REVIEW: GENETICS

Finding Genes That Underlie Complex Traits

Anne M. Glazier,[1] Joseph H. Nadeau,[2]* Timothy J. Aitman[1]*

Phenotypic variation among organisms is central to evolutionary adaptations underlying natural and artificial selection, and also determines individual susceptibility to common diseases. These types of complex traits pose special challenges for genetic analysis because of gene-gene and gene-environment interactions, genetic heterogeneity, low penetrance, and limited statistical power. Emerging genome resources and technologies are enabling systematic identification of genes underlying these complex traits. We propose standards for proof of gene discovery in complex traits and evaluate the nature of the genes identified to date. These proof-of-concept studies demonstrate the insights that can be expected from the accelerating pace of gene discovery in this field.

A
ll organisms vary in subtle and profound ways that involve every aspect of biological systems, including morphology, behavior, physiology, development, and susceptibility to common diseases. Many of these phenotypes are controlled by multiple genes and are therefore called multigenic or genetically complex traits, in contrast to phenotypes that are controlled by single genes (monogenic or Mendelian traits). The propensity of genetic background to modify the phenotypic expression of most if not all Mendelian traits suggests that few if any traits are truly monogenic and that instead most are genetically complex (1).

Many genes that control Mendelian traits, but relatively few genes underlying genetically complex traits, have been identified in the last 20 years (Fig. 1). Genes that contribute to complex traits (also known as quantitative trait loci or QTLs) pose special challenges that make gene discovery more difficult, including locus heterogeneity, epistasis, low penetrance, variable expressivity and pleiotropy, and limited statistical power (2–4). Prominent examples of these difficulties involve important diseases such as schizophrenia in humans, where claims of linkage discovery have been notoriously difficult to verify. The prospects for success have improved markedly, however, with the recent development of an extensive array of genome resources and technologies. Claims of gene discovery in complex traits require additional evidence, however. We propose standards of evidence that together establish the

[1]Physiological Genomics and Medicine Group, MRC Clinical Sciences Centre, Hammersmith Hospital, Imperial College Faculty of Medicine, Ducane Road, London W12 0NN, UK. [2]Department of Genetics and Center for Computational Genomics, Case Western Reserve University, Cleveland, OH 44106, USA.

*To whom correspondence should be addressed. E-mail: jhn4@po.cwru.edu, t.aitman@csc.mrc.ac.uk

formal burden of proof, and we then use these standards to evaluate the evidence for gene discovery in complex traits in a wide variety of organisms.

Burden of Proof

In Mendelian traits and diseases, the first step in gene discovery involves mapping the gene precisely and unambiguously to a small genetic interval. Typically, because of the strong relation between genotype and phenotype, single recombinants are sufficient to define minimal intervals of less than 1 cM. As a result, discovery of coding sequence variants that are found only in one of a small number of candidate genes in affected individuals usually provides adequate evidence to establish gene identity. The same certainties do not apply to genetically complex traits. We propose the following working criteria for establishment of gene discovery in studies of complex traits.

Step 1: Linkage and association. The first step is to establish statistically significant genome-wide evidence for linkage or association in a single study, or consistent suggestive evidence in several independent studies (5–7). The Lander-Kruglyak guidelines for significance thresholds address concerns about testing numerous genetic markers for linkage (multiple hypotheses) and about correlated inheritance patterns among linked markers (autocorrelation). They propose guidelines for identifying results that are statistically significant as well as those that warrant further investigation despite not reaching formal statistical significance. Permutation tests are an alternative method for establishing rigorous thresholds for statistical significance (8). However, because of the nature of complex traits, it is usual for minimal interval of a QTL—even in large human family collections or experimental crosses—to be restricted to no less than 10 to 30 cM in primary genome screens for genetic

linkage. A genetic interval of this size typically corresponds in humans to 10 to 30 Mb of DNA, or ~100 to 300 genes, which is far too many candidates to begin functional evaluation of each gene individually. To date, no complete genome-wide tests of association have been completed, although association studies offer considerable promise for studying complex traits in populations. In the absence of such proof-of-concept studies for genome-wide association, we focus in this review on those complex trait genes identified in whole-genome linkage studies.

