model (Allison 1997; Allison et al. 1999), the Fulker model (Fulker et al. 1999), and the orthogonal model (Abecasis et al. 2000). Page and Amos (1999) showed that these models are valid for the computation of linkage disequilibrium in the presence of population admixture. The Allison model (i.e., test 5) computes the test statistic with only those data for which parental genotypes are available. The Fulker model uses a combined linkage and association sib-pair analysis for quantitative traits, for which the siblings serve as controls and parental genotypes are not required. The algorithm partitions the genotype score into between-family and within-family components. The within-family component is free from confounding population-substructure effects. Abecasis et al. (2000) extended this approach to create the orthogonal model that was designed to accommodate any number of offspring and optionally to include parental genotypes if available. Analyses were performed with the QTDT program (version 2.2.1; for download binaries, see the Center for Statistical Genetics Web site) by Abecasis et al. (2000), which computes χ^2 and P values for each allele (present in $\geq$30 probands) of every marker. For all of the computations within QTDT, including variance components, markers were treated as polymorphic with multiple alleles. In contrast to the linkage studies described above, inclusion was not conditioned on the basis of a cutoff for the performance score. The entire range of reading scores from all offspring in the cohort with RD was included in the association and transmission disequilibrium studies. For the Fulker and orthogonal models, empirical significance levels were calculated from 1,000 Monte Carlo permutations (9,999 permutations for JA04/allele 1). Global P values were computed using the multiallelic option and the orthogonal model of association within QTDT. The multiallelic option aggregates rare alleles with a frequency <5% in the total genotyped sample and then estimates the individual effects for all other alleles that produce a single P value for each marker-phenotype pair.

Transmission disequilibrium was estimated in the presence of linkage with the Allison, Fulker, and orthogonal models by use of variance components and $\hat{\pi}$, the estimated proportion of alleles shared IBD. Through use of variance-components modeling, both with and without association modeling, the significance of polygenic effects, as well as evidence of linkage, population stratification, and total association, was evaluated. Simwalk2 (version 2.6.0; Sobel and Lange 1996) was used to calculate IBD for the sample with RD for the association studies. Heterozygosity values for the marker panel were estimated within our sample with pedstats, a program distributed with the QTDT download.

Intermarker Linkage-Disequilibrium Analysis

The graphical overview of linkage disequilibrium (GOLD) application by Abecasis and Cookson (2000) was used to analyze intermarker linkage disequilibrium. Founder haplotypes and recombinations were estimated

Table 3

***P* Values for Single-Allele and Multiallele TDT**

	P Value[b] for					
Marker[a]	DISC	OC Composite	PA Composite	OCH	PTP	PDL
JA06	.0084					
D6S2217			.0143			
D6S105					.0177	
D6S1260					.0181	
JA05						.0141
D6S1571			.0162 (.0446)			(.0396)
JA04		.0172	.0132	.0021 (.0331)	.0154	
D6S1691		.0109				
JA03		.0142		.0169 (.0296)		
D6S1663	.01					.0195
D6S506					.0077 (.0281)	

Note.—Only markers and phenotypes with significant *P* values are reported.

[a] Markers are in order from pter, at the top.

[b] The most significant single-allele nominal *P* values ($P < .02$) are reported, corresponding with figure 3. Multiallele nominal *P* values ($P < .05$) are in parentheses.

by Simwalk2. GOLD calculates several disequilibrium statistics independently of reading phenotypic data, including Lewontin's disequilibrium coefficient, *D′*. A graphical interface summarizes and displays the pairwise linkage-disequilibrium matrix along the marker map.

Results

Genotyping

There was 99% concordance of the allele calls for the two external CEPH controls and nine internal subject controls. The average marker heterozygosity in this sample with RD was 0.70 and ranged from 0.32 (AFM342xe5) to 0.87 (JA04) and 0.9 (D6S1691). These heterozygosity values are similar to those calculated for 21 whites from the Coriell Cell Repository.

