DEVELOPMENTAL DYSLEXIA: GENETIC DISSECTION OF A COMPLEX COGNITIVE TRAIT

Simon E. Fisher and John C. DeFries‡*

Developmental dyslexia, a specific impairment of reading ability despite adequate intelligence and educational opportunity, is one of the most frequent childhood disorders. Since the first documented cases at the beginning of the last century, it has become increasingly apparent that the reading problems of people with dyslexia form part of a heritable neurobiological syndrome. As for most cognitive and behavioural traits, phenotypic definition is fraught with difficulties and the genetic basis is complex, making the isolation of genetic risk factors a formidable challenge. Against such a background, it is notable that several recent studies have reported the localization of genes that influence dyslexia and other language-related traits. These investigations exploit novel research approaches that are relevant to many areas of human neurogenetics.

PHONEMES
Individual units of speech sound that combine to make words.

**Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford OX3 7BN, UK.*
‡Institute for Behavioral Genetics, University of Colorado, Boulder, Colorado 80309-0447, USA.
Correspondence to S.E.F.
e-mail: simon.fisher@well.ox.ac.uk
doi:10.1038/nrn936

The current perception of dyslexia as a neurological syndrome with a constitutional basis dates back to the original reports, in the mid-1890s, of what was then referred to as 'congenital word blindness'[1,2]. These initial accounts viewed the disorder as the developmental analogue of acquired loss of reading ability; it was already known that neurological damage to certain areas of the brain in adults could result in selective impairment in reading and writing (alexia). As implied by the use of terms such as 'word blindness' and 'strephosymbolia' (meaning 'twisted symbols'), early explanations of dyslexia posited that a basic deficit in visual processing was at the heart of the reading difficulties of the affected subjects[3]. Such theories proposed that unstable visual representations lead to errors of letter reversal (such as 'd' substituted for 'b') and transposition (such as 'god' instead of 'dog'). It is now widely accepted that dyslexia (also known as 'specific reading disability') is better characterized as a language-related condition in which reading problems stem largely from an impairment in the representation and manipulation of PHONEMES[4]. However, there remains a lack of consensus about the exact nature of the putative 'core deficit'; indeed, some researchers doubt that there can be an adequate explanation of aetiology in terms of a single underlying process. For example, the 'double-deficit' hypothesis proposes that dyslexia results from the combined effects of two independent deficits, one involving processing of phonemes, the other involving rapid naming of simple visual stimuli (colours, objects, digits or letters)[5]. An important criticism of pure phonological-deficit models of reading disability is that they cannot account for the full range of symptoms that are experienced by people with dyslexia. These include slight but demonstrable impairments in visual[6] and auditory[7] perception, and problems with motor coordination[8]. In recent years, several alternative theories have been formulated to explain this complex phenotypic profile, some of which invoke deficits in basic neuronal mechanisms that have an impact on multiple brain modalities[6–10] (BOX 1).

So, despite decades of comprehensive multidisciplinary investigation, including studies of neuropsychology, brain anatomy, neuroimaging and magnetoencephalography (reviewed extensively elsewhere[11]), the specific causal mechanisms that underlie developmental dyslexia are still obscure. Here, we will focus on a rapidly growing area of dyslexia research that

Box 1 | The neurological basis of dyslexia — is there a single underlying cause?

Initial explanations of 'congenital word-blindness' held that significant defects in the visual system were solely responsible for the letter and word reversals that were believed to epitomize dyslexic reading. This viewpoint turned out to be untenable. Although subtle abnormalities in specific aspects of visual processing have been shown in people with dyslexia[6], these are unlikely to cause reading problems directly. Over the years, evidence has accumulated to implicate language processing. When learning to read, we develop an explicit understanding that words can be broken down into constituent phonemes, which map to visually presented letter strings, known as graphemes. Phonological-deficit theories, which have dominated the field for some years, view dyslexia as a cognitive difficulty in processing phonemes[4]. There is indeed robust evidence that phonological skills of individuals with dyslexia are compromised, but how does this fit with the complexity of the phenotype, which includes an array of subtle sensory impairments and motor difficulties? Several differing (but related) models endeavour to tackle this thorny issue. For example, rapid-processing hypotheses propose that dyslexia arises from a basic deficit in processing rapidly successive and transient stimuli that enter the nervous system, affecting all modalities[10]. In such models, the phonological impairments that are responsible for reading difficulties stem from a lower-level inability to discriminate acoustic cues that are involved in distinguishing phonemes[7]. The magnocellular deficit theory is based on data from anatomical, psychophysical and imaging studies, which indicate that many people with dyslexia have mild anomalies in the magnocellular visual subsystem[6]. Magnocells are neurons concerned with motion perception and temporal resolution, and are important for the control of eye movements. Magnocellular pathways might exist in other sensory modalities, so a multi-modal magnocell deficit might account for the full range of symptoms that are associated with dyslexia, with reading difficulties resulting from a combination of visual and phonological impairment[9]. More recently, it has been suggested that dyslexia represents a general impairment in skill automatization that results from cerebellar dysfunction[8]. The debate continues.

MENDELIAN
A trait resulting from changes in a single gene that has a significant effect on the phenotype and is inherited in a simple pattern that is similar or identical to those described by Gregor Mendel. Also referred to as monogenic.

PROBAND
Usually, the person who serves as the starting point of a genetic study.

MONOZYGOTIC
Twins that develop from a single fertilized egg cell through its division into two genetically identical parts.

DIZYGOTIC
Twins that develop during the same pregnancy as the result of two separate eggs being fertilized by two separate sperm.

HERITABILITY
The proportion of variability in a particular characteristic that can be attributed to genetic influences. This is a statistical description that applies to a specific population and might change if the environment is altered.

SPECIFIC LANGUAGE IMPAIRMENT
A significant deficit in language development in children with normal non-verbal intelligence that cannot be attributed to hearing loss, inadequate educational opportunity or obvious neurological impairment.

ATTENTION-DEFICIT/ HYPERACTIVITY DISORDER
A common disorder with childhood onset, in which persistent inattention and/or hyperactive–impulsive behaviour leads to impaired social and/or academic functioning.

CANDIDATE GENE
A gene that encodes a protein, the expected or known function of which indicates that it might be responsible for a disease or trait in a population of individuals. Pure candidate-gene approaches do not exploit or require information on chromosomal location (in contrast to 'positional cloning').

might offer a new route to elucidating the aetiology of the syndrome — the field of molecular genetics. This field has already proved to be enormously powerful in isolating causal mechanisms for numerous simple MENDELIAN disorders, and is now being applied to common, complex traits such as heart disease, diabetes, psychiatric disorders and specific learning disabilities. As we describe below, this is an exciting time for research into the genetics of dyslexia. Converging advances in phenotypic dissection, high-throughput genetic analyses and statistical methodologies have greatly improved our prospects of pinpointing genes that are involved in this and other language-related traits. We will discuss how such developments have contributed to recent successes in the field, but also highlight the limitations and assess the future potential of the molecular genetic approach for studying dyslexia.

Reading deficits are heritable

Soon after the publication of the original case studies of dyslexia at the turn of the last century, several reports noted that the condition tends to run in families[12,13]. The first large-scale family study[14] was carried out in 1950; since then, many systematic investigations have documented an increased risk of reading and spelling problems in the relatives of PROBANDS with dyslexia[15–19]. Familial clustering of a trait is consistent with the involvement of genetic factors, but could also be accounted for by environmental influences that are common to subjects within a family. The relative contributions of genetic influences and shared family environment can be dissected in twin studies. It has been shown robustly that concordance for a qualitative diagnosis of dyslexia is significantly higher in MONOZYGOTIC (MZ) twins, who have a virtually identical genetic makeup, than it is in DIZYGOTIC (DZ) twins, who (like ordinary siblings) share about half of their segregating alleles[20–22]. A large-scale study of twins with dyslexia yielded a concordance rate of 68% in MZ twins, as compared with 38% in DZ twins, indicating a substantial genetic component[23]. However, there can be intrinsic drawbacks to genetic studies of complex cognitive traits if the studies are based on all-or-none definitions of affection status. Qualitative diagnoses of dyslexia are often derived from a subject's scores on quantitative reading-related measures, which vary continuously throughout the general population (BOX 2). As such, there has been debate over whether dyslexia represents a discrete clinical entity, or simply corresponds to the extreme lower tail of normal variability in reading ability[24].

As an alternative to using dichotomous definitions of dyslexia, some genetic studies have adopted techniques that involve the direct analysis of continuous indices of severity. DeFries and Fulker developed a method (DeFries–Fulker regression) that exploits twin data to evaluate the HERITABILITY of extreme deficits in a measure of interest[25]. This is referred to as 'group heritability' (or h^2_g) to distinguish it from estimates of heritability for individual variation in the normal range of ability. The basic DeFries–Fulker technique targets one end of the distribution and involves the selection of twin pairs in which at least one member has an extreme score (that is, cases in which at least one twin performs very poorly on a continuous measure of reading ability). A statistical test is then used to assess whether the scores of co-twins regress towards the unselected population mean as a function of zygosity (DZ versus MZ). If DZ co-twins are more similar to the general population than MZ co-twins, then this points to a role for genetic factors. Direct estimates of h^2_g can be derived from such analyses. For example, in a large set of twin pairs with reading difficulties from Colorado, h^2_g was estimated to be ~50% for a composite score of overall reading performance[26]. The DeFries–Fulker regression method has had a wide impact on the field of childhood learning disorders, having been used to assess the heritability of

POLYMORPHIC GENETIC MARKERS
Naturally occurring variants in DNA sequence that can be used to track the inheritance pattern of a particular chromosomal location.

POSITIONAL CLONING
A strategy for the identification of disease genes on the basis of marker inheritance data from affected families that does not require any prior knowledge of the underlying biological pathways or gene function (in contrast to 'candidate-gene' approaches). In recent years, a blend of positional cloning and candidate-gene approaches (sometimes referred to as a 'positional-candidate' strategy) has often been used, involving the combined use of data on map location and expected gene function.

GENOTYPE
The genetic constitution of an individual. This can refer to the entire complement of genetic material or to a specific gene (or set of genes).

PHENOTYPE
The appearance of an individual in terms of a particular characteristic (physical, biochemical, physiological and so on), resulting from interactions between the individual's genotype and the environment.

PENETRANCE
The probability that an individual with a particular genotype manifests a given phenotype. Complete penetrance corresponds to the situation in which every individual with the same specific genotype manifests the phenotype in question.

PHENOCOPIES
People who manifest the same phenotype as other individuals of a particular genotype, but do not possess this genotype themselves. For example, this might occur when environmental influences alone evoke a developmental trait that has a similar genetic counterpart.

OLIGOGENICITY
When a few different genes work together to contribute to a particular phenotype.

psychometric measures in dyslexia[27], SPECIFIC LANGUAGE IMPAIRMENT[28] and ATTENTION-DEFICIT/HYPERACTIVITY DISORDER (ADHD)[29]. Moreover, such quantitative approaches, including an extension of DeFries–Fulker regression, have allowed key advances in the genetic mapping of reading disability and other language-related traits.

Genetics of dyslexia — a complex problem

In the absence of a solid understanding of the mechanisms that underlie dyslexia, there is no *a priori* reason to expect that any particular gene of known function will be a risk factor. A few theoretical accounts of trait aetiology might indicate possible subsets of genes that could be targeted for study (such as the controversial immune-disorder hypothesis that we discuss below). However, in general, there are no compelling cases that would sufficiently limit the genetic search to make a pure CANDIDATE GENE approach cost-effective. An alternative strategy is to track the inheritance of different chromosomal regions in families that segregate dyslexia, to map the location of putative genetic risk factors. This technique (linkage mapping) assesses whether POLYMORPHIC GENETIC MARKERS in particular genomic regions are 'linked' to the trait of interest. Sufficient reliable linkage data might highlight small areas of the genome that contain a manageable subset of genes. Further investigation of such genes should then allow the identification of specific gene variants in those regions that are involved in trait susceptibility. In the past 25 years, POSITIONAL CLONING has become an extremely fruitful research strategy for the investigation of monogenic disorders, exploiting the simple inheritance patterns that are observed for such traits.