Step 2: Fine-mapping. The next step is to reduce as much as possible the size of the critical interval. This can be done with the use of high-resolution crosses, congenic strains, near-isogenic lines, and progeny testing, or by linkage disequilibrium (LD) mapping in experimental crosses, family-based studies, or case-control studies. Initial low-resolution linkage studies typically establish the map location to a resolution that is sufficiently precise to justify further study. By contrast, the goal of a high-resolution study is to reduce the size of the candidate interval sufficiently that the number of candidate genes is modest and functional studies can be undertaken. These approaches may be used to reduce the minimal interval to less than 1 cM (9–12). For conclusive proof in LD studies, dense genetic markers covering the entire minimal interval should then be tested for disequilibrium with the trait phenotype in several populations. The density of markers required depends on the extent of local LD, but the recent evidence for haplotype blocks in humans, mice, and probably other species may simplify these studies (13). Considerable theoretical and empirical work is under way to determine what single-nucleotide polymorphism (SNP) density is optimal for genome-wide and regional association studies.

Step 3: Sequence analysis. DNA sequence analysis within the interval is needed to identify candidate nucleotide variants. Despite considerable effort, minimal QTL intervals often include several genes and numerous DNA sequence variants; some of these reside in coding regions, and others are located in flanking genomic DNA. Some QTLs result from single nucleotide lesions; others result from several variant nucleotides, either in the same gene or in closely linked and perhaps functionally unrelated genes (9, 14–17). As a result, each candidate nucleotide variant as well as all combinations of candidate nucleotides in one or several genes must be iden-

tified, prioritized, and functionally tested. This process is very different from that used to identify traditional Mendelian traits and presents a greater logistical challenge.

Step 4: Functional tests of candidate genes. The most conclusive evidence is a demonstration that replacement of the variant nucleotide results in swapping one phenotypic variant for another. This test can be based on knock-in technology (*18*) or a combination of gene targeting to create an engineered deficiency followed by transgenic complementation with the nucleotide or combination of nucleotides that is being tested. For cellular phenotypes, in vitro functional tests may be appropriate. There are at least two limitations: Transgenic and gene-targeting technologies are not available for many species, and some variants may be specific to particular species or heavily dependent on genetic background, in which case functional tests might not be informative.

Circumstantial evidence. In parallel with the distinction, based on statistical criteria, between suggestive and significant genetic linkage (*5*), we propose additional classes of evidence that together make a compelling case. This evidence could include appropriate tissue expression pattern and cellular distribution, similar phenotypes associated with naturally occurring or engineered mutations

says of gene expression combined with other formal evidence (steps 1 to 3 above)—may provide a sufficient wealth of evidence for the establishment of gene discovery.

Complicating factors. It has been speculated that complex traits result more often from noncoding regulatory variants than from coding sequence variants (*19–21*). If this is the case, searches restricted to coding sequences may fail to reveal the causal nucleotide variants, even if the correct gene has been screened. Noncoding regulatory variants pose special problems. In coding regions, the functional consequences of variants are readily assessed as missense, nonsense, splicing, and other polymorphisms. By contrast, interpreting the consequences of noncoding sequence variants is more complicated, if only because the relationship between promoter or intergenic sequence variation, gene expression level, and trait phenotype is less well understood than the relationship between coding DNA sequence and protein function. These factors explain why geneticists may be reluctant to embark on screens of regulatory and intergenic regions even though functional proof of regulatory sequence variants may be achieved by techniques such as yeast or bacterial artificial chromosome transgenesis (*22*). These factors are also a potential source of bias toward

Identified Genes in Complex Traits

With the use of the above criteria, increasing numbers of genes and allelic variants underlying complex traits have been identified from genome-wide linkage studies (Fig. 1). Most of the identified genes and variants come from studies of model organisms and plants (Table 1).

Plants. Several attributes of plant genetics make it possible to obtain strong evidence of gene identity. First, crosses often involve very large numbers of meioses (up to 10,000) that enable precise QTL localization. Second, because the ratio of physical to genetic distance is generally smaller in plants than in mammals—for example, 250,000 base pairs (250 kbp) per cM in *Arabidopsis* versus 1970 kbp/cM in mice (table S2)—each crossover provides greater mapping resolution. Third, genetic transformation techniques available in several plant species make it feasible to test whether candidate nucleotides are responsible for the phenotypic variants.

