European Child & Adolescent Psychiatry 8:
Suppl. 3, III/47–III/51 © Steinkopff Verlag 1999

S. E. Fisher
J. F. Stein
A. P. Monaco

A genome-wide search strategy for identifying quantitative trait loci involved in reading and spelling disability (developmental dyslexia)

Simon E. Fisher (✉) · Anthony P. Monaco
The Wellcome Trust Centre
for Human Genetics
University of Oxford
Windmill Road Headington
Oxford, OX3 7BN, UK

John F. Stein
Dept. of Physiology
University of Oxford
Parks Road
Oxford, OX1 3PT, UK

Abstract Family and twin studies of developmental dyslexia have consistently shown that there is a significant heritable component for this disorder. However, any genetic basis for the trait is likely to be complex, involving reduced penetrance, phenocopy, heterogeneity and oligogenic inheritance. This complexity results in reduced power for traditional parametric linkage analysis, where specification of the correct genetic model is important. One strategy is to focus on large multigenerational pedigrees with severe phenotypes and/or apparent simple Mendelian inheritance, as has been successfully demonstrated for speech and language impairment. This approach is limited by the scarcity of such families. An alternative which has recently become feasible due to the development of high-throughput genotyping techniques is the analysis of large numbers of sib-pairs using allele-sharing methodology. This paper outlines our strategy for conducting a systematic genome-wide search for genes involved in dyslexia in a large number of affected sib-pair familites from the UK. We use a series of psychometric tests to obtain different quantitative measures of reading deficit, which should correlate with different components of the dyslexia phenotype, such as phonological awareness and orthographic coding ability. This enables us to use QTL (quantitative trait locus) mapping as a powerful tool for localising genes which may contribute to reading and spelling disability.

Key words Developmental dyslexia – sib-pair studies – allele-sharing – quantitative trait locus – semi-automated genotyping

Developmental dyslexia is a complex trait

The familial clustering of reading and spelling disability was first noted at the turn of the century (27), nearly a decade after Morgan's initial description of what was then referred to as 'congenital word-blindness' (19). Since Hallgren's large family study in 1950, there have been numerous reports supporting the role of genetic factors in the aetiology of developmental dyslexia (5, 13, 17, 18, 20), with results of a large-scale twin study suggesting a heritability of over 50% (6).

These studies have also shown that at the genetic level dyslexia is a 'complex trait', a term used to refer to any phenotype that does not display a classical Mendelian inheritance pattern which can be attributed to a single

European Child & Adolescent Psychiatry, Vol. 8, Suppl. 3 (1999)

locus (16). This breakdown of simple phenotype-genotype correspondence is due to a number of factors. Segregation analyses of reading disability (17, 18, 20) suggest that the disorder is genetically heterogeneous, meaning that different genes may influence the trait in different families. In addition, there may be reduced penetrance, where some individuals who inherit a predisposing allele do not develop the disorder, and/or phenocopy, where individuals who do not inherit a predisposing allele do manifest the trait (as a consequence of environmental or random factors). Finally, it is likely that reading disability is 'oligogenic', involving the simultaneous presence of predisposing alleles at several different loci whose interaction results in elevated risk of developing dyslexia.

Another important issue for genetic analysis of dyslexia is definition of phenotype, which may vary between studies. There remains much debate about the precise nature of the disorder, with different theories favouring visual problems, phonological difficulties, or an underlying temporal processing deficit affecting multiple modalities as core deficits in reading disability (25). Furthermore, there are age-dependent variations in penetrance and expressivity, with many dyslexics showing compensation by the time they reach adulthood.

The success of traditional parametric linkage analysis is usually dependent on specification of the correct genetic model. For complex traits such as dyslexia, the problems outlined above greatly reduce the power of such studies. As a consequence, there has been relatively limited progress in identifying consistently replicable linkages for reading disability using these parametric techniques (1, 10, 21, 26). In recent years, however, alternative strategies have been developed to deal with complex traits (16, 29), and these show greater promise for genetic analysis of dyslexia.