With a few rare exceptions, the transmission of dyslexia and other language-related traits in affected families tends to be complex; there is no straightforward correspondence between a subject's genetic makeup (GENOTYPE) and his or her cognitive abilities (PHENOTYPE)[30,31]. For example, a family might contain one or more individuals who inherit a high-risk genotype but do not develop problems (cases of incomplete PENETRANCE). Conversely, there might be subjects who are clearly affected, even though they have a low-risk genotype (PHENOCOPIES). Genotype–phenotype concordance is further eroded by heterogeneity — distinct genetic loci implicated in different families — and OLIGOGENICITY — allelic variants at multiple loci contributing to increased risk. In combination, these factors severely limit the POWER of traditional linkage mapping, which assumes single-gene inheritance and relies on the precise specification of transmission pattern, penetrance levels and phenocopy rates.

The problems that are associated with genetic complexity are further compounded by constraints at the phenotypic level. Delineation of the dyslexia phenotype for genetic studies is restricted by a lack of consensus as to the physiological, behavioural and cognitive correlates of the disorder[30]. As a consequence, different investigations into genetic aetiology have generally used distinct diagnostic tools and classification criteria. Sometimes, this is the inevitable outcome of language differences; although most genetic linkage studies have involved English-speaking cohorts[32–37], investigations have also been conducted in Danish[38], German[39,40], Norwegian[41] and Finnish[42] families. But regardless of the native language, operational definitions of dyslexia have varied markedly, yielding increased heterogeneity. This could raise questions when trying to interpret and integrate data from multiple studies. For example, subjects that have 'phonological coding dyslexia' in one study sample[36] might not be directly comparable to those reported to have 'spelling disability'[40] or 'reading/writing disability associated with severe speech delay'[39] in others.

Box 2 | Defining dyslexia for genetic studies

What is dyslexia? A standard answer would be something like "dyslexia is a specific, significant impairment in reading ability that is not explained by deficits in general intelligence, opportunity, motivation or sensory acuity". The deceptive simplicity of this definition breaks down as soon as one examines it in detail. How do we decide what constitutes 'specific, significant' impairment in a subject's reading? What level of intellectual deficit would be considered adequate to 'explain' poor literacy? Do subtle abnormalities in auditory or visual processing that are only detectable in experimental situations (BOX 1) represent deficits in 'sensory acuity' that would invalidate a positive diagnosis?

Clinical diagnoses of dyslexia usually derive from applying thresholds to psychometric measures that are normally distributed in unselected populations. A commonly used definition requires 'significant' (say, –2 standard deviations) discrepancy between observed reading ability (assessed by standardized word-recognition tests) and that expected on the basis of IQ[84]. The cogency of diagnosis on the basis of IQ–achievement discrepancy has been challenged. For example, the measured IQ of dyslexic children declines with age, and is closely related to socioeconomic status, so that children who are older or of lower socioeconomic status are less likely to be diagnosed by discrepancy criteria[85]. Alternative methods that make no assumptions about IQ–achievement relationships, such as those requiring a significant lag in reading age[86], are also flawed. A shortcoming of most classification schemes is the use of thresholds, which are established in an arbitrary manner; the question of whether dyslexia is a pathological condition or represents the tail of a normal curve remains unresolved[24,87]. Furthermore, psychometric profiles can vary greatly among people with dyslexia and at different developmental stages. Adolescents and adults with dyslexia can 'compensate'; they seem to have normal word-recognition skills, but the underlying deficits persist. These deficits can be shown with appropriate tests, such as those that tap spelling, reading rate or phonological skills[88]. Therefore, the choice of diagnostic measure can be crucial. In many situations, clinical all-or-none diagnoses of dyslexia might not be optimal for genetic research, as they do not capture the complex essence of the phenotype[30].

Table 1 | Linkage methods that have been used to investigate dyslexia

Method	Phenotype	Sample	Advantages	Disadvantages	References*
Traditional parametric (model based)	Qualitative	Extended families with multiple affected members	Most powerful method for detecting linkage in pedigrees with simple inheritance, if genetic model is correctly specified	Depends on accurate specification of mode of inheritance, penetrance, phenocopy etc.; assumes monogenic transmission; large families with simple inheritance patterns are rare; can be limited by dichotomous classification of complex trait	32,41,42
Allele-sharing nonparametric (model free)	Qualitative	Extended or nuclear families with multiple affected members	No assumptions about mode of inheritance, penetrance, phenocopy etc.; nuclear families easy to collect	Large sample sizes needed to yield sufficient power; some methods sensitive to specification of allele frequencies; can be limited by dichotomous classification of complex trait	35,40–42
Basic Haseman–Elston regression	Quantitative	Phenotyped sib-pairs	No assumptions about mode of inheritance etc.; sample easy to collect; exploits continuous nature of trait; simple to implement	Large sample sizes needed to yield sufficient power; does not exploit all available trait variability; difficult to accommodate multiple sib-ships	37,63
DeFries–Fulker regression	Quantitative	Phenotyped sib-pairs; extreme proband	No assumptions about mode of inheritance etc.; sample easy to collect; exploits continuous nature of trait; simple to implement; well suited to selected samples	Large sample sizes needed to yield sufficient power; does not exploit all available trait variability; difficult to accommodate multiple sib-ships; no *a priori* basis for choosing level of selection	33,34,50–52
Variance-components partitioning	Quantitative	Phenotyped sib-pairs or extended families	No assumptions about mode of inheritance etc.; sample easy to collect; exploits most of observed variability in trait; incorporates all pedigree members simultaneously	Large sample sizes needed to yield sufficient power; computationally intense; tests of significance assume multivariate normality	37,50

*Key examples of dyslexia linkage studies that have successfully applied each of these methods. Note that the terms 'nonparametric' or 'model free' are not entirely accurate; although such approaches do not depend on the previous specification of penetrance, phenocopy or transmission, they do sometimes involve the estimation of certain parameters and/or rely on various assumptions about the genetic model. Nevertheless, nonparametric/model-free methods are much less restrictive than the fully parametric approach.

This leads us to another area of controversy. Is dyslexia a single trait or a cluster of related subtypes with distinct aetiologies (which are likely to involve different subsets of genes)? Castles and Coltheart proposed the existence of two forms of dyslexia that are analogous to subtypes that were formerly documented in alexia cases[43]. They based their classification scheme on the idea that skilled readers can use two discrete routes for decoding text, one involving processing of the individual phonemes that make up a word, and the other exploiting direct recognition of whole-word letter patterns (orthography) without apparent need for phonological mediation. Psychometric testing of children with dyslexia identified some individuals who seemed to have selective deficits in either the phonological or the orthographic route. Castles and Coltheart defined these as cases of phonological and surface dyslexia, respectively[43]. Although the underlying assumption of a 'dual-route' reading model has been criticized, twin studies of phonological and surface subgroups have indicated greater heritability of reading deficits for the former, whereas shared environment is important for the latter[44], supporting the idea of divergent aetiologies. But although a significant number of poor readers fit the characteristics of either proposed subtype, most cases have difficulty with both phonological and orthographic tasks (as discussed further below).

A final problem for phenotypic definition is that the nature and severity of deficits might vary at different developmental stages of the life of the person with dyslexia. This troublesome issue is usually disregarded by molecular studies (BOX 2). It could be addressed in the future by obtaining longitudinal data at multiple time points from each subject in a study.

Mapping genes for dyslexia

In recent years, innovations in three areas have contributed to success in localizing genes for dyslexia and other language-related traits. These are QUANTITATIVE TRAIT LOCUS (QTL) mapping, phenotypic dissection and high-throughput genome-wide scanning. Most current linkage studies of dyslexia use one or more of these to facilitate gene mapping.

QTL mapping. As discussed above, analyses of continuous indices of severity in twins have been important for assessing genetic contributions to reading and language deficits[21,22,27,28]. Direct use of the same quantitative measures in combination with molecular genetic data provides a strategy for localizing potential risk factors to particular chromosomal regions. QTL mapping is one form of what are often referred to as 'nonparametric' or 'model-free' linkage methods (TABLE 1). These tend to be more suitable for complex genetic traits than 'parametric' linkage methods, as they do not rely on assumptions of monogenic inheritance, estimates of penetrance levels, phenocopy rates and gene frequencies, or the precise specification of transmission (recessive, dominant, sex-linked and so on)[45]. Furthermore, they can better handle unknown levels of heterogeneity and oligogenicity. The trade-off for nonparametric techniques is that they usually require very large data sets (several hundred nuclear families) to yield sufficient power for gene mapping[45].

POWER
The probability of correctly rejecting the null hypothesis when it is truly false. For linkage studies, the null hypothesis is that of 'no linkage', so the power represents the probability of correctly detecting a genuine linkage.

QUANTITATIVE TRAIT LOCUS (QTL). A genetic locus or chromosomal region that contributes to variability in a complex quantitative trait (such as body weight), as identified by statistical analysis.

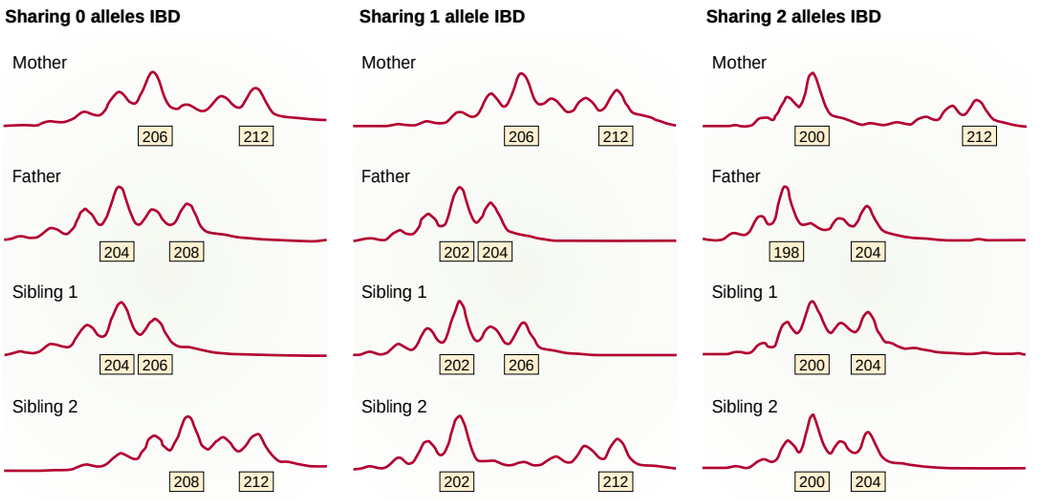


Figure 1 | **IBD allele sharing can be assessed using polymorphic genetic markers.** Examples of genotype data from three nuclear families, showing the three types of identical-by-descent (IBD) allele sharing (0, 1 or 2). Genotypes such as these are generated using high-throughput fluorescence-based genotyping technology[56]. Numbers in boxes underneath the peaks correspond to sizes (in base pairs) of the alleles, automatically called by genotyping software. IBD estimates derived from such genotype data are subsequently used for linkage analysis. Adapted, with permission, from REF. 56 © 1999 Steinkopff Verlag.