Linkage Analyses

Results of single-point linkage analyses are summarized in table 1. There is evidence for linkage with all reading and language phenotypes, including OC (by HE, $P = .002$ at D6S2238), OCH (by DFB, $P = .0005$ at D6S105), HCH (by DFB, $P = .0047$ at JA02), PD (by NHE, $P = .0008$ at D6S2217), PA (by NHE, $P = .0016$ at D6S2217), PTP (by HE, $P = .002$ at D6S1686), PDL (by HE, $P = .02$ at JA03), WR (by HE, $P = .018$ at JA03), TWR (by DFA, $P = .0011$ at D6S2238), PWR (by NHE, $P = .0045$ at D6S461), and DISC (by DFA, $P = .012$ at JA03). Note that single-point analyses reveal several peaks in this chromosomal region, and several markers are consistently significant for linkage—such as JA03 (2.31 cM; for OC, OCH, HCH, PD, PA, PDL, WR, TWR, PWR, and DISC), D6S1281 (4.77 cM; for OCH, HCH, PA, PTP, and WR), D6S105 (6.80 cM; for OCH, HCH, PD, WR, TWR, PWR, and DISC), and D6S2217 (7.40 cM; for OCH, HCH, PD, PA, PTP, WR, TWR, and DISC).

There is modest evidence for multipoint linkage with OC composite (by DFB, $P = .008$ at D6S461 [3.17 cM]), OCH (by HE, $P = .006$ at JA01 [0.0 cM], and $P = .008$ at D6S461 [3.17 cM]), HCH (by HE, $P = .004$ at D6S2233–D6S2238 [5.07–5.37 cM]), and TWR (by DFA, $P = .0025$ at D6S461 [3.17 cM]) (table 2). The multipoint linkage data presented in figure 2 show that, although it is not possible to discriminate the precise location of the QTL in this region, it is likely that the QTL is located approximately 3.17 cM from the beginning of the region, around D6S461.

Transmission-Disequilibrium Analyses

Results of association analyses are presented in figure 3 and table 3. Only markers with significant allelic transmission disequilibrium ($P \leq .02$) from the QTDT analysis by use of the orthogonal model of association with polygenic, environmental, and major additive locus variance components are shown. There is a peak of transmission disequilibrium at allele 1 of JA04 with OCH ($\chi^2 = 9.48$; $P = .0021$; empirical $P = .0033$). JA04 is located 1 Mb centromeric of D6S461, which is the lo-

Table 4

Linkage Disequilibrium JA04/Allele 1 and OCH

Model	χ^2	*P*	No. of Probands
Orthogonal	9.48	.0021	93
Fulker	9.43	.0021	63
Allison	4.98	.0257	69

Table 5

Linkage Disequilibrium JA04/Allele 1 and PTP

Model	χ^2	P[a]	No. of Probands
Orthogonal	5.87	.0154	86
Fulker	2.20	NS	57
Allison	2.55	NS	62

[a] NS = not significant.

cation of the QTL by the multipoint linkage analyses described above. Transmission disequilibrium with other phenotypes—PTP ($\chi^2 = 5.87$; $P = .0154$; empirical $P = .0310$) and the two composite scores, OC composite ($\chi^2 = 5.68$; $P = .0172$; empirical $P = .0170$) and PA composite ($\chi^2 = 6.14$; $P = .0132$; empirical $P = .0150$)—is similar to other markers in the region. Although there is evidence of transmission disequilibrium with several phenotypes at markers distributed over the entire 10 Mb, none is as pronounced as JA04/allele 1 with OCH. The results of the QTDT analyses for JA04/allele 1 and OCH by use of three different models of association—the orthogonal, Fulker, and Allison models—are presented in table 4. All three models use the environmental, polygenic, and major additive locus variance components, but, within QTDT, only the orthogonal and Fulker models can calculate empirical P values. The results of the same three association models for JA04/allele 1 and PTP are presented in table 5.

With evidence for transmission disequilibrium at JA04/allele 1, which is the most common allele in the RD cohort, we completed the multiallelic test within QTDT, to look for transmission disequilibrium for the whole marker. In these analyses, the multiallelic option aggregated the 6 alleles of JA04 that have a frequency <5% into a single allele and then estimated the individual effects for the 10 other alleles, to produce a single global P value for all 16 alleles. By use of the multiallelic option with the orthogonal model of association and variance components, evidence for transmission disequilibrium at JA04 was most significant for OCH ($\chi^2_6 = 13.72$; $P = .0330$).