Strategies for identifying loci involved in developmental learning disorders

We are adopting two complementary approaches in our studies of genetics of childhood learning disorders:

Extended pedigrees

Identification of large multigenerational families with multiple affected members showing apparent simple inheritance of the trait can greatly aid the search for the gene or genes involved. For complex traits such pedigrees may be extremely rare, but can provide substantially increased power in linkage analysis. Similarly, studies may focus on particularly severe sub-types of the phenotype in question. A recent example in the field of developmental learning disorders is our work on a large three generation pedigree segregating a severe speech and language disorder (8). This family, known as KE, showed fully penetrant autosomal dominant monogenic transmission of the trait, enabling us to localise the gene responsible (termed SPCH1) to a 5.6-cM interval of chromosome 7. The results were highly significant (maximum lod score of 6.6, where a lod of 3 is the cut-off for significant linkage for monogenic inheritance) demonstrating the value of finding such families. However, further studies are needed to discover whether SPCH1 is involved in less severe and more common forms of language impairment, or only implicated in the rare sub-type found in the KE family.

Large-scale sib-pair studies

Alternatively, one can focus on the ascertainment of large numbers of sib-pairs and use allele-sharing methods to identify chromosomal regions which are likely to contain genes influencing the trait of interest. Given the significant heritability of reading disability, small nuclear families with multiple affected sibs are relatively easy to identify, and it is this strategy which we are currently employing for our dyslexia study. Sib-pair analysis does not require prior specification of parameters such as mode of transmission, penetrance or phenocopy rate, and it copes well with problems of heterogeneity or oligogenic inheritance, making it a very valuable tool for investigating complex traits (16, 29). Essentially this approach involves determination of the number of alleles (0, 1 or 2) shared IBD (identical by descent) between two sibs at a particular locus (Fig. 1). In qualitative sib-pair analysis, affected sib-pairs will show increased IBD sharing over that expected by random Mendelian segregation if the locus under investigation is in the vicinity of a gene involved in the disorder. Given genotypic data for the families it is possible to estimate the maximum likelihood (ML) values for z_0 (the proportion of sib-pairs sharing 0 alleles IBD), z_1 (those sharing only 1 allele) and z_2 (those sharing 2 alleles). A linkage statistic is obtained by comparing the likelihood of the observed data arising under these ML estimates of (z_0, z_1, z_2) to the likelihood under the expected Mendelian proportions (¼, ½, ¼). For a trait that can be defined using a quantitative measure, QTL (quantitative trait locus) mapping offers increased power (2, 3, 14, 15, 29). In this form of sib-pair analysis, siblings who share more alleles IBD should be more similar in phenotype than those sharing fewer alleles at a locus influencing the trait. If σ^2 is defined as the variance of D (the