Proof of gene discovery has been obtained for six plant complex traits: two in rice, two in tomato, one in *Arabidopsis*, and one in maize, with formal complementation in four of these cases (Table 1). For example, large experimental crosses were used to generate maps for the rice photoperiod QTLs *Hd1* and *Hd6*, fine-mapping was carried out in both cases to 26-kbp critical regions in nearly isogenic lines, sequence variants were identified within genes in each critical region (*Se1* and *CK2α*), and complementation was achieved by transformation with wild-type genomic clones (*11, 12*). The fruit size QTL *fw2.2* was mapped in tomato by analysis of recombinants, cosmid complementation, and genomic sequencing to identify the *OFRX* gene as the underlying gene (*10*), although *fw2.2* allelic differences may result from changes in the coding or upstream noncoding regulatory regions.

Saccharomyces cerevisiae. The extremely low ratio of physical to genetic distance in yeast (3 kbp/cM; table S2) allows very high resolution mapping with relatively few meioses. A genome scan, based in part on reciprocal hemizygosity mapping, revealed a QTL for high-temperature growth (Htg) (*17*). Oligonucleotide arrays provided genetic markers showing linkage to an Htg QTL, the QTL was fine-mapped, and detailed sequence analysis revealed several nucleotides that differ between Htg$^+$ and Htg$^-$ strains. Isogenic strains that each differed only in the alleles of one of three genes (*MKT1*, *END3*, and *RHO2*) from the QTL region differed in growth characteristics depending on the alleles of these three tightly linked genes.

Drosophila. Bristle number and alcohol dehydrogenase (ADH) activity in *Drosophila* have been paradigms for QTL analysis for several decades (*19*). QTL mapping localized
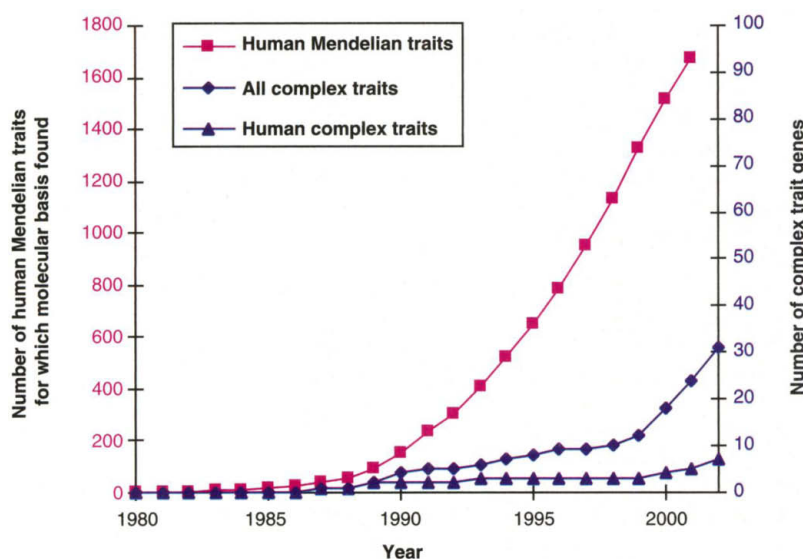

Fig. 1. Identification of genes underlying human Mendelian traits and genetically complex traits in humans and other species. Cumulative data for human Mendelian trait genes (to 2001) include all major genes causing a Mendelian disorder in which causal variants have been identified (*58, 59*). This reflects mutations in a total of 1336 genes. Complex trait genes were identified by the whole-genome screen approach and denote cumulative year-on-year data described in this review.

in other species, or strong mechanistic support for the causal relationship between variant nucleotide, altered protein expression or function, and phenotype. In species where in vivo functional tests are not possible, other lines of evidence—for example, in vitro complementation tests or reporter gene as-

identifying functionally significant coding rather than noncoding regulatory sequence variants. Progress is being made, however, with the recent report of regulatory genetic variants that control the level or pattern of expression of many genes in inbred strains of mice (*23*).

genes controlling abdominal and sternopleural bristle to the *Achaete-scute* (*ASC*), *scabrous* (*sca*), and *Delta* (*Dl*) genes (*19*). Evidence that each of these genes contributes to variation in bristle number is based on several observations: (i) These genes have important roles in the development of external sensory organs including bristles (*19*), (ii) spontaneous mutations affecting bristle number fail to complement mutant alleles in these genes, and (iii) restriction fragments at these gene loci are in linkage disequilibrium with bristle number in wild stocks (*19*). However, complementation of variant nucleotides and alleles remains to be undertaken.