QTDT was also used to evaluate population stratification by comparing between- versus within-family associations. For all alleles, at all markers, P values were >.03, showing no population stratification. We then computed total association (not a TDT) with OCH and found the peak, once again, at JA04/allele 1 ($\chi^2 = 11.49$; $P = .0007$).

Heritability Estimation by Use of Variance Components

To estimate the significance of individual variance components on the overall transmission disequilibrium, we performed a series of tests in which two alternative variance models (a null model and a full model) without any association modeling were compared. The significance of a polygenic component in the heritability of each reading phenotype was examined by the comparison of a null model that included only environmental variance, V_e, with a full model that included both environmental and polygenic variance, V_g. Heritability, h^2, was estimated as $V_g/(V_e + V_g)$. Although there was evidence for a polygenic component in the heritability of all the phenotypes, including OCH ($\chi^2 = 6.84$; $P = .0089$) and OC composite ($\chi^2 = 10.52$; $P = .0012$), the strongest evidence was for PTP ($\chi^2 = 29.12$; $P = 7 \times 10^{-8}$) and PA composite ($\chi^2 = 28.95$; $P = 7 \times 10^{-8}$). Review of the parameter estimates for these evaluations also supports a polygenic effect that is greater than the environmental effect for both PTP ($V_e = 0.208$; $V_g = 2.662$; $h^2 = 0.928$) and PA composite ($V_e = 0.275$; $V_g = 2.068$; $h^2 = 0.883$).

Variance-Components Analyses

To dissect the proportion of transmission disequilibrium that could be ascribed to linkage, we examined the significance of the variance due to an additive major locus both with and without association modeling. First, without association, a null model ($V_e + V_g$) was compared to a full model ($V_e + V_g + V_a$) that also included an additive major locus component, V_a. There was no evidence for linkage with OCH at JA04 by use of this method. However, there was weak evidence ($P < .1$) for linkage with PTP at three of the four markers with significant transmission disequilibrium: JA04 ($P = .0741$), D6S1260 ($P = .0094$), and D6S105 ($P = .0591$). With association modeling included in the analysis, there was evidence of residual linkage with PTP only at D6S1260 ($P < .03$).

Intermarker Linkage Disequilibrium

The graphical output from GOLD describes intermarker linkage disequilibrium within the sample with RD and is presented in figure 4. D' is graphed to illustrate the pairwise disequilibrium between each of the 29 markers covering the region. There is a paucity of intermarker linkage disequilibrium in the telomeric half of the marker panel. In contrast, there is significant intermarker linkage disequilibrium that begins at D6S1281 and extends toward the centromere through JA06.

Discussion

The overall goal of these studies was to characterize the RD QTL on 6p21.3-22 in terms of phenotype, heritability, and location. We used several independent analyses of data that were generated from a dense marker panel that corresponds with the reported peaks of link-