Typically, such methods rely on estimating whether related subjects have inherited identical copies of a polymorphic genetic marker from a common ancestor (FIG. 1). For example, at any particular marker, a pair of siblings will share, on average, 50% of their alleles identical-by-descent (IBD), owing to random segregation. Qualitative nonparametric approaches test whether a chromosomal region shows elevated IBD allele sharing in related subjects who are concordant for a disorder. However, they retain one potential limitation of other qualitative approaches in adopting all-or-none classifications of affection status. Quantitative nonparametric approaches (QTL methods) evaluate whether there is a significant correlation between genetic similarity (indexed by IBD allele sharing) and phenotypic resemblance (assessed through a comparison of quantitative scores) for related people in the chromosomal region of interest. As they directly exploit additional phenotypic information that is available from quantitative data, QTL approaches often have advantages over qualitative nonparametric approaches, assuming that the measures are reliable and accurate indices of severity[30,31]. However, as outlined in TABLE 1, the alternative strategies that are available for phenotypic definition and statistical analysis might have different strengths and weaknesses depending on the nature of the study sample.

Simple implementations of QTL mapping use regression analysis in sib-pairs to assess phenotype–genotype relationships (FIG. 2). The original Haseman–Elston method regresses the square of the difference in siblings' phenotypic scores against the number of alleles shared IBD at a particular marker[46]. Alternatively, an extension of DeFries–Fulker regression is ideal for investigating samples selected from one tail of a normal distribution, which can offer increased power for detecting genetic effects. The DeFries–Fulker linkage method assesses whether co-sibs of individuals with extreme phenotypic scores regress towards the unselected population mean as a function of IBD at the marker under investigation[47]. Regression-based methods are straightforward to apply (*t*-tests of appropriate regression coefficients yield estimates for significance of linkage), but they do not exploit all the available trait information. Moreover, there is disagreement over how best to accommodate nuclear families that contain multiple sib-ships, in which alternative pairings of sibs are not fully independent, or more complex extended pedigrees.

A complementary approach that is based on variance components simultaneously evaluates all relationships in a family and makes use of almost all observed phenotypic variability, but is computationally intense[48]. Trait variability is partitioned into components due to major-gene, unlinked POLYGENIC and residual environmental factors, using maximum-likelihood estimation. To assess linkage, the likelihood of the data under the null hypothesis (no major-gene effect) is compared with that when the major-gene component is unconstrained (FIG. 2). Nominal estimates of significance, which are derived from likelihood-ratio tests, are based on an assumption of multivariate normality. This is likely to be invalid for many data sets, particularly selected samples, and might lead to false-positive evidence for linkage (increased type-I error) or reduced power to detect a real effect (increased type-II error)[49]. Various options exist for surmounting this problem, including the use of simulations to derive empirical-based estimates of significance[50]. Haseman–Elston, DeFries–Fulker and variance-components methods have been used successfully to localize putative dyslexia risk factors, such as that on chromosome 6p (REFS 33,34,37,50,51).

POLYGENIC
The effects of a large number of different genes, each of which has a slight influence on the phenotypic outcome.

GRAPHEME
A written symbol, or group of symbols, that is used to represent a specific phoneme.

MULTIPOINT ANALYSIS
The use of data obtained from multiple neighbouring genetic markers on the same chromosome to extract linkage information at many points across a genomic region.

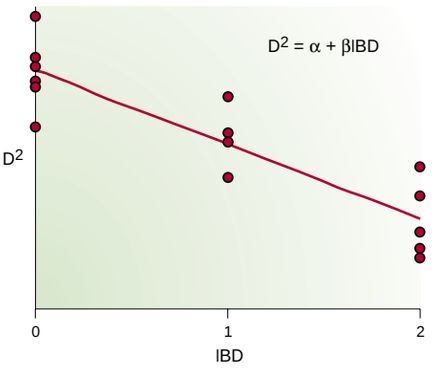
SINGLE-POINT ANALYSIS
The investigation of linkage at one point on a chromosome, using data from a single marker.

LOD SCORE
Linkage mapping involves comparing two likelihoods. The first is the likelihood of the data, under the hypothesis that there is linkage between inheritance of the trait and that of the chromosomal region in question. The second is the likelihood of the data, under the null hypothesis that there is no linkage. The lod score is the logarithm of the likelihood ratio; if it exceeds a given threshold, the null hypothesis can be rejected.

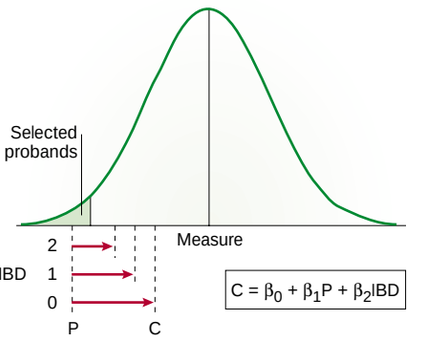
Phenotypic dissection. Many genetic studies of dyslexia have focused on what might be referred to as 'global' deficit (see BOX 2), using general diagnoses or quantitative analyses of overall indices of severity (for example, on the basis of scores on standardized tests of word recognition or spelling ability). Recently, there has been a move towards complementary approaches in which the dyslexia profile is dissected into distinct but related phenotypic components[35,37,50–52]. This dissection is driven by theories about the nature of the reading process, but the validity of using such hypothetical components is well supported by cognitive-psychological and psychometric studies. Tests have been developed that are believed to tap predominantly each putative component. For example, phoneme awareness, defined as the capacity to reflect explicitly on the individual elements of speech, can be assessed with oral tasks that do not involve the visual processing of print. These might include tests requiring phoneme deletion[51] ("say 'prot' without the 'r' sound" — 'pot') or the construction of 'spoonerisms'[50] ("swap around the first sounds of these two words: 'cat sad'" — 'sat cad'). Phonological decoding, the ability to convert written GRAPHEME units into their corresponding phonemes, is usually evaluated through oral reading of pronounceable words that lack real meaning (non-words), such as 'teg' or 'latsar'[43]. Recognition of whole-word orthography (orthographic coding) can be measured through the oral reading of words that violate standard letter-sound conventions of English, such as 'meringue' and 'yacht'. These irregular words cannot be read correctly using phoneme–grapheme conversion rules, so success should principally reflect orthographic-processing ability[43]. Orthographic skills can also be assessed using forced-choice tasks that require rapid recognition of a correctly spelt target word versus a phonologically identical non-word ('rain' versus 'rane')[51]. Another ability that seems to be impaired in many people with dyslexia involves the rapid serial naming of visual stimuli (rapid automized naming)[5,35,52].

The relationship between hypothetical components and their relative importance for reading and spelling problems in dyslexia remains an active area of study[27,52–55]. Twin studies indicate that group deficits in quantitative measures of phonological, orthographic and rapid-naming skills of people with dyslexia are all significantly heritable[27,52]. It is important to realize that the components are overlapping, not independent; inter-trait correlations between reading- and language-related measures are usually moderate to high[27,55]. Furthermore, analysis with a bivariate extension of the DeFries–Fulker regression indicates that a substantial proportion of the observed covariance is due to genetic factors that are common to all components[27,52]. However, such investigations also indicate the existence of genetic effects that uniquely influence the independent variance of different measures (for example, phonological versus orthographic)[27]. There might even be distinct genetic influences acting on accuracy and latency deficits that can be observed for the same component[27].

a



b



c

	Total variance			Log likelihood
	Major gene	Polygenic	Environment	
Null hypothesis	0.00	0.62	0.38	L(0)
Unlinked locus*	0.00	0.62	0.38	L(A)
Linked locus*	0.33	0.29	0.38	L(A)

*Alternative hypothesis

Chi-square = $-2 \times (L(0) - L(A))$

Figure 2 | **Methods for QTL-based linkage mapping in humans. a** | The basic Haseman–Elston method evaluates the relationship between differences in siblings' scores (D) and their identical-by-descent (IBD) allele status. A *t*-test of regression coefficient β yields an estimate of significance[46]. **b** | DeFries–Fulker regression requires the selection of 'probands' (P), followed by an assessment of whether co-sibs (C) regress further towards the unselected mean as IBD status decreases. A *t*-test of regression coefficient β_2 yields an estimate of significance[47]. **c** | Variance-components analysis involves partitioning the total variability into major-gene, polygenic and environmental factors[48]. Under the null hypothesis, the likelihood of the data is maximized, with the major-gene component constrained at 0. Under the alternative hypothesis, the likelihood of the data is maximized without this constraint. If there is no linkage between trait variability and IBD status at the locus in question, then the major-gene effect under the alternative hypothesis remains 0. L(0) and L(A) are log likelihoods of the data under the null hypothesis and the alternative hypothesis, respectively. Evidence for linkage is assessed by a likelihood-ratio test. This provides a valid test of linkage significance (given certain assumptions[49]). However, when analysing the top results from a genome scan, the maximum-likelihood values of the components under linkage are not themselves meaningful and will give biased estimates of effect size[78]. The figures shown here are used simply to illustrate the approach and are not taken from real analyses.

Table 2 | Targeted linkage studies of chromosome 15 in developmental dyslexia

Authors (year)	Families (country)	Treatment of phenotype	Linkage method	Summary of findings	Reference
Smith *et al.* (1983)	9 (USA)	Qualitative, global	Parametric	Lod of 3.24 with marker *cen15* (chromosome 15 centromeric heteromorphism)	32
Bisgaard *et al.* (1987)	5 (Denmark)	Qualitative, global	Parametric	Exclusion of linkage to *cen15* (lod of –3.42)	38
Smith *et al.* (1991)	19 (USA)	Qualitative, global Quantitative, global	HE* HE	$P = 0.009$ for RFLP marker *ynz90*, mapping between 15q15 and 15qter $P = 0.03$ for RFLP marker *ju201*, mapping between 15q15 and 15qter	63
Fulker *et al.* (1991)	19 (USA)	Quantitative, global	DF	$P < 0.005$ for markers *ynz90* and *ju201*	64
Rabin *et al.* (1993)	9 (USA)	Qualitative, global	Parametric	Exclusion of linkage to multiple RFLP markers in proximal region of 15q	62
Grigorenko *et al.* (1997)	6 (USA)	Qualitative, components	Parametric Nonparametric	Lod of 3.15 with microsatellite marker *D15S143* in 15q21 for single-word reading; no report of linkage to phoneme awareness, phonological decoding, rapid automized naming or IQ–reading discrepancy Generally negative; a significant result was said to be obtained for *D15S128* in 15q11, but no details were presented	35
Schulte-Körne *et al.* (1998)	7 (Germany)	Qualitative, global	Parametric Nonparametric	Multipoint analyses gave $P = 0.0042$ at *D15S132* in 15q21 Multipoint analyses gave $P = 0.03$ at *D15S143*	40

All studies initially identified extended families on the basis of multiple affected individuals in different generations, but the subsequent analyses varied in several respects. Qualitative affection status or quantitative measures of deficit were used; most studies adopted a global assessment of the phenotype, whereas one study (that of Grigorenko *et al.*, 1997) dissected the phenotype into hypothetical components. The linkage method was either parametric model-based linkage analysis or nonparametric linkage analysis of allele sharing in affected relatives. DF, DeFries–Fulker regression analysis of selected sib-pairs; HE, Haseman–Elston regression analysis of sib-pairs; RFLP, restriction fragment length polymorphism. *Haseman–Elston analysis can be performed for a qualitative all-or-none diagnosis by treating this as a measure where 0 = unaffected and 1 = affected.