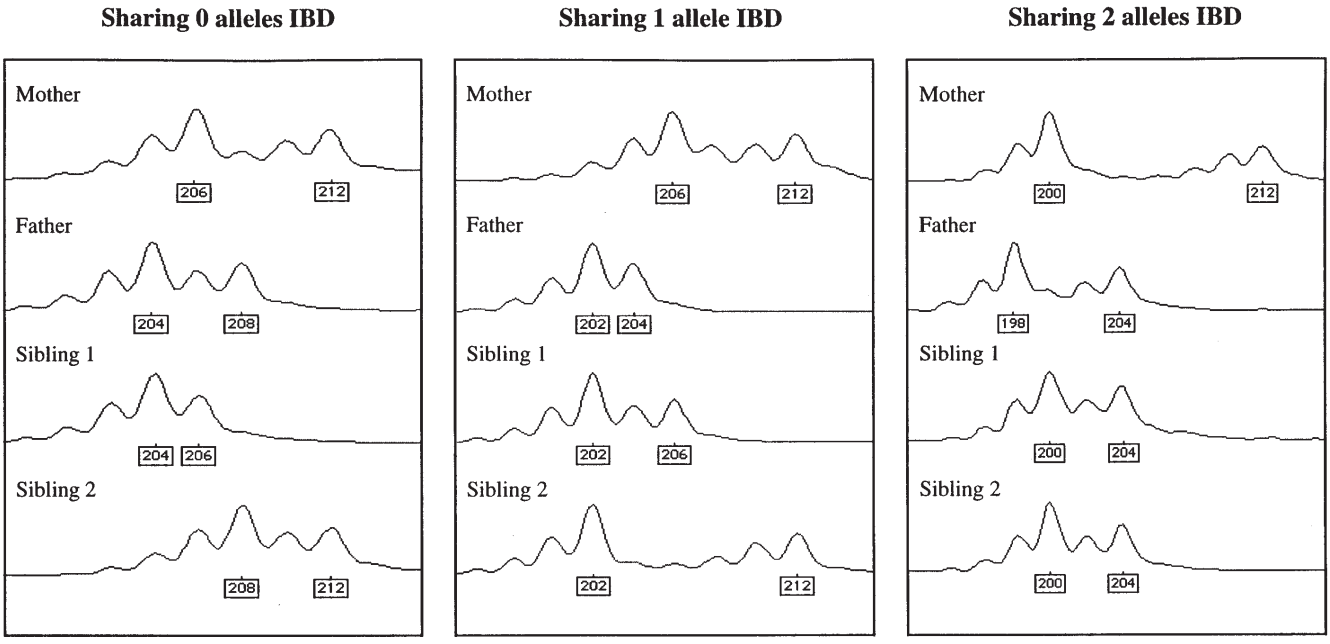

Fig. 1 Sib-pair allele-sharing detected by fluorescent genotyping. Example electropherograms from three families, illustrating the three types of IBD sharing that can be identified with the semi-automated genotyping system described in the text. These genotypes were generated using fluorescently labelled primers for PCR of a polymorphic dinucleotide repeat marker, followed by detection on an ABI 373 machine, and processing using computer software. Numbers in boxes underneath the peaks correspond to sizes (in base-pairs) of alleles, automatically called by the software. Electropherograms such as these are manually checked to confirm that allele-calling is accurate.

difference between phenotypes of the two siblings in a sib pair), then this involves testing the hypothesis that σ_0^2 (the variance for sibs sharing 0 alleles IBD) > σ_1^2 (sibs sharing 1 allele) > σ_2^2(sibs sharing 2 alleles). The original implementation of this method used regression of D^2 against expected values of sharing (14), but the approach has recently been extended to allow direct ML estimation of σ_0^2, σ_1^2 and σ_2^2 given the family data (15). In addition, for both qualitative and quantitative methods, it is now possible to perform complete multipoint analysis using information from all genetic markers to derive IBD status at each point along a chromosome (15). Allele-sharing methods for application to extended pedigrees have also been developed (29).

A systematic genome-wide search for genes involved in dyslexia

To date, all previous linkage studies of reading disability have targeted specific regions of the genome for investigation (1–3, 10, 12, 21, 24, 26), for example, on the basis of possible associations between dyslexia and autoimmune disorders (11). With recent advances in genotyping technology (see below) and the development of sib-pair methodology, a systematic genome-wide screening approach has become a feasible option (29). A two tier strategy provides an efficient way of conducting the screen. In the first stage, over 400 highly polymorphic genetic markers, spaced at ~10cM intervals throughout the genome, are used to screen a set of 100–200 sib-pairs. Evidence for linkage is assessed at each marker using the allele-sharing techniques summarised above. Chromosomal regions where linkage statistics exceed a certain threshold are the focus for further study, via genotyping of extra markers from these regions and in a replication set of an additional 100 or so sib-pairs. The first stage employs a relatively low statistical threshold in order to increase power, whilst the second stage, requiring replication in a second wave of families, is crucial for reducing the number of false positives identified by the screen.

High-throughput semi-automated genotyping

The genome screen thus requires genotyping of several hundred individuals with over 400 markers (i.e., generation of ~150 000 genotypes). The recent development of high-throughput fluorescent-based methodology has

European Child & Adolescent Psychiatry, Vol. 8, Suppl. 3 (1999)

made it possible to achieve this goal in a realistic time frame (22). The microsatellite markers for the genome scan contain tandem repeats of di-nucleotide sequence flanked by unique sequence, and are highly polymorphic due to variation in the number of repeat units (7). The polymerase chain reaction (PCR) is used to amplify across microsatellite repeats with fluorescently labelled oligonucleotide primers directed to the unique flanking sequence. PCR products are size-fractionated on a polyacrylamide gel and the fluorescent signals are detected by a laser-based system (ABI 373 or 377 machine). Computer software is then used to convert the raw data into size estimates for the PCR products generated (see Fig. 1). A key feature of this system is the ability to detect four different fluorescent dyes, with distinct emission spectra, simultaneously and independently. One dye is reserved for labelling of an internal size standard, which is included in each lane of the gel, allowing precise sizing of products for each sample. Different microsatellite markers with overlapping size-ranges can be labelled with each of the remaining three dyes, and the PCR products can be pooled, run in the same lane of the gel and detected independently, thus, increasing the capacity of each gel threefold. Furthermore, by pooling PCR products with non-overlapping size ranges, such that the entire size range of the gel is used, it is possible to run up to 24 different markers in a single lane. Therefore, this pooling system (referred to as 'multiplexing') allows the generation of over 1 000 genotypes from one 48-lane gel run. In addition, the organisation of DNA samples into 96-well microtitre plates facilitates the use of robotics for generating PCR reactions and post-PCR pooling, and the gel analysis software performs automatic size-calling of products. (This process is 'semi-automated' however, since the genotyping plots (Fig. 1) must still be checked for errors manually.)