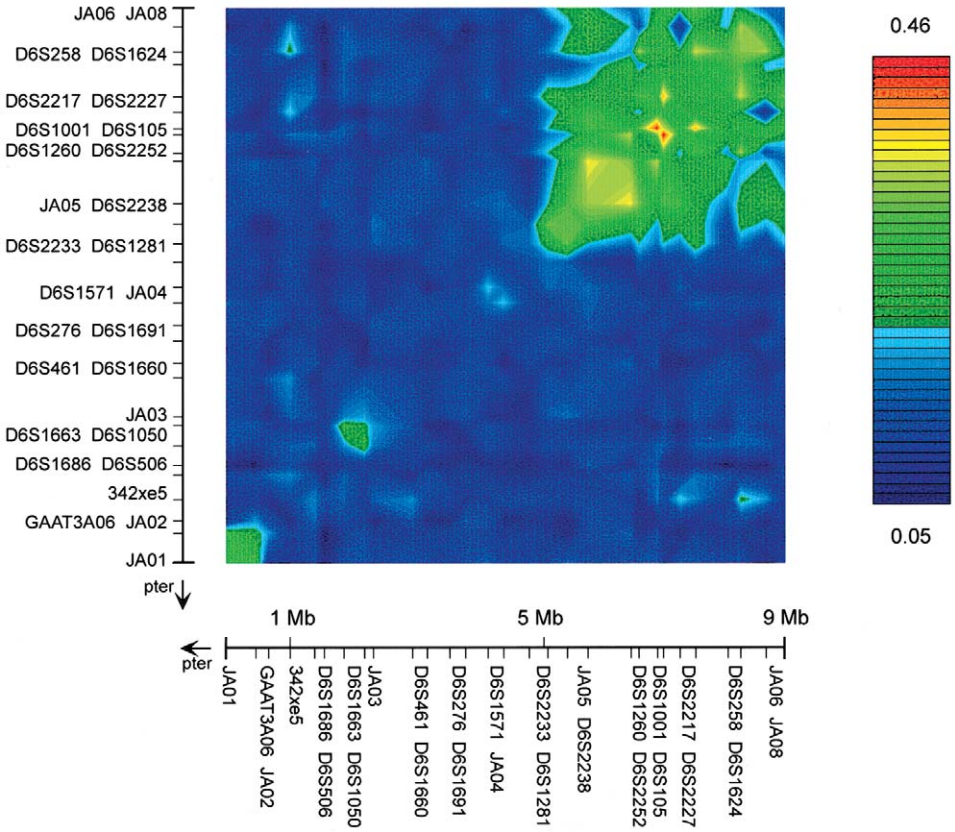

Figure 4 Graphical output from GOLD (Abecasis and Cookson 2000), showing intermarker linkage disequilibrium (*D′*) independent of reading phenotype data.

age, in order (1) to characterize further linkage to the QTL, (2) to define more accurately the location and the effect size of the QTL, and (3) to identify a peak of association within the linkage region.

Linkage Analyses

The primary aim of the Haseman-Elston and DeFries-Fulker linkage analyses was to use a high-density marker map to confirm evidence of RD linkage in a combined sample composed of two previously reported samples. Secondary aims were to establish—with more accuracy than previous reports had—the significance of the finding, as well as the location and the effect size of the QTL.

Results of the linkage analyses were consistent: there was evidence for linkage to this small 6p region for all reading and language phenotypes. Nonetheless, secondary results (i.e., significance and location) varied, depending on the phenotype and the method of analysis employed. Although linkage evidence was found for all phenotypes (except for intelligence quotient), the significance of the finding and the estimated location of the QTL varied for the different phenotypes. Linkage results also varied widely in the degree of statistical significance for the different linkage tests (HE, NHE, DFB, and DFA). The reason for this variability in results across phenotypes and tests probably lies in the nature of the phenotypes themselves and differences in the tests. It is known that certain parameters—such as the sibling correlation, the QTL genetic variance, the residual shared sibling variance, and the nonshared sibling variance—affect the power of linkage (and association) tests (Sham et al. 2000). Although the reading and language phenotypes analyzed in this study were all correlated, these parameters—especially the pattern of genetic and environmental influences (Gayán and Olson 2001)—are somewhat different for each phenotype. These parameter differences, as well as the sample size differences that arose from the selection criteria, could account for the variability of results observed. Although the families were selected for a history of reading problems in at least one sibling, this criterion did not assure a low performance in the specific reading and language tasks analyzed. The sample-size differences reflected the fact that our sample was selected for history of reading problems, which were more related to WR and DISC. Furthermore,

the small (relative to power estimates) sample analyzed could also explain, in part, the variability in results.

Linkage evidence was relatively flat along this small (8.8-cM) chromosomal region and thus could not refine a location for the QTL. Peaks, especially in the single-point analyses, were found all along the region. Nonetheless, a peak at 3.17 cM (D6S461) is apparent for many phenotypes from figure 2, which suggests this as the most likely location for the QTL.

The estimates of heritability, h_a^2(QTL), provided by the DFB linkage test, ranged from 0.12, for PDL and WR, to 0.66, for OC. For PWR and DISC, the heritability was estimated as 0, because the regression coefficient was of the sign opposite to linkage. The average heritability of the QTL for the 11 reading and language phenotypes was 0.27.

Given that the genetic correlations among these reading and language tasks are significant, ranging from 0.28 to 0.99 (Gayán and Olson 2001) and that this 6p QTL seems to affect all these variables, it is probable that this QTL explains some of the genetic covariation among these reading and language deficits. However, because of the high density and close proximity of the markers in this panel, as well as the small sample size, it is difficult to determine the exact location of this QTL by use of linkage studies alone.