Genome-wide scanning. The most thorough way of identifying genetic linkage involves a systematic search of all chromosomes, which requires the analysis of several hundred polymorphic markers in numerous subjects[56]. Using MULTIPOINT methods, it is possible to infer the IBD status of chromosomal intervals between markers, allowing the assessment of linkage at virtually all points across the genome (in contrast to SINGLE-POINT analyses, which evaluate linkage at each marker in isolation)[57]. Genome-wide searches used to be prohibitively labour-intensive and time-consuming, beyond the capabilities of many laboratories. After the development of high-throughput genotyping technologies, they now represent a standard tool for the analysis of both monogenic and complex traits[56]. For dyslexia and other language-related traits, genome-wide scans have been carried out either in single extended pedigrees[41,42,58] or in large samples of sib-pairs[50,59]. These scans involve analyses of multiple independently segregating genomic regions. The concomitant multiple testing results in a substantial increase in the risk of identifying false-positive results (linkage observations that are due to chance, rather than real aetiological effects). It is therefore an accepted procedure to adopt stringent thresholds for declaring the identification of significant linkage. LOD SCORES or *P*-values are the common currencies for describing strength of linkage results. Traditionally, a lod score of 3 is deemed to be 'significant' in parametric analysis of monogenic pedigrees, but it has been argued that a threshold of 3.6 is more suitable for sib-pair allele-sharing methods[60]. This threshold can be shown to correspond to a nominal *P*-value of 0.00002, constituting a strict cut-off to guard strongly against false-positive linkages. Note that many genome-wide scans of complex traits do not yield such strong results[61], and that, even if they do, a proportion of 'significant' linkages will still turn out to be chance events, so replication in independent studies remains crucial[60].

CHROMOSOMAL HETEROMORPHISM
Natural variation in the shape or staining pattern of a chromosome, as viewed under the microscope.

CENTROMERE
The constricted region of a chromosome that includes the site of attachment to the mitotic or meiotic spindle. Geneticists divide the chromosome into 'short' and 'long' arms, which are separated by this centromere.

CENTIMORGAN
A standard measure of genetic distance that is derived from observations of recombination between neighbouring loci. The relationship to actual physical distance along a chromosome varies throughout the genome; on average, 1 centimorgan corresponds to around one million bases of DNA.

Targeted linkage studies of dyslexia

Several molecular genetic studies of dyslexia have focused on specific chromosomal regions or have screened a limited proportion of the genome with some success. In 1983, Smith *et al.* carried out the first linkage study of dyslexia, investigating the handful of CHROMOSOMAL HETEROMORPHISMS and protein polymorphisms that were available at the time[32]. A parametric analysis of extended families with three-generation histories of reading disability yielded significant linkage to the chromosome 15 CENTROMERE, originating mainly from one family (TABLE 2; FIG. 3). Subsequent reports could not support the initial findings[38,62]. One investigation re-analysed the family that had given strongest evidence in the original study, but used more highly polymorphic DNA-based markers. The new data excluded linkage to the centromere and neighbouring regions of 15q in this and other families[62]. In 1991, Smith and co-workers reported a follow-up to their 1983 study, involving nonparametric analyses of new markers in an expanded sample of kindreds[63]. Although linkage was observed to markers on 15q, these mapped 90–120 CENTIMORGANS from the original site of interest[63,64]. Grigorenko and colleagues[35] investigated qualitatively defined component phenotypes in extended families for this region. They observed significant linkage with single-word reading at one marker, *D15S143* in 15q21, when using

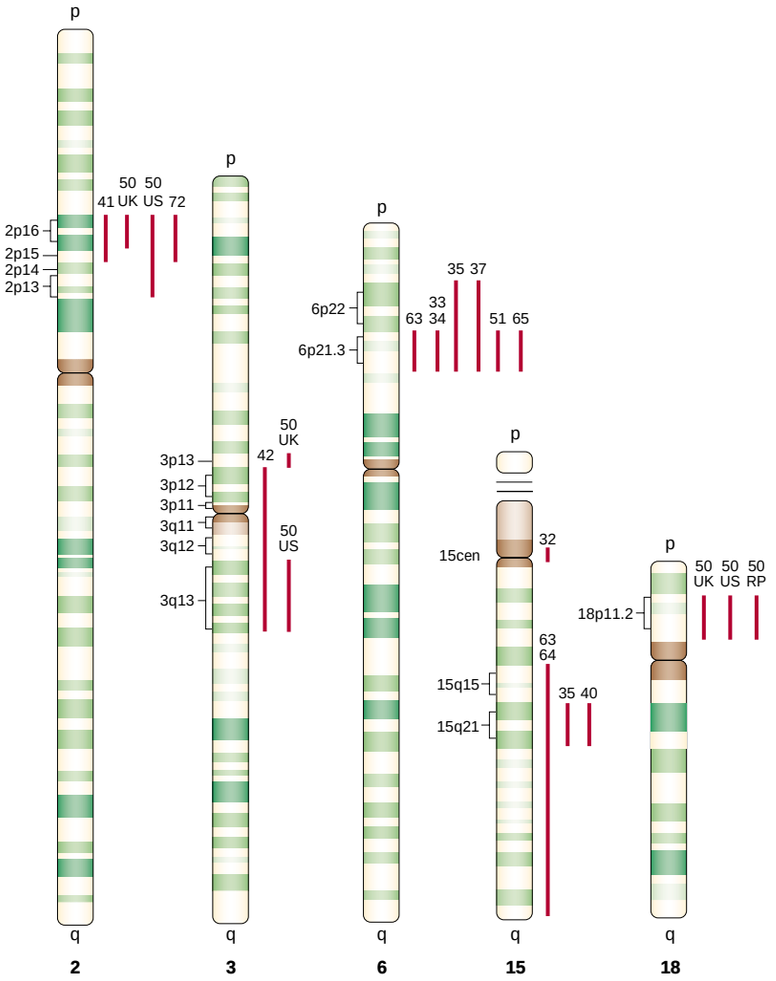


Figure 3 | **Replicated regions of chromosomes 2, 3, 6, 15 and 18 implicated by linkage studies of dyslexia.** Ideograms of each chromosome are shown with the cytogenetic bands of interest indicated. Each chromosome has a short (p) arm and a long (q) arm, which are separated by a centromere. Red bars indicate approximate positions of positive regions of linkage, with the relevant citation number of the study shown above. REF. 50 included two independent genome scans (using samples from the United Kingdom and the United States) and a further replication set (RP). Further details of each study are given in TABLES 2–4 and in the main text.

single-point parametric methods[35]. However, they noted that other markers mapping close to *D15S143* gave highly negative lod scores. (No multipoint parametric results were reported.) Linkage was not observed for other component phenotypes, or with complementary nonparametric analysis of the same data. But, in 1998, an independent investigation of spelling disability in German families supported linkage to 15q21, with parametric and nonparametric methods[40]. Overall, targeted linkage studies indicate the presence of a gene in 15q that influences dyslexia, but inconsistencies in the reported data must be addressed before the evidence can be considered compelling.

An investigation of the Smith *et al.*[63] kindreds that had implicated 15q also indicated linkage to protein polymorphisms in the vicinity of the human leukocyte antigen (HLA) COMPLEX on chromosome 6p. Cardon and colleagues[33,34] analysed sib-ships from these kindreds and a sample of DZ twins, using QTL methods and DNA-based markers, thereby obtaining evidence for the 6p21.3 locus in each data set. Linkage of reading-related phenotypes to 6p21.3 has since become one of the most replicable findings in the genetics of human cognition (TABLE 3). Evidence from several independent data sets shows remarkable agreement about the probable position of the risk locus[33–35,37,50–52,65] (FIG. 3). Samples implicating 6p have been obtained from diverse sources and studied with a variety of methods, including qualitative analyses of extended pedigrees[35,65] and quantitative approaches in sib-pairs or DZ twins[33,34,37,50–52]. These findings are intriguing in view of one often-contested theory of dyslexia aetiology, which posits a direct connection between immune dysfunction and reading problems[66]. However, data from the same samples that support genetic linkage of reading-related phenotypes to 6p21.3 fail to support a connection (mediated genetically or otherwise) between immune disorders and dyslexia[67]. It seems likely that the putative dyslexia risk factor on 6p is not a gene of the immune system, and that its location adjacent to a cluster of immune-related genes has no aetiological significance. Although evidence in support of an involvement of 6p21.3 is more robust than that for 15q21, there have been reports of non-replication in large samples[36,40,68]. Still, these negative studies did not formally exclude the involvement of 6p. For example, a QTL-based study of Canadian families did not find significant evidence of linkage, probably owing to genetic heterogeneity. But the authors did note that some of their results were weakly supportive, including a lod score of 0.82 in a region that was consistent with the positive studies[68].

Targeted studies[62,69,70] have also led to suggestions of other potential sites of dyslexia linkage on 1p34–36 and 6q12. At present, the data that support these sites are weaker than those for loci on 6p21.3 and 15q21, so further investigation is required[69,70].

A genome-wide perspective

As highlighted here, integrating data from multiple molecular investigations of dyslexia can be problematic, even when similar chromosomal regions seem to be implicated. The interpretation of results is complicated by discrepancies in the criteria for sample recruitment, phenotypic definition, marker selection and analytical method. Moreover, most studies have focused on small subsets of the human genome, often analysing the same marker data from these limited regions with several different analytical approaches. There are substantial benefits to be gained from obtaining a genome-wide perspective of linkage in any given sample of families[56]. In targeted studies, a locus with an important effect on the phenotype could remain undetected, simply because the relevant chromosomal region was never examined. Furthermore, with genome-wide data, it is possible to assess the general behaviour of linkage statistics in the sample under investigation, allowing a comparison of positive results to background levels of

HLA COMPLEX
A well-studied region of chromosome 6p that contains many loci, such as the human leukocyte antigen (HLA) genes, which encode key components of the immune system. Also known as the major histocompatibility complex (MHC).

Table 3 | Targeted linkage studies of chromosome 6p in developmental dyslexia

Authors (year)	Sample (country)	Treatment of phenotype	Linkage method	Summary of findings	References
Smith *et al.* (1991)	19 extended families (USA)	Qualitative, global	HE	$P < 0.02$ with marker GLO1 (a protein polymorphism in the red cell enzyme glyoxylase 1), mapping to 6p21.3	63
		Quantitative, global	HE	$P < 0.0001$ with marker BF (a protein polymorphism in properdin factor), which maps to 6p21.3	
Cardon *et al.* (1994, 1995)*	19 extended families (USA)	Quantitative, global	DF	Interval mapping identified peak linkage ($P = 0.04$) between DNA markers *D6S105* and *TNFB* in 6p21.3	33,34
	46 DZ twin pairs (USA)*	Quantitative, global	DF	Interval mapping identified peak linkage ($P = 0.009$) between *D6S105* and *TNFB**	
Grigorenko *et al.* (1997)	6 extended families (USA)	Qualitative, components	Parametric	No significant linkage for any phenotypes examined (phoneme awareness, phonological decoding, rapid automized naming, single-word reading or IQ–reading discrepancy)	35
			Nonparametric	Multipoint analyses gave $P < 0.005$ in *D6S109–D6S306* interval of 6p22.3–21.3 for each phenotype; strongest results for phoneme awareness ($P < 0.000001$), weakest for single-word reading ($P < 0.005$)	
Schulte-Körne *et al.* (1998)	7 extended families (Germany)	Qualitative, global	Parametric	No significant linkage; multipoint analyses gave lod of –0.95 between *D6S105* and *D6S464* in 6p21.3	40
			Nonparametric	No significant linkage; multipoint analyses gave $P = 0.21$ between *D6S105* and *D6S464*	
Field and Kaplan (1998)	79 families, nuclear and extended (Canada)	Qualitative, global	Parametric and nonparametric	Absence of linkage to several markers in 6p23–21.3	36
Fisher *et al.* (1999)	82 nuclear families (UK)	Quantitative, components	HE	Multipoint analyses identified peak linkage between *D6276* and *D6S105* in 6p21.3 for tests of phonological decoding ($P = 0.007$) and orthographic coding ($P = 0.0006$)	37
			VC	Multipoint analyses identified peak linkage between *D6276* and *D6S105* for phonological decoding ($P = 0.004$) and orthographic coding ($P = 0.007$)	
Gayán *et al.* (1999)	79 twin-based families (USA)	Quantitative, components	DF	Multipoint analyses identified peak linkage close to *D6276* for several reading- and language-related measures; strongest results were lods of 1.46 for phoneme awareness, 2.42 for phonological decoding and 3.1 for orthographic choice	51
Petryshen *et al.* (2000)	79 families (Canada)	Quantitative, components	HE	No significant linkage with measures of phoneme awareness, phonological coding, rapid automized naming or spelling; however, spelling gave P=0.07 at *TNFB*	68
			VC	Weak evidence of linkage to 6p23–21.3 for phonological coding, rapid automized naming and spelling (spelling gave a lod of 0.82 close to *TNFB*)	
Grigorenko *et al.* (2000)	8 extended families (USA)	Qualitative, components	Nonparametric	Support for linkage to *D6S464–D6S273* region in 6p21.3 for a variety of phenotypes, including single-word reading, vocabulary and spelling	65

Qualitative affection status or quantitative measures of deficit were used; some studies adopted a global assessment of the phenotype, whereas others dissected the phenotype into hypothetical components. The linkage method was either parametric model-based linkage analysis or nonparametric linkage analysis of allele sharing in affected relatives. DF, DeFries–Fulker regression analysis of selected sib-pairs; HE, Haseman–Elston regression analysis of sib-pairs; VC, variance-components analysis. *The report by Cardon *et al.* (1994), involving 50 twin pairs, included four pairs who were later discovered to be monozygotic. Re-analysis of the data excluding these four monozygotic twins led to a reduction in the significance of linkage and the results were published in a 1995 correction. The results given in this table are taken from the corrected report.