Several nucleotide variants affect catalytic efficiency and protein level of *Adh*, the major gene controlling alcohol dehydrogenase activity (*14*). Sequence analysis and transformation experiments showed that catalytic efficiency is determined by a single amino acid variant that accounts for the difference between "slow" and "fast" activity variants (*24*). Genetic control of protein levels was mapped by transformation experiments in *Adh*-negative strains to a 2.3-kbp restriction fragment, which was then dissected into three separate, interacting fragments, each with multiple nucleotide variants that influence protein levels in transformed flies (*14*).

Cattle. On the basis of a cross involving 1158 progeny, a QTL for milk composition was mapped to the centromeric end of chromosome 14 in cows (*25*). Linkage disequilibrium and physical mapping showed that *DGAT1* is contained within the critical region (*26*). *DGAT1* catalyzes the final step in triglyceride synthesis, and complete inhibition of lactation is observed in *Dgat1*-deficient mice. Four polymorphisms are found in the *DGAT1* gene, three of which cosegregate with the phenotype. One of these polymorphisms is in an intron and another in the 3' untranslated region (3'UTR), leading the authors to propose that the third polymorphism, a nonconservative $Lys^{232} \rightarrow Ala$ substitution, underlies variation in milk yield and composition. However, a direct effect of the substitution on enzyme activity has not yet been demonstrated.

Rodents. Despite the numerous complex traits in mice and rats that have been analyzed by genome-wide linkage studies, few of the underlying genes have been identified. Of the eight mouse complex trait genes and three rat complex trait genes identified to date, four mouse and two rat QTLs have been formally proven by complementation.

Among 18 type 1 (autoimmune) diabetes susceptibility loci in the non-obese (NOD) mouse, the strongest gene effect involves a QTL within the major histocompatibility complex (MHC). Transgenesis was used to show that both the unique class II *I-A* gene and the null *I-E* gene are causally associated with disease susceptibility (*27–29*).

Mice with a mutation in the adenomatous

Table 1. Molecular basis of complex trait genes localized initially in genome-wide linkage studies for various species. The genes listed were identified according to the criteria of proof described in the text but may not include every complex trait gene identified. See table S1 for further details, including full references.

Trait	Gene	Molecular basis	Formal complementation?
		Human	
Type 1 diabetes	HLA-DQB	Aspartic acid at position 57 is protective	No
Type 1 diabetes	HLA-DQA	Multiple amino acid substitutions	No
Type 2 diabetes	CAPN10	Intronic haplotype	No
Crohn's disease	CARD15*	Multiple amino acid substitutions, single nucleotide insertion at nucleotide position 3020 causing frameshift and protein truncation	No
Alzheimer's disease	ApoE	Arginine substituted for cysteine at position 112	No
Asthma	ADAM33	Coding and noncoding haplotype pairs	No
Angiotensin I–converting enzyme levels	ACE	Multiple intragenic SNPs	No
		Cattle	
Milk yield, milk composition	DGAT1	Alanine substituted for lysine at position 232	No
		Mouse	
Intestinal neoplasia, Mom1	Pla2g2a	Insertion of thymidine at nucleotide position 208	Yes
Type 1 diabetes	I-Eα	Promoter deletion	Yes
	I-Aβ	Histidine substituted for proline at position 56 Serine substituted for aspartic acid at position 57	Yes
Hearing loss, moth1	Mtap1a	Multiple amino acid substitutions	Yes
Type 1 diabetes	Il2	Proline substituted for serine at position 6	No
Experimental allergic asthma	C5	Deletion of 2 nucleotides causing frameshift in coding sequence and protein truncation	No
Saccharin preference	Tas1r3	Alanine substituted for threonine at position 55 Threonine substituted for isoleucine at position 60	No
Plasmacytoma susceptibility	P16^{INK4a}	Proline substituted for histidine at position 18 Isoleucine substituted for valine at position 51	No
		Rat	
Insulin action, fatty acid metabolism	Cd36	Chromosomal deletion creating chimeric nonfunctional protein	Yes
Type 1 diabetes	Cblb	Stop codon substituted for arginine at position 455	Yes
Blood pressure and 18-hydroxylase activity	Cyp11b1	Leucine substituted for valine at position 381 Leucine substituted for isoleucine at position 384	No
		Drosophila	
ADH activity	Adh	Threonine substituted for lysine at position 192	Yes
ADH concentration	Adh	Intronic, 3'UTR, and silent exonic substitutions	Yes
Bristle number	scabrous, Delta, Achaete-scute	Insertion/deletion or SNP variants within putative regulatory regions	No
		Tomato	
Fruit size, fw2.2	ORFX	Promoter and/or missense coding sequence variants	Yes
Sugar content, Brix9-2-5	Lin5	Missense coding and/or regulatory intronic variants	No
		Rice	
Photoperiod sensitivity, Hd6	CK2α	Stop codon substituted for lysine at position 91	Yes
Photoperiod sensitivity, Hd1	Se1	Deletion of 43 nucleotides within exon 1 Insertion of 433 nucleotides within intron 1 Deletion of 2 nucleotides within exon 2	Yes
		Maize	
Apical dominance	Tb1	Multiple variants in regulatory region QQ	No
		Arabidopsis	
Flowering time, EDI	CRY2	Methionine substituted for valine at position 367	Yes
		S. cerevisiae	
High-temperature growth (Htg)	MKT1, END3, RHO2	Multiple amino acid substitutions	Yes