Since results are presented for multiple phenotypes, a correction of statistical significance would be appropriate. However, all the phenotypic measures are correlated, and the markers are closely spaced; thus, a Bonferroni correction would be too conservative, because the assumption of independence is not applicable. As a result, we report the nominal and empirical *P* values (if possible) and sample sizes with no adjustment for multiple testing in both the linkage and association analyses.

Association Analyses

With evidence for significant linkage and genetic heritability, we performed association studies and variance-components analyses by use of TDT, aiming to refine the locus and to define the boundaries of the QTL.

Results of the association analysis were also consistent: regardless of the association model, the strongest peak of transmission disequilibrium was at JA04/allele 1 with OCH. It was also the peak for total association, and it is the only allele to show significant transmission disequilibrium with two reading phenotypes (OCH and PTP) and two related composite scores (OC composite and PA composite).

Similar to the Haseman-Elston and DeFries-Fulker linkage analyses, the variance due to environmental components, polygenic components, and an additive major gene effect was different for each phenotype. There was evidence for a polygenic effect on the heritability of all the phenotypes, but the strongest evidence was with PTP and the corresponding PA composite. Analysis of the regression parameters for the phoneme-related phenotypes showed a polygenic effect that was significantly greater than an environmental effect. Consistent with this finding, Gayán and Olson (2001), using twin data, have provided evidence for large genetic effects and smaller environmental effects for group deficits in all these reading and language skills.

The boundaries of the association peak were less clear. Only JA03, which is 2 Mb telomeric to JA04, had some minimally significant transmission disequilibrium with OCH ($\chi^2 = 5.55$; $P = .0185$), suggesting rapid decay between markers in this area. Could PTP be used to corroborate the OCH peak and also to define the boundaries? Using variance-components analyses, we asked which of the four markers that all share equivalent transmission disequilibrium with PTP—D6S506, JA04, D6S1260, or D6S105—is nearest to the QTL for this phenotype. According to Fulker et al. (1999), the disappearance of any evidence for linkage when association modeling is added in a variance-components analysis suggests that both JA04 and D6S105 may be at or in strong linkage disequilibrium with the QTL. Conversely, the evidence for variance-components linkage at D6S1260 diminished but did not completely vanish, disqualifying it as the QTL but suggesting instead that it may be in disequilibrium with the QTL. In fact, there was significant intermarker disequilibrium between D6S1260 and D6S105, as shown in figure 4 (for further discussion, see below). D6S506, the marker that shows the strongest transmission disequilibrium with PTP, had no variance-components linkage either with or without association modeling; therefore, the analysis is noncontributory for this marker. Overall variance-components analysis suggests a centromeric boundary that perhaps extends as far as D6S105.

Why does the peak of linkage, at D6S461, not include JA04? In actuality, the *T* scores for both D6S461 and JA04 are similar for most of the phenotypes in figure 2. There is also independent evidence for linkage at JA04. Both the Fulker model and the orthogonal model perform a joint analysis of linkage and association by combination of the maximum-likelihood approach and the common biometrical model for the simultaneous analysis of means and covariance matrices (Fulker et al. 1999). Therefore, the χ^2 value for JA04 shows significant transmission disequilibrium in the presence of linkage, although the size of the cohort may be too small for detection of linkage by use of variance components and other nonparametric linkage methods.

Finally, why do all the linkage curves acutely flatten at 5.5 cM, beginning at JA05 and extending toward the centromere? There were 28 recombinations in the entire sample, but none were centromeric of JA05. Similarly,

markers in this same region show intense intermarker linkage disequilibrium (fig. 4), at lengths far beyond the human-genome average of 60 kb that has been described by Reich et al. (2001). Malfroy et al. (1997) described remarkable recombination suppression that begins just 300 kb telomeric of JA05, at the *HFE* gene (near D6S1260), and extends 6 Mb centromeric to human leukocyte antigen–A, within the class I region of the major histocompatibility class. D6S105 and D6S1260, located 1 Mb centromeric of JA05, would be included in this zone of recombination suppression, perhaps thereby accounting for the linkage disequilibrium between these two markers that is discussed above. Furthermore, localized recombination suppression could also explain the uniform flattening of the linkage curves and the difficulty in definition of a precise centromeric boundary to the association peak, and it suggests that fine localization centromeric of JA05, if necessary, would require an even greater sample size.