'noise' across all chromosomes. For example, methods that are prone to yield false-positive evidence for linkage could be seen to yield equally high lod scores at numerous chromosomal sites.

Genome-wide scans have been carried out in two large extended pedigrees in which inheritance is consistent with AUTOSOMAL DOMINANT transmission[41,42]. The first (from Norway) yielded significant linkage to 2p15–16, whereas the second (from Finland) strongly implicated 3p12–q13 (TABLE 4). Neither study reported linkage to sites that were indicated by previous investigations. In each family, although linkage evidence was convincing, phenotype–genotype correspondence was incomplete, with cases of phenocopy and/or non-penetrance[41,42]. Again, these molecular studies directly show the genetic complexity of dyslexia. Heterogeneity, reduced penetrance and phenocopies are evident, even when studying single multigenerational pedigrees with apparently simple inheritance. Furthermore, these kinds of family are scarce, raising the question of whether the relevant genetic effects will generalize to the wider population of people with dyslexia. To identify loci that are important for the latter, Fisher and colleagues[50] performed genome-wide scans in two large sets of nuclear families affected by dyslexia, from the United Kingdom and the United States, using QTL methodology and simulations to assess significance of the results. Previous targeted investigations of these family collections had provided evidence in support of the 6p21.3 locus[33,34,37,51], which could now be assessed in the context of the remainder of the genome. The study continued to support the importance of 6p21.3 in dyslexia, but indicated several other regions on various chromosomes that might similarly be involved in

AUTOSOMAL DOMINANT
One type of inheritance pattern that is observed for monogenic traits. Autosomes are any chromosomes in a cell that are not sex chromosomes. Autosomal dominant transmission results when an abnormal copy of an autosomal gene from a single parent gives rise to the trait, even though the copy inherited from the other parent is normal.

Table 4 | Genome-wide scans for loci influencing developmental dyslexia

Authors (year)	Sample (country)	Treatment of phenotype	Summary of findings	Reference
Fagerheim *et al.* (1999)	1 large extended family with apparent autosomal dominant inheritance (Norway)	Qualitative, global	• Genome-wide parametric analyses: significant linkage to 2p15–16; maximum pointwise lod scores of 2.92, 3.54 or 4.32, depending on inclusion criteria • Nonparametric analyses of 2p15–16: multipoint *P*-values of 0.016, 0.023 or 0.0009, depending on inclusion criteria • Co-segregation: of 18 genotyped family members with positive current diagnosis and/or history of dyslexia, 15 were IBD for 2p15–16 (i.e. 3 possible cases of phenocopy); one child who inherited the 2p risk genotype appeared to be unaffected (potential case of non-penetrance)	41
Nopola-Hemmi *et al.* (2001)	1 large extended family with apparent autosomal dominant inheritance (Finland)	Qualitative, global	• Genome-wide nonparametric analyses in subset of family: most significant result for 3p12–q13 region, with $P = 0.0017$ • Follow-up nonparametric analyses of 3p12–q13 in full extended family: $P = 0.00006$ • Parametric analyses of 3p12–q13 in full extended family: significant multipoint lod score of 3.84 • Co-segregation: of 21 dyslexic members, 19 were IBD for 3p12–q13 (i.e. 2 cases of phenocopy); simulations showed that this would occur by chance in < 1/1,000 genome scans	42
Fisher *et al.* (2002)	89 nuclear families (UK)	Quantitative, components	• Genome-wide QTL analyses of phoneme awareness, phonological decoding, orthographic processing, single-word reading: implicated several regions; most significant results on chromosomes 2, 3, 6, 9, 11, 18 and X; included 2p16 and 3p13 regions close to those found in previous scans; region on 18p11.2 gave empirical multipoint $P < 0.00001$ with single-word reading using VC approach	50
	119 twin-based families (USA)	Quantitative, components	• Genome-wide QTL analyses of phoneme awareness, phonological decoding, orthographic processing, single-word reading: implicated several regions; most significant results on 2, 3, 4, 8, 13, 18 and 21; included 2p15 and 3q13 regions implicated by previous scans; region on 18p11.2 gave empirical multipoint $P < 0.0004$ with single-word reading using DF approach	
	84 nuclear families (UK)	Quantitative, components	• QTL analysis of 18p11.2 for component measures in independent data set: replication of linkage; most significant empirical multipoint $P < 0.0005$ for VC analysis of phoneme awareness; combined analysis of all 173 UK families confirmed that 18p11.2 influences multiple measures	

Qualitative affection status or quantitative measures of deficit were used; some studies adopted a global assessment of the phenotype, whereas others dissected the phenotype into hypothetical components. The linkage method was either parametric model-based linkage analysis or nonparametric linkage analysis of allele sharing in affected relatives. DF, DeFries–Fulker regression analysis of selected sib-pairs; IBD, identical-by-descent; QTL, quantitative trait locus; VC, variance-components analysis.

trait susceptibility (TABLE 4). Notably, both the UK and the US samples implicated regions on 2p15–16 and 3p12–q13, indicating that effects at these loci might indeed be relevant to common forms of dyslexia, rather than being restricted to rare multigenerational pedigrees with simpler transmission[50,71]. This concordance is especially encouraging, given the disparity between the methods in the Norwegian/Finnish investigations and the UK/US genome scans. Furthermore, an independent study of Canadian families replicated the linkage at the 2p locus with both qualitative and quantitative methods, strengthening the case that this represents a susceptibility locus[72].

The key finding of the genome-wide searches by Fisher *et al.*[50] was the observation of strong linkage to 18p11.2 in the UK families. This same region was among the most significant results in the US sample, and also linked in a third independent set of families that were investigated. This study illustrates the value of QTL-based genome-wide scans in large samples, detecting at least four potential dyslexia susceptibility loci (on chromosomes 2, 3, 6 and 18) for which there is independent verification in multiple data sets. However, note that, as in all complex-trait analyses, until the relevant gene variants are pinpointed, there still remains a possibility that one or more of these loci might turn out to represent false-positive findings.

Alternative approaches to detecting dyslexia loci

Linkage studies are complemented by the investigation of people with chromosomal abnormalities, such as TRANSLOCATIONS, DUPLICATIONS, DELETIONS or INVERSIONS, that are associated with the trait of interest. For example, translocations might disrupt a gene at a breakage site or lead to the fusion of two normally unrelated genes; in some cases, this can be traced as the cause of the disorder. Many chromosomal alterations have no phenotypic consequences or alter the expression of genes that map up to one million bases away from a BREAKPOINT (known as a position effect), so caution is needed when interpreting such data. Nopola-Hemmi *et al.*[73] described two independent families in which people with dyslexia had inherited translocations that involved the same region of 15q21, consistent with that previously implicated by linkage studies. In each family, there were also children who were unaffected despite inheriting a rearrangement, indicating a significantly reduced penetrance. However, the convergence of independent linkage and translocation data provides further support for a 15q locus influencing dyslexia, and detailed analysis of the translocations might aid in the isolation of the putative risk gene. Another study reported co-segregation of a translocation involving 1p22 and 2q31 with retarded speech development and dyslexia in three members of a family[39]. However, the associated breakpoints map a

TRANSLOCATION
A genetic rearrangement in which part of a chromosome is detached by breakage and becomes attached to another part of the same chromosome, or to a different chromosome.

DUPLICATION
A genetic rearrangement that involves the doubling or repetition of part of a chromosome.

DELETION
A genetic rearrangement that involves the loss of part of a chromosome.

INVERSION
A genetic rearrangement in which part of a chromosome is reversed, so that the genes within that part are in inverse order.

BREAKPOINT
The specific site of chromosomal breakage that is associated with a particular chromosomal rearrangement.

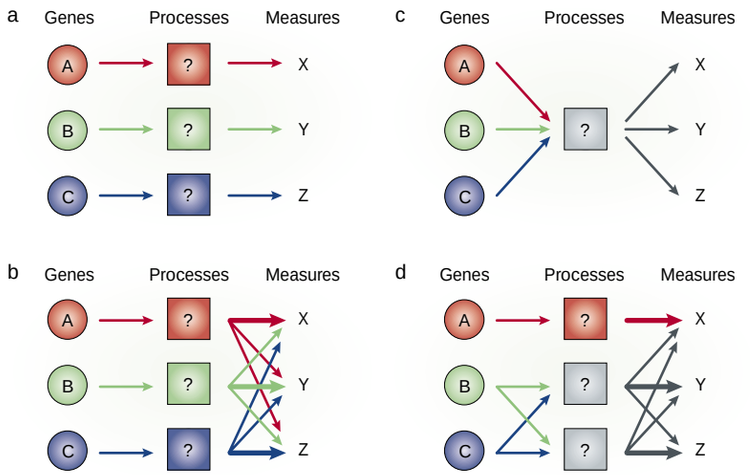


Figure 4 | **Genetic dissection of dyslexia.** Schematic simplified representations of pathways that map specific genes (A, B and C) to phenotypic measures of different aspects of reading ability (X, Y and Z) through unknown neurological mechanisms (indicated by '?'). **a** | The simplest models involve straightforward one-to-one relationships between specific genes and different measures, implying the presence of unique underlying brain processes (shown in red, green and blue). Both genetic and psychometric data indicate that this is unlikely to be valid. **b** | A more complex model acknowledges that, although the language- and reading-related measures that are used for the phenotypic assessment of dyslexia might tap predominantly one or other hypothetical component, they do not represent pure indicators of isolated brain processes. So, even if genes map to separable neurological mechanisms, simple relationships will not be revealed. **c** | There are moderate to high correlations between most language- and reading-related measures, and twin studies indicate that the correlated variance is highly heritable. So, it is possible that multiple genes influence common mechanisms that influence all measures. **d** | Twin studies also indicate the presence of heritable variability that is independent for each measure. It is therefore most likely that the underlying aetiology of dyslexia involves a complex interplay between specific and mixed effects. In reality, all these models are gross oversimplifications, as they ignore the role of the environment and the possibility of interactions between different genes or different neurological mechanisms. They are intended merely to illustrate some of the key issues for researchers seeking a genetic explanation of dyslexia.

substantial distance from regions of chromosomes 1 and 2 that are highlighted by other dyslexia studies; too far to be explained by position effects.

Another approach that is beginning to be applied to dyslexia is association analysis[74]. Whereas linkage assesses phenotype–genotype co-segregation within a family, association looks for correlations between specific alleles and a trait at the population level. The latter has greater power to detect genes of minor effect and allows risk variants to be mapped at higher resolution. Positive evidence of association could indicate that the allele itself is a risk variant, but can also be observed for alleles at loci that are in LINKAGE DISEQUILIBRIUM with the true susceptibility gene. So far, targeted association studies of dyslexia have been undertaken for small regions of 15q (REF. 75), 2q (REF. 71) and 6p (REF. 76); it is not yet feasible to carry out studies of this type on a genome-wide scale. This is because association approaches require a much higher density of marker coverage than linkage methods to ensure the reliable detection of a genetic effect. Furthermore, the appropriate selection of markers is influenced by region-specific patterns of linkage disequilibrium, which vary substantially throughout the genome[74].