Defining the dyslexia phenotype

Families for our study were ascertained through the dyslexia clinic at the Royal Berkshire Hospital in the UK. Probands were selected whose BAS (British Ability Scales) reading scores were ≥ 2 standard deviations below that predicted from their BAS Similarities (verbal comprehension) or Matrices (non-verbal comprehension) scores (23, 28). We then chose families where there was evidence of reading difficulty in one or more sibs of the proband (e.g., on the basis of school history or parental report). We thereby identified 90 families, comprising nearly 200 sib-pairs (many families contain more than two sibs), who are being genotyped in the first stage of the scan. We are currently collecting further families who will form the replication set for the second stage.

We have used a strict diagnostic criteria of a 2 SD discrepancy between IQ and expected reading ability to derive a categorical classification of dyslexia affection status for each sib (23, 28). This enables us to employ a qualitative sib-pair approach where we look for regions of increased sharing between affected sibs. However, dichotomisation of a complex disorder such as dyslexia is not an ideal strategy, since it discards a large amount of phenotypic detail which may be informative for genetic analysis. Therefore, regardless of affection status, all sibs are assessed with an extensive battery of psychometric tests, and the scores obtained are used for QTL mapping. This type of analysis was previously applied by Cardon et al. to demonstrate linkage to 6p21.3 in a sample of sib-pairs where one sib was reading disabled (2, 3). However, whilst Cardon et al. used a single composite discriminant score for their QTL analysis, we are investigating a number of different quantitative measures which should correlate with different components of the dyslexia phenotype. These include IQ-reading discrepancy, the BAS spelling test, 'spoonerisms' (to assess phonological awareness), as well as tests of non-word reading (which should correlate with phonological coding ability) and irregular word reading (as an index of orthographic coding ability) (4). This is comparable to a recent study of extended pedigrees by Grigorenko et al. (12), in which they fractionated the overall dyslexia deficit into several partly overlapping, but partly distinct, phenotypic definitions, and proposed that different components were linked to different chromosomal regions, on 6 and 15. In contrast to the Grigorenko study, which used categorical definitions of deficit, we are analysing the quantitative measures directly. In addition, we will be screening the entire genome, rather than targeting specific chromosomal regions.

Conclusion

Large-scale genotyping and the development of statistical methods for sib-pair analysis provide increased power for localising genes involved in reading and spelling disability. In particular we propose a QTL approach using several measures of different aspects of the dyslexia phenotype as the optimal strategy for such studies. Investigation of chromosome 6p in our UK set of families supports this conclusion (9). After completing our genome screen, it is likely that additional techniques such as linkage disequilibrium mapping (discussed elsewhere) will be

needed to refine the map position of loci identified. Over recent years there has been substantial progress in construction of comprehensive integrated maps of the genome, such that it is now possible to rapidly isolate candidate genes in regions of interest. Ultimately, identification and characterisation of genes implicated in the development of reading disability may increase our understanding and aid the treatment of this disorder.