In conclusion, there is a peak of association with an RD QTL that is centered at JA04 and that overlaps with a region of linkage described by these studies and others. The boundaries of the peak could not be precisely determined because of rapid decay in linkage disequilibrium around JA04, recombination suppression and intense intermarker disequilibrium around the centromeric markers, and the difference in linkage and the effects that variance components have on each individual phenotype. Despite these limitations, the data suggest that the most likely location of the QTL is within a 4-Mb region surrounding JA04. Additional studies with a larger cohort and an even denser marker panel could further refine the association, identify candidate genes responsible for these effects, and potentially identify the precise haplotypes that carry the susceptibility alleles.

Acknowledgments

We wish to thank Goncalo Abecasis (University of Michigan, Ann Arbor), for help with the QTDT analysis, and Elizabeth Pugh (The Center for Inherited Disease Research, Baltimore), for guidance with our high-throughput genotyping. Support for J.R.G. was through The Charles H. Hood Foundation and National Institutes of Health (NIH) grants 2P01 HD21887-11, R01 DA12849, R01 DA12690, and U01 HD98002; support for J.A. was through the March of Dimes Birth Defects Foundation and NIH grant 2P01 HD21887-11; support for T.-W.W. was through NIH grant 2P01 HD21887-11; support for D.E.K. was through NIH grant 2P01 HD21887-11; support for J.C.D. was through NIH grant P50 HD27802; and support for B.F.P. was through NIH grant MH38820.

Electronic-Database Information

The URL for data in this article is as follows:

Center for Statistical Genetics, http://www.sph.umich.edu/csg/abecasis/QTDT/download/ (for QTDT download binaries)

References

Abecasis GR, Cardon LR, Cookson WO (2000) A general test of association for quantitative traits in nuclear families. Am J Hum Genet 66:279–292

Abecasis GR, Cherny SS, Cookson WO, Cardon LR (2002) Merlin—rapid analysis of dense genetic maps using sparse gene flow trees. Nat Genet 30:97–101

Abecasis GR, Cookson WO (2000) GOLD—graphical overview of linkage disequilibrium. Bioinformatics 16:182–183

Ahn J, Won T-W, Zia A, Reutter H, Kaplan D, Sparks R, Gruen J (2001) Peaks of linkage are localized by a BAC/PAC contig of the 6p reading disability locus. Genomics 78:19–29

Allison DB (1997) Transmission-disequilibrium tests for quantitative traits. Am J Hum Genet 60:676–690

Allison DB, Heo M, Kaplan N, Martin ER (1999) Sibling-based tests of linkage and association for quantitative traits. Am J Hum Genet 64:1754–1763

Cambien F (1994) The angiotensin-converting enzyme (ACE) genetic polymorphism: its relationship with plasma ACE level and myocardial infarction. Clin Genet 46:94–101

Cardon LR, Smith SD, Fulker DW, Kimberling WJ, Pennington BF, DeFries JC (1994) Quantitative trait locus for reading disability on chromosome 6. Science 266:276–279

——— (1995) Quantitative trait locus for reading disability: correction. Science 268:1553

Critchley M (1970) The dyslexic child. Charles C Thomas, Springfield, IL

DeFries JC, Filipek PA, Fulker DW, Olson RK, Pennington BF, Smith SD, Wise BW (1997) Colorado Learning Disabilities Research Center. Learning Disabilities: a Multidisciplinary Journal 8:7–19

DeFries JC, Fulker DW (1985) Multiple regression analysis of twin data. Behav Genet 15:467–473

Douglas JA, Boehnke M, Lange K (2000) A multipoint method for detecting genotyping errors and mutations in sibling-pair linkage data. Am J Hum Genet 66:1287–1297

Dunn LM, Markwardt FC (1970) Peabody individual achievement test examiner's manual. American Guidance Service, Circle Pines, MN

Elston RC, Buxbaum S, Jacobs KB, Olson JM (2000) Haseman and Elston revisited. Genet Epidemiol 19:1–17

Fagerheim T, Raeymaekers P, Tonnessen FE, Pedersen M, Tranebjaerg L, Lubs HA (1999) A new gene (*DYX3*) for dyslexia is located on chromosome 2. J Med Genet 36:664–669

Falk CT, Rubinstein P (1987) Haplotype relative risks: an easy reliable way to construct a proper control sample for risk calculations. Ann Hum Genet 51:227–233