LINKAGE DISEQUILIBRIUM
Non-random association between specific allelic variants at one genetic locus and those at another genetic locus that maps nearby.

EFFECT SIZE
A standardized measure of effect that is adopted when different scales are used to measure an outcome. In QTL analyses, the effect size is the proportion of variability in a measure that is attributable to the genetic locus of interest.

A molecular genetic dissection of dyslexia?

A recent idea that captured the imagination of many researchers was the proposal that different components of this complex phenotype could be linked to distinct genetic loci[35]. The corollary is that molecular investigations might reveal simple one-to-one mapping between certain cognitive processes and specific gene variants, which combine to give the overall dyslexic phenotype (FIG. 4). The hope that the genetic aetiology of dyslexia could be thus dissected initially arose from analyses of chromosomes 6 and 15 by Grigorenko and colleagues[35]. After observing variable patterns of linkage for different component phenotypes at each locus, they suggested the possibility of separable genetic effects, such that phoneme awareness mapped to chromosome 6, whereas single-word reading mapped to 15. This interpretation has been criticized[30,36,37,77]. First, Grigorenko *et al.* identified significant linkage to 6p with every phenotype under investigation[37,77]. Maximum lod scores at this locus varied for each component, with phoneme awareness highest and single-word reading lowest, but the differences in linkage evidence were not significant. Second, variability in the profiles of results at each locus might have resulted from the use of alternative statistical methodologies; linkage to chromosome 6 was observed only with nonparametric techniques, whereas linkage to chromosome 15 was found only by parametric analysis[37]. Third, the validity of viewing single-word reading as a component phenotype, cognitively separate from phoneme awareness, has been questioned[77].

Other studies have failed to support the idea that loci on 6 and 15 predominantly influence any single component[37,40,50–52,65,73,76]. For example, QTL-based sib-pair analyses consistently indicate that the 6p locus affects several aspects of dyslexia, including phonological and orthographic processing and rapid naming[37,50–52,76]. Most importantly, the authors of the Grigorenko *et al.*[35] report carried out a follow-up study in an expanded sample, including new families and further subjects from the original pedigrees[65]. In this later investigation, linkage to 6p became strongest for single-word reading, and much weaker for phoneme awareness (that is, an inverse pattern to the initial findings). As the authors point out, this highlights one of the key methodological challenges that researchers in the field face[65]. Comparing magnitudes of univariate linkage for each component is not a reliable way of evaluating relative EFFECT SIZES of a particular locus on different aspects of the phenotype[37,50,65]. Linkage levels might be influenced by factors that are unrelated to the underlying genetic effect, such as age distribution of the sample, sensitivity of psychometric testing, or even simple stochastic variability resulting from inadequate sample size[37,50,65]. It has also been shown that the process of maximizing lod scores in a particular data set leads to a bias in estimating effect size, which can be particularly problematic when evaluating data from genome-wide scans[78].

These concerns are further illustrated by the identification of a potential susceptibility locus on chromosome 18 in two separate QTL-based genome-wide scans[50]. In the UK scan sample of Fisher *et al.*[50], there

Box 3 | A gene mutated in a speech and language disorder

Although the boundaries between speech/language disorders and dyslexia are not always clear, the former involve gross language problems, detected by standard tests of grammar, syntax and/or vocabulary[31]. People with dyslexia often perform in the normal range on these tests, but show deficits in language-related processes such as the manipulation of phonemes. In 1998, Fisher *et al* [58] reported linkage studies of a unique three-generation family, known as KE, in which severe speech and language impairment was inherited as an autosomal dominant monogenic trait. Using traditional approaches, the locus was mapped to a small interval of 7q31, which co-segregated perfectly with the disorder; that is, there was 100% concordance between genotype and affection status, with no phenocopies and full penetrance.

The researchers constructed a sequence-based map of genes in 7q31, and used it to direct mutation analyses[80]. The search was aided by the identification of CS, an unrelated child with similar problems to those of family KE, which were associated with a translocation involving 7q31. Localization of the translocation breakpoint indicated that it disrupted the gene encoding FOXP2, a transcription factor containing a forkhead/winged-helix DNA-binding domain[81]. FOXP2 belongs to a family of proteins that are key regulators of gene expression during embryogenesis, and have been implicated in various developmental disorders in humans and mice[89–93]. Sequencing of *FOXP2* in the KE family revealed a point mutation in all affected individuals, altering an amino-acid residue in a crucial part of the DNA-binding domain of the protein[81].

The phenotype associated with FOXP2 disruption is severe, involving difficulty in controlling the fine mouth movements that are required for speech[94], coupled with deficits in many aspects of language processing and grammatical skill[95]. Some individuals in the KE family have reduced non-verbal intelligence[95], but these general cognitive deficits do not tend to co-segregate with the disorder. Variants in the coding region of FOXP2 do not seem to be a main cause of more common forms of speech and language impairment[82]. However, functional studies of FOXP2 might offer new insight into neurological mechanisms that are important for an individual to acquire speech and language[96,97].

was strong linkage of 18p11.2 to single-word reading, but substantially weaker evidence for other measures. Taken at face value and disregarding the issues discussed above, it might be reasonable to conclude that this locus is 'specific' to single-word reading. However, analysis of the US sample indicated more wide-ranging effects of 18p11.2 on multiple measures. A further independent sample of UK families was specifically investigated at 18p11.2, and linkage was replicated, but in this case, the most significant evidence arose from a measure of phoneme awareness[50]. Combined analyses of all UK families supported the view that the 18p locus is a general risk factor for dyslexia. Observed differences in linkage profiles probably arose from stochastic variability owing to small sample sizes, along with the effect-size biases associated with genome scanning[50,78].

It is clear from heritability and linkage studies that simple relationships between hypothetical components of the reading process and molecular genetic data are unlikely to exist (FIG. 4). However, there is still considerable interest in the possibility of the 'genetic dissection' of dyslexia. As discussed above, attempts to address this through univariate analyses have been ineffectual, and approaches are needed that can accurately estimate relative effect sizes of a particular locus for different reading- and language-related measures. Bivariate linkage analyses are already being carried out[52,79], but perhaps the most promise lies with fully multivariate linkage methods that can simultaneously incorporate data from all measures. However, for a comprehensive dissection of the genetic aetiology of dyslexia, we will first need to find the risk genes themselves, and to examine their effects in very large samples.

TRANSCRIPTION FACTOR
A DNA-binding protein that regulates gene expression.

Can we identify susceptibility genes for dyslexia?

Although the concordance of linkage results is encouraging, finding the particular genetic variants that influence dyslexia remains a daunting task. Linkage typically implicates chromosomal regions that contain hundreds of candidate genes, so success will probably depend on complementary data from chromosomal abnormalities and/or association-based analyses, as well as the use of much larger sample sizes — thousands, rather than hundreds, of nuclear families. The availability of comprehensive data from human genomic sequencing projects will have a great impact on these efforts[31]. So far, no specific dyslexia risk gene has been identified, but studies of speech and language deficits have been more fruitful, yielding a gene responsible for a rare and severe form of the disorder[58,80,81] (BOX 3). This serves as a pertinent example of how investigations of these types of trait might ultimately succeed and yield new insight into underlying processes. The main caveat is that genetic effects that are implicated in rare cases might not necessarily explain variability in the wider population[82].

Identification of specific gene variants that contribute to dyslexia will have many ramifications[30]. These include the possibility of early identification of those at elevated risk, allowing environmental intervention at a young age and the diagnosis of phenotypically ambiguous cases. Note that genotype–phenotype correspondence will usually be far from perfect, so any predictions will be probabilistic (concordance rates of reading difficulties in MZ twins are only ~2/3). A main goal after identifying crucial genes will be to increase our understanding of the molecular pathology of dyslexia, with the hope that this will also shed light on the mechanisms that are involved in normal reading- and language-related processes. Even if a gene is only implicated in a rare form of the trait, it will still provide a valuable entry point into the relevant developmental pathways[31]. The nature of functional studies will depend largely on the types of protein that the genes encode, be they involved in metabolism, structure, signalling, transcriptional regulation or some other cellular process[83]. For example, if a susceptibility gene is found to encode a cell-surface receptor, then a variety of techniques can be used to isolate the proteins with which it interacts and to dissect the relevant signalling pathways. Alternatively, for a susceptibility gene that encodes a TRANSCRIPTION FACTOR, it is possible to exploit new methods to identify the downstream targets in neuronal development. Isolation of key genetic pathways that are implicated in reading- and language-related disorders might help to bridge the gaps between other levels of study, such as brain imaging and neuropsychology, bringing us closer to a comprehensive explanation of the aetiology of dyslexia.