References

1. Bisgaard ML, Eiberg H, Moller N, Niebuhr E, Mohr J (1987) Dyslexia and chromosome 15 heteromorphism: negative lod score in a Danish sample. Clin Genet 32:118–119
2. Cardon LR, Smith SD, Fulker DW, Kimberling WJ, Pennington BF, DeFries JC (1994) Quantitative trait locus for reading disability on chromsome 6. Science 266:276–279
3. Cardon LR, Smith SD, Fulker DW, Kimberling WJ, Pennington BF, DeFries JC (1995) Quantitative trait locus for reading disability on chromosome 6: a correction. Science 268:1553
4. Castles A, Coltheart M (1993) Varieties of developmental dyslexia. Cognition 47:149–180
5. DeFries JC, Fulker DW, LaBuda MC (1987) Evidence for a genetic aetiology in reading disability of twins. Nature 329:537–539
6. DeFries JC, Gillis JJ (1993) Genetics of reading disability. In: Plomin R, McClearn G (eds) Nature, Nurture, and Psychology. APA Press, Washington, DC
7. Dib C, Faure S, Fizames C, Samson D, Drouot N, Vignal A, et al (1996) A comprehensive genetic map of the human genome based on 5,264 microsatellites. Nature 380:152–154
8. Fisher SE, Vargha-Khadem F, Watkins KE, Monaco AP, Pembrey ME (1998) Localisation of a gene implicated in a severe speech and language disorder. Nat Genet 18:168–170
9. Fisher SE, Marlow AJ, Lamb J, Maestrini E, Williams DF, Richardson AJ, Weeks DE et al. (1999) A quantitative-trait locus on chromosome 6p influences different aspects of developmental dyslexia. Am J Hum Genet 64:146–156
10. Froster U, Schulte-Körne G, Hebebrand J, Remschmidt H (1993) Cosegregation of balanced translocation (1;2) with retarded speech development and dyslexia. Lancet 342:178–179
11. Geschwind N, Behan PO (1982) Left handedness: Association with immune disease, migraine and developmental learning disorder. Proc Natl Acad Sci USA 79:5097–5100
12. 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
13. Hallgren B (1950) Specific dyslexia ("congenital word blindness"): a clinical and genetic study. Acta Psychiatr Neurol 65:2–289
14. Haseman JK, Elston RC (1972) The investigation of linkage between a quantitative trait and a marker locus. Behav Genet 2:3–19
15. Kruglyak L, Lander ES (1995) Complete multipoint sib-pair analysis of qualitative and quantitative traits. Am J Hum Genet 57:439–454
16. Lander ES, Schork NJ (1994) Genetic dissection of complex traits. Science 265:2037–2048
17. Lewitter FI, DeFries JC, Elston RC (1980) Genetic models of reading disabilities. Behav Genet 10:9–30
18. Lubs HA, Rabin M, Feldman E, Jallad BJ, Kushch A, Gross-Glenn K, et al (1993) Familial dyslexia: genetic and medical findings in eleven 3-generation families. Annal Dyslexia 43:44–60
19. Morgan WP (1896) A case of congenital word-blindness (inability to learn to read) Br Med J 2:1543–1544
20. Pennington BF, Gilger JW, Pauls D, Smith SA, Smith SD, DeFries JC (1991) Evidence for major gene transmission of developmental dyslexia. JAMA 266:1527–1534
21. Rabin M, Wen XL, Hepburn M, Lubs HA, Feldman E, Duara R (1993) Suggestive linkage of developmental dyslexia to chromsome 1p34-p36. Lancet 342:178
22. Reed PW, Davies JL, Copeman JB, Bennett ST, Palmer SM, Pritchard LE, et al (1994) Chromosome-specific microsatellite sets for fluorescence-based, semi-automated genome mapping. Nat Genet 7:390–395
23. Rutter M, Yule W (1975) The concept of specific reading retardation. J Child Psych 16:181–197
24. Schulte-Körne G, Grimm T, Nöthen MM, Müller-Myhsok B, Propping P, Remschmidt H (1997) Evidence for linkage of spelling disability to chromosome 15. Am J Med Genet 74:661 (abstract)
25. Smith SD, Gilger JW, Pennington BF (1996) Dyslexia and other specific learning disorders. In: Rimoin DL, Connor JM, Pyeritz RE (eds) Principles and Practice of Medical Genetics, Churchill Livingston, New York, pp 1767–1789
26. Smith SD, Kimberling WJ, Pennington BF, Lubs HA (1983) Specific reading disability: identification of an inherited form through linkage analysis. Science 219:1345
27. Thomas CJ (1905) Congenital "word-blindness" and its treatment. Opthalmoscope 3:380–385
28. Thompson ME (1982) The assessment of children with specific reading disabilities (dyslexia) using the British Ability Scales. Br J Psych 73:461–478
29. Weeks DE, Lathrop GM (1995) Polygenic disease: methods for mapping complex disease traits. Trends Genet 11:513–519