Field LL, Kaplan BJ (1998) Absence of linkage of phonological coding dyslexia to chromosome 6p23- p21.3 in a large family data set. Am J Hum Genet 63:1448–1456

Fisher SE, Francks C, Marlow AJ, MacPhie IL, Newbury DF, Cardon LR, Ishikawa-Brush Y, Richardson AJ, Talcott JB, Gayán J, Olson RK, Pennington BF, Smith SD, DeFries JC, Stein JF, Monaco AP (2002) Independent genome-wide scans identify a chromosome 18 quantitative-trait locus influencing dyslexia. Nat Genet 30:86–91

Fisher SE, Marlow AJ, Lamb J, Maestrini E, Williams DF, Richardson AJ, Weeks DE, Stein JF, Monaco AP (1999) A quantitative-trait locus on chromosome 6p influences dif-

ferent aspects of developmental dyslexia. Am J Hum Genet 64:146–156

Fulker D, Cardon L, DeFries J, Kimberling W, Pennington B, Smith S (1991) Multiple regression analysis of sib-pair data on reading to detect quantitative trait loci. Reading and Writing: an Interdisciplinary Journal 3:299–313

Fulker DW, Cherny SS, Sham PC, Hewitt JK (1999) Combined linkage and association sib-pair analysis for quantitative traits. Am J Hum Genet 64:259–267

Gayán J, Olson RK (2001) Genetic and environmental influences on orthographic and phonological skills in children with reading disabilities. Dev Neuropsychol 20:483–507

Gayán J, Smith SD, Cherny SS, Cardon LR, Fulker DW, Brower AM, Olson RK, Pennington BF, DeFries JC (1999) Quantitative-trait locus for specific language and reading deficits on chromosome 6p. Am J Hum Genet 64:157–164

Grigorenko EL, Wood FB, Meyer MS, Hart LA, Speed WC, Shuster A, Pauls DL (1997) Susceptibility loci for distinct components of developmental dyslexia on chromosomes 6 and 15. Am J Hum Genet 60:27–39

Grigorenko EL, Wood FB, Meyer MS, Pauls DL (2000) Chromosome 6p influences on different dyslexia-related cognitive processes: further confirmation. Am J Hum Genet 66: 715–723

Haseman JK, Elston RC (1972) The investigation of linkage between a quantitative trait and a marker locus. Behav Genet 2:3–19

Hata A, Namikawa C, Sasaki M, Sato K, Nakamura T, Tamura K, Lalouel JM (1994) Angiotensinogen as a risk factor for essential hypertension in Japan. J Clin Invest 93:1285–1287

Hugot JP, Chamaillard M, Zouali H, Lesage S, Cezard JP, Belaiche J, Almer S, Tysk C, O'Morain CA, Gassull M, Binder V, Finkel Y, Cortot A, Modigliani R, Laurent-Puig P, Gower-Rousseau C, Macry J, Colombel JF, Sahbatou M, Thomas G (2001) Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. Nature 411: 599–603

Julier C, Lucassen A, Villedieu P, Delepine M, Levy-Marchal C, Danze PM, Bianchi F, Boitard C, Froguel P, Bell J, Lathrop GM (1994) Multiple DNA variant association analysis: application to the insulin gene region in type I diabetes. Am J Hum Genet 55:1247–1254

Lerner JW (1989) Educational interventions in learning disabilities. J Am Acad Child Adolesc Psychiatry 28:326–331

Lessem JL, Cherny SS (2001) DeFries-Fulker multiple regression analysis of sibship QTL data: a SAS macro. Bioinformatics 17:371–372

Malfroy L, Roth MP, Carrington M, Borot N, Volz A, Ziegler A, Coppin H (1997) Heterogeneity in rates of recombination in the 6-Mb region telomeric to the human major histocompatibility complex. Genomics 43:226–231

Morris DW, Robinson L, Turic D, Duke M, Webb V, Milham C, Hopkin E, Pound K, Fernando S, Easton M, Hamshere M, Williams N, McGuffin P, Stevenson J, Krawczak M, Owen MJ, O'Donovan MC, Williams J (2000) Family-based association mapping provides evidence for a gene for reading disability on chromosome 15q. Hum Mol Genet 9:843–848