1. Hinshelwood, J. Word blindness and visual memories. *Lancet* **2**, 1566–1570 (1895).
2. Morgan, W. P. A case of congenital word blindness. *Br. Med. J.* **2**, 1378 (1896).
3. Orton, S. T. Word-blindness in school children. *Arch. Neurol. Psychiatr.* **14**, 582–615 (1925).
4. Snowling, M. J. From language to reading and dyslexia. *Dyslexia* **7**, 37–46 (2001).
5. Wolf, M. & Bowers, P. G. Naming-speed processes and developmental reading disabilities: an introduction to the special issue on the double-deficit hypothesis. *J. Learn. Disabil.* **33**, 322–324 (2000).
6. Eden, G. F. *et al.* Abnormal processing of visual motion in dyslexia revealed by functional brain imaging. *Nature* **382**, 66–69 (1996).
7. Temple, E. *et al.* Disruption of the neural response to rapid acoustic stimuli in dyslexia: evidence from functional MRI. *Proc. Natl Acad. Sci. USA* **97**, 13907–13912 (2000).
8. Nicolson, R. I., Fawcett, A. J. & Dean, P. Developmental dyslexia: the cerebellar deficit hypothesis. *Trends Neurosci.* **24**, 508–511 (2001).
9. Stein, J. & Walsh, V. To see but not to read; the magnocellular theory of dyslexia. *Trends Neurosci.* **20**, 147–152 (1997).
10. Hari, R. & Renvall, H. Impaired processing of rapid stimulus sequences in dyslexia. *Trends Cogn. Sci.* **5**, 525–532 (2001).
11. Habib, M. The neurological basis of developmental dyslexia. An overview and working hypothesis. *Brain* **123**, 2373–2399 (2000).
A comprehensive discussion of multidisciplinary investigations into the aetiology of dyslexia.
12. Thomas, C. J. Congenital 'word-blindness' and its treatment. *Ophthalmoscope* **3**, 380–385 (1905).
13. Stephenson, S. Six cases of congenital word-blindness affecting three generations of one family. *Ophthalmoscope* **5**, 482–484 (1907).
14. Hallgren, B. Specific dyslexia ('congenital word blindness'): a clinical and genetic study. *Acta Psychiatr. Neurol. Scand.* **65** (Suppl.), 1–287 (1950).
15. Finucci, J. M., Guthrie, J. T., Childs, A. L., Abbey, H. & Childs, B. The genetics of specific reading disability. *Ann. Hum. Genet.* **40**, 1–23 (1976).
16. Lewitter, F. I., DeFries, J. C. & Elston, R. C. Genetic models of reading disabilities. *Behav. Genet.* **10**, 9–30 (1980).
17. Vogler, G. P., DeFries, J. C. & Decker, S. N. Family history as an indicator of risk for reading disability. *J. Learn. Disabil.* **18**, 419–421 (1985).
18. Pennington, B. F. *et al.* Evidence for major gene transmission of developmental dyslexia. *JAMA* **266**, 1527–1534 (1991).
19. Wolff, P. H. & Melngailis, I. Family patterns of developmental dyslexia: clinical findings. *Am. J. Med. Genet.* **54**, 122–131 (1994).
20. Bakwin, H. Reading disability in twins. *Dev. Med. Child Neurol.* **15**, 184–187 (1973).
21. Stevenson, J., Graham, P., Fredman, G. & McLoughlin, V. A twin study of genetic influences on reading and spelling ability and disability. *J. Child Psychol. Psychiatry* **28**, 229–247 (1987).
22. DeFries, J. C., Fulker, D. W. & LaBuda, M. C. Evidence for a genetic aetiology in reading disability of twins. *Nature* **329**, 537–539 (1987).
23. DeFries, J. C. & Alarcón, M. Genetics of specific reading disability. *Ment. Retard. Dev. Disabil. Res. Rev.* **2**, 39–47 (1996).
24. Pennington, B. F. & Lefly, D. L. Early reading development in children at family risk for dyslexia. *Child Dev.* **72**, 816–833 (2001).
25. DeFries, J. C. & Fulker, D. W. Multiple regression analysis of twin data. *Behav. Genet.* **15**, 467–473 (1985).
26. DeFries, J. C. & Gillis, J. J. in *Nature, Nurture, and Psychology* (eds Plomin, R. & McClearn, G.) 121–145 (American Psychiatric Association, Washington DC, 1993).
27. Gayán, J. & Olson, R. K. Genetic and environmental influences on orthographic and phonological skills in children with reading disabilities. *Dev. Neuropsychol.* **20**, 483–507 (2001).
28. Bishop, D. V. M. *et al.* Different origin of auditory and phonological processing problems in children with language impairment: evidence from a twin study. *J. Speech Lang. Hear. Res.* **42**, 155–168 (1999).
29. Stevenson, J. Evidence for a genetic etiology in hyperactivity in children. *Behav. Genet.* **22**, 337–344 (1992).
30. Fisher, S. E. & Smith, S. D. in *Dyslexia: Theory and Good Practice* (ed. Fawcett, A. J.) 39–64 (Whurr, London, UK, 2001).
31. Fisher, S. E. in *Behavioral Genetics in the Postgenomic Era* (eds Plomin, R., DeFries, J. C., Craig, I. W. & McGuffin, P.) 205–226 (American Psychiatric Association, Washington DC, 2002).
32. Smith, S. D., Kimberling, W. J., Pennington, B. F. & Lubs, H. A. Specific reading disability: identification of an inherited form through linkage analysis. *Science* **219**, 1345 (1983).
33. Cardon, L. R. *et al.* Quantitative trait locus for reading disability on chromosome 6. *Science* **266**, 276–279 (1994).
An early demonstration of the value of applying QTL mapping methods to continuous measures of cognitive ability. Strong evidence was provided for a locus on 6p, which was subsequently verified in several independent populations.
34. Cardon, L. R. *et al.* Quantitative trait locus for reading disability: correction. *Science* **268**, 1553 (1995).
35. Grigorenko, E. L. *et al.* Susceptibility loci for distinct components of developmental dyslexia on chromosomes 6 and 15. *Am. J. Hum. Genet.* **60**, 27–39 (1997).
This paper proposed the intriguing idea that differing aspects of the dyslexia profile might link to distinct genetic loci. Although some conclusions from this study have been criticized, it raised the key question of whether we can reliably dissect complex cognitive phenotypes using genetic linkage data.
36. Field, L. L. & Kaplan, B. J. 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 (1998).
37. Fisher, S. E. *et al.* A quantitative trait locus on chromosome 6p influences different aspects of developmental dyslexia. *Am. J. Hum. Genet.* **64**, 146–156 (1999).
38. Bisgaard, M. L., Eiberg, H., Moller, N., Niebuhr, E. & Mohr, J. Dyslexia and chromosome 15 heteromorphism: negative lod score in a Danish sample. *Clin. Genet.* **32**, 118–119 (1987).
39. Froster, U., Schulte-Körne, G., Hebebrand, J. & Remschmidt, H. Cosegregation of balanced translocation (1;2) with retarded speech development and dyslexia. *Lancet* **342**, 178–179 (1993).
40. Schulte-Körne, G. *et al.* Evidence for linkage of spelling disability to chromosome 15. *Am. J. Hum. Genet.* **63**, 279–282 (1998).
41. Fagerheim, T. *et al.* A new gene (*DYX3*) for dyslexia is located on chromosome 2. *J. Med. Genet.* **36**, 664–669 (1999).
The successful genome-wide application of traditional parametric methods led to the localization of a dyslexia risk gene in a single large pedigree.
42. Nopola-Hemmi, J. *et al.* A dominant gene for developmental dyslexia on chromosome 3. *J. Med. Genet.* **38**, 658–664 (2001).
A genome-wide study of a large extended family identified a susceptibility locus for dyslexia on chromosome 3. References 41 and 42 provide a clear example of genetic heterogeneity, even when investigating multigenerational pedigrees.
43. Castles, A. & Coltheart, M. Varieties of developmental dyslexia. *Cognition* **47**, 149–180 (1993).
44. Castles, A., Datta, H., Gayán, J. & Olson, R. K. Varieties of developmental reading disorder: genetic and environmental influences. *J. Exp. Child Psychol.* **72**, 73–94 (1999).
45. Lander, E. S. & Schork, N. J. Genetic dissection of complex traits. *Science* **265**, 2037–2048 (1994).
An excellent introduction to key concepts of complex genetic analysis.
46. Haseman, J. K. & Elston, R. C. The investigation of linkage between a quantitative trait and a marker locus. *Behav. Genet.* **2**, 3–19 (1972).
47. Cardon, L. R. & Fulker, D. W. The power of interval mapping of quantitative trait loci, using selected sib pairs. *Am. J. Hum. Genet.* **55**, 825–833 (1994).
48. Amos, C. I. Robust variance-components approach for assessing genetic linkage in pedigrees. *Am. J. Hum. Genet.* **54**, 535–543 (1994).
49. Allison, D. B. *et al.* Testing the robustness of the likelihood-ratio test in a variance-component quantitative-trait loci-mapping procedure. *Am. J. Hum. Genet.* **65**, 531–544 (1999).
50. Fisher, S. E. *et al.* Independent genome-wide scans identify a chromosome 18 quantitative-trait locus influencing dyslexia. *Nature Genet.* **30**, 86–91 (2002).
This paper reported the first QTL-based genome-wide linkage scans for dyslexia, yielding robust evidence for a chromosome 18 locus influencing dyslexia in three independent samples of sib-pairs, and implicating a number of other potential loci of interest.
51. Gayán, J. *et al.* Quantitative trait locus for specific language and reading deficits on chromosome 6p. *Am. J. Hum. Genet.* **64**, 157–164 (1999).
52. Davis, C. J. *et al.* Etiology of reading difficulties and rapid naming: the Colorado Twin Study of Reading Disability. *Behav. Genet.* **31**, 625–635 (2001).
53. Wijsman, E. M. *et al.* Segregation analysis of phenotypic components of learning disabilities. I. Nonword memory and digit span. *Am. J. Hum. Genet.* **67**, 631–646 (2000).
54. Raskind, W. H., Hsu, L., Berninger, V. W., Thomson, J. B. & Wijsman, E. M. Familial aggregation of dyslexia phenotypes. *Behav. Genet.* **30**, 385–396 (2000).
55. Marlow, A. J. *et al.* Investigation of quantitative measures related to reading disability in a large sample of sib-pairs from the UK. *Behav. Genet.* **31**, 219–230 (2001).
56. Fisher, S. E., Stein, J. F. & Monaco, A. P. A genome-wide search strategy for identifying quantitative trait loci involved in reading and spelling disability (developmental dyslexia). *Eur. Child Adolesc. Psychiatry* **8** (Suppl. 3), 47–51 (1999).
57. Kruglyak, L. & Lander, E. S. Complete multipoint sib-pair analysis of qualitative and quantitative traits. *Am. J. Hum. Genet.* **57**, 439–454 (1995).
58. Fisher, S. E., Vargha-Khadem, F., Watkins, K. E., Monaco, A. P. & Pembrey, M. E. Localisation of a gene implicated in a severe speech and language disorder. *Nature Genet.* **18**, 168–170 (1998).
59. The SLI consortium. A genomewide scan identifies two novel loci involved in specific language impairment. *Am. J. Hum. Genet.* **70**, 384–398 (2002).
60. Lander, E. & Kruglyak, L. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nature Genet.* **11**, 241–247 (1995).
A technical overview of central concerns in the sound interpretation of linkage data when analysing complex traits.
61. Altmüller, J., Palmer, L. J., Fischer, G., Scherb, H. & Wjst, M. Genomewide scans of complex human diseases: true linkage is hard to find. *Am. J. Hum. Genet.* **69**, 936–950 (2001).
62. Rabin, M. *et al.* Suggestive linkage of developmental dyslexia to chromosome 1p34–p36. *Lancet* **342**, 178 (1993).
63. Smith, S. D., Kimberling, W. J. & Pennington, B. F. Screening for multiple genes influencing dyslexia. *Read. Writ.* **3**, 285–298 (1991).
64. Fulker, D. W. *et al.* Multiple regression of sib-pair data on reading to detect quantitative trait loci. *Read. Writ.* **3**, 299–313 (1991).
65. Grigorenko, E. L., Wood, F. B., Meyer, M. S. & Pauls, D. L. Chromosome 6p influences on different dyslexia-related cognitive processes: further confirmation. *Am. J. Hum. Genet.* **66**, 715–723 (2000).
66. Geschwind, N. & Behan, P. Left-handedness: association with immune disease, migraine, and developmental learning disorder. *Proc. Natl Acad. Sci. USA* **79**, 5097–5100 (1982).
67. Gilger, J. W. *et al.* A twin and family study of the association between immune system dysfunction and dyslexia using blood serum immunoassay and survey data. *Brain Cogn.* **36**, 310–333 (1998).
68. Petryshen, T. L., Kaplan, B. J., Liu, M. F. & Field, L. L. 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 (2000).
69. Grigorenko E. L. *et al.* Linkage studies suggest a possible locus for developmental dyslexia on chromosome 1p. *Am. J. Med. Genet.* **105**, 120–129 (2001).
70. Petryshen T. L. *et al.* Evidence for a susceptibility locus on chromosome 6q influencing phonological coding dyslexia. *Am. J. Med. Genet.* **105**, 507–517 (2001).
71. Francks, C. *et al.* Quantitative association analysis within the chromosome 2p12–16 dyslexia susceptibility region: microsatellite markers and candidate genes *SEMA4F* and *OTX1*. *Psychiatr. Genet.* **12**, 35–41 (2002).
72. Petryshen, T. L., Kaplan, B. J., Hughes, M. L., Tzenova, J. & Field, L. L. Supportive evidence for the *DYX3* dyslexia susceptibility gene in Canadian families. *J. Med. Genet.* **39**, 125–126 (2002).
73. Nopola-Hemmi, J. *et al.* Two translocations of chromosome 15q associated with dyslexia. *J. Med. Genet.* **37**, 771–775 (2000).
74. Cardon, L. R. & Bell, J. I. Association study designs for complex diseases. *Nature Rev. Genet.* **2**, 91–99 (2001).
75. Morris, D. W. *et al.* Family-based association mapping provides evidence for a gene for reading disability on chromosome 15q. *Hum. Mol. Genet.* **9**, 843–848 (2000).
Converging evidence for a locus on 15q has been revealed by complementary approaches to linkage mapping, including this report of association and a study of chromosomal abnormalities (reference 73).
76. Kaplan, D. E. *et al.* Evidence for linkage and association with reading disability on 6p21.3–22. *Am. J. Hum. Genet.* **70**, 1287–1298 (2002).
77. Pennington, B. F. Using genetics to dissect cognition. *Am. J. Hum. Genet.* **60**, 13–16 (1997).
An insightful critique of the suggestion that genes might show simple mapping to individual cognitive processes that underlie dyslexia.

78. Goring, H. H., Terwilliger, J. D. & Blangero, J. Large upward bias in estimation of locus-specific effects from genomewide scans. *Am. J. Hum. Genet.* **69**, 1357–1369 (2001).
79. Willcutt, E. G. *et al.* Quantitative trait locus for reading disability on chromosome 6p is pleiotropic for attention-deficit/hyperactivity disorder. *Am. J. Med. Genet.* **114**, 260–268 (2002).
80. Lai, C. S. L. *et al.* The *SPCH1* region on human 7q31: genomic characterization of the critical interval and localization of translocations associated with speech and language disorder. *Am. J. Hum. Genet.* **67**, 357–368 (2000).
81. Lai, C. S. L., Fisher, S. E., Hurst, J. A., Vargha-Khadem, F. & Monaco, A. P. A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature* **413**, 519–523 (2001).
This paper reports the identification of the *FOXP2* gene and shows that its disruption causes one form of speech and language impairment. This is the only known case of a direct link between a specific gene and this type of developmental disorder.
82. Newbury, D. F. *et al. FOXP2* is not a major susceptibility gene for autism or Specific Language Impairment (SLI). *Am. J. Hum. Genet.* **70**, 1318–1327 (2002).
83. Nokelainen, P. & Flint, J. Genetic effects on human cognition: lessons from the study of mental retardation syndromes. *J. Neurol. Neurosurg. Psychiatry* **72**, 287–296 (2002).
84. Thomson, M. E. The assessment of children with specific reading disabilities (dyslexia) using the British Ability Scales. *Br. J. Psychiatry* **73**, 461–478 (1982).
85. Siegel, L. S. & Himel, N. Socioeconomic status, age and the classification of dyslexics and poor readers: the dangers of using IQ scores in the definition of reading disability. *Dyslexia* **4**, 90–103 (1998).
86. Pennington, B. F., Gilger, J. W., Olson, R. K. & DeFries, J. C. The external validity of age- versus IQ-discrepancy definitions of reading disability: lessons from a twin study. *J. Learn. Disabil.* **25**, 562–573 (1992).
87. Shapiro, B. K. Specific reading disability: a multiplanar view. *Ment. Retard. Dev. Disabil. Res. Rev.* **7**, 13–20 (2001).
88. Shaywitz, S. E. *et al.* Persistence of dyslexia: the Connecticut Longitudinal Study at adolescence. *Pediatrics* **104**, 1351–1359 (1999).
89. Nehls, M., Pfeifer, D., Schorpp, M., Hedrich, H. & Boehm, T. New member of the winged-helix protein family disrupted in mouse and rat nude mutations. *Nature* **372**, 103–107 (1994).
90. Nishimura, D. Y. *et al.* The forkhead transcription factor gene *FKHL7* is responsible for glaucoma phenotypes which map to 6p25. *Nature Genet.* **19**, 140–147 (1998).
91. Fang, J. *et al.* Mutations in *FOXC2* (*MFH-1*), a forkhead family transcription factor, are responsible for the hereditary lymphedema–distichiasis syndrome. *Am. J. Hum. Genet.* **67**, 1382–1388 (2000).
92. Brunkow, M. E. *et al.* Disruption of a new forkhead/winged-helix protein, scurfin, results in the fatal lymphoproliferative disorder of the scurfy mouse. *Nature Genet.* **27**, 68–73 (2001).
93. Crisponi, L. *et al.* The putative forkhead transcription factor *FOXL2* is mutated in blepharophimosis/ptosis/epicanthus inversus syndrome. *Nature Genet.* **27**, 159–166 (2001).
94. Hurst, J. A., Baraitser, M., Auger, E., Graham, F. & Norell, S. An extended family with a dominantly inherited speech disorder. *Dev. Med. Child Neurol.* **32**, 347–355 (1990).
95. Vargha-Khadem, F., Watkins, K., Alcock, K., Fletcher, P. & Passingham, R. Praxic and nonverbal cognitive deficits in a large family with a genetically transmitted speech and language disorder. *Proc. Natl Acad. Sci. USA* **92**, 930–933 (1995).
96. Bishop, D. V. M. Putting language genes in perspective. *Trends Genet.* **18**, 57–59 (2002).
An interesting discussion of what the discovery of *FOXP2* might tell us about speech and language development.
97. Fisher S. E. in *Neurosciences at the Postgenomic Era* (eds Mallet, J. & Christen, Y.) (Springer–Verlag, Heidelberg, Germany, 2002).

Acknowledgements
S.E.F. is a Royal Society Research Fellow. J.C.D. is supported by a centre grant from the National Institute of Child Health and Human Development.

Online links

DATABASES
The following terms in this article are linked online to:
LocusLink: http://www.ncbi.nlm.nih.gov/LocusLink/
FOXP2
OMIM: http://www.ncbi.nlm.nih.gov/Omim/
ADHD | dyslexia
OMIM Gene Map:
http://www.ncbi.nlm.nih.gov/Omim/searchmap.html
1p22 | 1p34–36 | 2p15–16 | 2q31 | 3p12–q13 | 6p21.3 | 6q12 | 15q21 | 18p11.2

FURTHER INFORMATION
Encyclopedia of Life Sciences: http://www.els.net/
language | quantitative genetics
FOXP2 in Speech and Language Disorder:
http://www.well.ox.ac.uk/monaco/spch1cecilia.shtml
Genetics of Developmental Dyslexia:
http://www.well.ox.ac.uk/monaco/dyslexiasimon.shtml
Genetics of Specific Language Impairment:
http://www.well.ox.ac.uk/monaco/dianne/index.shtml
MIT Encyclopedia of Cognitive Sciences:
http://cognet.mit.edu/MITECS/
dyslexia | language impairment, developmental | reading | visual word recognition
Access to this interactive links box is free online.

Online links

LocusLink
http://www.ncbi.nlm.nih.gov/LocusLink/

FOXP2
http://www.ncbi.nlm.nih.gov/LocusLink/LocRpt.cgi?l=93986

OMIM
http://www.ncbi.nlm.nih.gov/Omim/

ADHD
http://www.ncbi.nlm.nih.gov/htbin-post/Omim/dispmim?143465

dyslexia
http://www.ncbi.nlm.nih.gov/htbin-post/Omim/getmim?search=dyslexia&field=title

OMIM Gene Map
http://www.ncbi.nlm.nih.gov/Omim/searchmap.html

1p22
http://www.ncbi.nlm.nih.gov/htbin-post/Omim/getmap?chromosome=1p22

1p34–36
http://www.ncbi.nlm.nih.gov/htbin-post/Omim/getmap?chromosome=1p34-36

2p15–16
http://www.ncbi.nlm.nih.gov/htbin-post/Omim/getmap?chromosome=2p15-16

2q31
http://www.ncbi.nlm.nih.gov/htbin-post/Omim/getmap?chromosome=2q31

3p12–q13
http://www.ncbi.nlm.nih.gov/htbin-post/Omim/getmap?chromosome=3p12-q13

6p21.3
http://www.ncbi.nlm.nih.gov/htbin-post/Omim/getmap?chromosome=6p21.3

6q12
http://www.ncbi.nlm.nih.gov/htbin-post/Omim/getmap?chromosome=6q12

15q21
http://www.ncbi.nlm.nih.gov/htbin-post/Omim/getmap?chromosome=15q21

18p11.2
http://www.ncbi.nlm.nih.gov/htbin-post/Omim/getmap?chromosome=18p11.2

Encyclopedia of Life Sciences
http://www.els.net/

language
http://www.els.net/els/els/els/return.html?artid=A0000151&freeid=e5a693d6cd26d43afd5061a0ce27a619

quantitative genetics
http://www.els.net/els/els/els/return.html?artid=A0001785&freeid=97e4bb55052414b15431199b8bc79d0a

FOXP2 in Speech and Language Disorder
http://www.well.ox.ac.uk/monaco/spch1cecilia.shtml

Genetics of Developmental Dyslexia
http://www.well.ox.ac.uk/monaco/dyslexiasimon.shtml

Genetics of Specific Language Impairment
http://www.well.ox.ac.uk/monaco/dianne/index.shtml

MIT Encyclopedia of Cognitive Sciences
http://cognet.mit.edu/MITECS/

dyslexia
http://cognet.mit.edu/MITECS/Entry/galaburda1

language impairment, developmental
http://cognet.mit.edu/MITECS/Entry/tallal

reading
http://cognet.mit.edu/MITECS/Entry/adams

visual word recognition
http://cognet.mit.edu/MITECS/Entry/seidenberg

Biographies

Simon E. Fisher is a postdoctoral fellow in neurogenetics at the Wellcome Trust Centre for Human Genetics (WTCHG), University of Oxford, UK. During his D.Phil. (in the Genetics Laboratory at Oxford), he specialized in the positional cloning of human disease genes and identified the gene that is mutated in a hereditary kidney-stone disorder. For the past six years, in Anthony Monaco's group at the WTCHG, he has led a research team searching for genetic variants that influence childhood learning disabilities. In that time, the team has completed the first large-scale genome-wide scans for dyslexia, language disorders and attention-deficit/hyperactivity disorder (ADHD), and has isolated a gene that is mutated in severe speech and language impairment. Simon Fisher was recently awarded a Royal Society Research Fellowship to head his own group with the aim of dissecting the neurological pathways involved in speech and language disorders.

John C. DeFries is Professor of Psychology, Director of the Colorado Learning Disabilities Research Center, and past Director (1981–2001) of the Institute for Behavioral Genetics (IBG), University of Colorado, Boulder. He received a Ph.D. in agriculture from the University of Illinois in 1961, moved to IBG in 1967, and founded the journal *Behavior Genetics* with Steven G. Vandenberg in 1970. He has co-authored eight books, including four editions of the standard textbook in the field of behavioural genetics, over 250 journal articles and two edited books that are in press. He served as President of the Behavior Genetics Association from 1982 to 1983, received the association's Th. Dobzhansky Award for Outstanding Research in 1992, and became a Fellow of the American Association for the Advancement of Science (Section J, Psychology) in 1994.

At a glance

- Despite decades of multidisciplinary investigation, the biological basis of dyslexia — a specific impairment of reading ability — remains obscure. But a series of recent studies has emphasized the contribution of genetic factors to this disorder.
- Dyslexia runs in families, and studies of monozygotic and dizygotic twins have provided valuable insights into the heritability of the condition. Methods developed for these studies have also aided in the genetic mapping of this reading disability.
- For several reasons, the genetic analysis of dyslexia is complex. For example, there is no straightforward correspondence between genotype and phenotype, and phenotypic variations can depend on the developmental stage of the subject. Similarly, there is a lack of consensus on the definition of dyslexia, and on whether it is a single trait or a cluster of traits with distinct aetiologies.
- Successful localization of genes that influence dyslexia has been aided by innovations in three areas. First, methods have been developed for mapping genes that contribute to quantitative variability in reading performance. Second, researchers are dissecting the phenotypic profile into distinct but related components for genetic study. Third, it is now possible to scan all chromosomes of the genome when searching for genes that influence complex traits such as dyslexia.
- Targeted linkage studies of dyslexia have provided strong evidence that two chromosomal regions — 15q21 and 6p21 — are involved in this syndrome. Similarly, genome-wide scans have identified further regions on chromosomes 2, 3 and 18 that seem to be linked to dyslexia in multiple independent sets of families.
- Although the linkage results highlight chromosomal regions that are involved in dyslexia susceptibility, finding individual genes that are affected remains a daunting task. So far, no specific dyslexia gene has been identified, but studies of speech and language deficits have found a gene — *FOXP2* — that is responsible for a rare form of the disorder.