Nöthen MM, Schulte-Körne G, Grimm T, Cichon S, Vogt IR, Müller-Myhsok B, Propping P, Remschmidt H (1999) Genetic linkage analysis with dyslexia: evidence for linkage of spelling disability to chromosome 15. Eur Child Adolesc Psychiatry 8 Suppl 3:56–59

Ogura Y, Bonen DK, Inohara N, Nicolae DL, Chen FF, Ramos R, Britton H, Moran T, Karaliuskas R, Duerr RH, Achkar JP, Brant SR, Bayless TM, Kirschner BS, Hanauer SB, Nunez G, Cho JH (2001) A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. Nature 411: 603–606

Olson RK, Forsberg H, Wise B (1994) Genes, environment, and the development of orthographic skills. In: Berninger VW (ed) The varieties of orthographic knowledge. I. Theoretical and developmental issues. Kluwer Academic, Dordrecht, The Netherlands, pp 27–71

Olson R, Wise B, Conners F, Rack J, Fulker D (1989) Specific deficits in component reading and language skills: Genetic and environmental influences. J Learn Disabil 22:339–348

Page GP, Amos CI (1999) Comparison of linkage-disequilibrium methods for localization of genes influencing quantitative traits in humans. Am J Hum Genet 64:1194–1205

Petryshen TL, Kaplan BJ, Liu MF, Field LL (2000) Absence of significant linkage between phonological coding dyslexia and chromosome 6p23-21.3, as determined by use of quantitative-trait methods: confirmation of qualitative analyses. Am J Hum Genet 66:708–714

Rabin M, Wen XL, Hepburn M, Lubs HA, Feldman E, Duara R (1993) Suggestive linkage of developmental dyslexia to chromosome 1p34-p36. Lancet 342:178

Reich DE, Cargill M, Bolk S, Ireland J, Sabeti PC, Richter DJ, Lavery T, Kouyoumjian R, Farhadian SF, Ward R, Lander ES (2001) Linkage disequilibrium in the human genome. Nature 411:199–204

Schachter F, Faure-Delanef L, Guenot F, Rouger H, Froguel P, Lesueur-Ginot L, Cohen D (1994) Genetic associations with human longevity at the *APOE* and *ACE* loci. Nat Genet 6: 29–32

Schork NJ, Fallin D, Thiel B, Xu X, Broeckel U, Jacob HJ, Cohen D (2001) The future of genetic case-control studies. Adv Genet 42:191–212

Sham PC, Cherny SS, Purcell S, Hewitt JK (2000) Power of linkage versus association analysis of quantitative traits, by use of variance-components models, for sibship data. Am J Hum Genet 66:1616–1630

Shaywitz SE, Shaywitz BA, Pugh KR, Fulbright RK, Constable RT, Mencl WE, Shankweiler DP, Liberman AM, Skudlarski P, Fletcher JM, Katz L, Marchione KE, Lacadie C, Gatenby C, Gore JC (1998) Functional disruption in the organization of the brain for reading in dyslexia. Proc Natl Acad Sci USA 95:2636–2641

Smith S, Kimberling W, Pennington B (1991) Screening for multiple genes influencing dyslexia. Reading Writing 3:285–298

Smith SD, Kimberling WJ, Pennington BF, Lubs HA (1983) Specific reading disability: identification of an inherited form through linkage analysis. Science 219:1345–1347

Sobel E, Lange K (1996) Descent graphs in pedigree analysis: applications to haplotyping, location scores, and marker-sharing statistics. Am J Hum Genet 58:1323–1337

Spielman RS, McGinnis RE, Ewens WJ (1993) Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). Am J Hum Genet 52:506–516

Terwilliger JD, Ott J (1992) A haplotype-based "haplotype

relative risk" approach to detecting allelic associations. Hum Hered 42:337–346

Wechsler D (1974) Wechsler's intelligence scale for children, revised examiner's manual. The Psychological Corporation, New York

——— (1981) Wechsler adult intelligence scale, revised examiner's manual. The Psychological Corporation, New York

Yoshida K, Ishigami T, Nakazawa I, Ohno A, Tamura K, Fukuoka M, Mizushima S, Umemura S (2000) Association of essential hypertension in elderly Japanese with I/D polymorphism of the angiotensin-converting enzyme (*ACE*) gene. J Hum Genet 45:294–298