*CARD15 was originally reported under the gene name NOD2.

polyposis coli gene (*Apc*) are susceptible to intestinal polyps that can lead to colon cancer. The discovery that susceptibility depends on genetic background led to the mapping of a QTL called *Mom1*, which acts as a strong modifier gene of the *Apc* phenotype. The finding that the secretory phospholipase (*Pla2g2a*) gene cosegregates with the *Mom1* phenotype, identification of a single base pair insertion in the *Pla2g2a* coding sequence (*30*), and construction of a transgenic mouse with a *Pla2g2a*-containing cosmid (*31*) together demonstrated that *Pla2g2a* confers resistance to polyp formation in *Apc* mutant mice. Interestingly, a second closely linked phospholipase, *Pla2g4*, also confers resistance to polyp formation in the small intestine (*32*), raising the possibility that the *Mom1* QTL results from the joint action of both phospholipases; this finding calls into question the originally proposed mechanism by which *Pla2g2a* suppresses polyp formation.

Mutations in the tubby (*tub*) gene cause obesity, retinal degeneration, and hearing loss (*33, 34*). A modifier gene (*moth1*) protects tubby mice from hearing loss (*35*). The critical region was reduced to 0.17 cM with crosses involving 1780 progeny (*9*). DNA sequence analysis identified multiple substitutions in the *Mtap1a* cDNA, and a combination of gene targeting and transgenesis showed that a protective allele of *Mtap1a* rescues hearing loss (*9*).

A combined linkage and microarray analysis identified several genes that are differentially expressed in hypertensive, normotensive, and congenic spontaneously hypertensive rats (SHR) (*15, 36*). From this analysis, a biological candidate, *Cd36*, was found to map to a QTL for defective insulin action and fatty acid metabolism, and a deletion of *Cd36* was associated with this SHR phenotype (*15, 37*). Transgenesis was used to complement the *Cd36*-deficient phenotype in SHR, although the same adipocyte traits used for QTL mapping were not used in the transgenic complementation test (*38*).

In the Komeda diabetes-prone (KDP) model of type 1 diabetes, QTL mapping localized a non-MHC gene to a 3.0-cM region of chromosome 11 (*39, 40*). A nonsense mutation in the *Cblb* gene, a member of the Cbl/Sli family of ubiquitin-protein ligases, is found in Komeda rats (*40*), and *Cblb*-deficient mice have an autoimmune phenotype (*41*). Transgenic rescue demonstrated that *Cblb* contributes to the diabetes-prone phenotype in Komeda rats (*40*).

Humans. Genetic linkage of type 1 diabetes to the MHC genes *HLA-DR* and *-DQ* was established more than 20 years ago. Several lines of evidence have placed the role of DQβ^{57} in susceptibility to type 1 diabetes beyond reasonable doubt (*42–44*): (i) conservation of diabetes-encoded sus-

ceptibility at amino acid 57 in the mouse ortholog of the *HLA-DQ* gene in the NOD mouse (*42*), (ii) consistent association between DQβ^{57} and type 1 diabetes in different populations (*43*), and (iii) the finding that amino acid 57 is key to the structure of the DQB molecule (*44*). Conserved association between the *HLA-DQA* gene and type 1 diabetes has also been demonstrated in several populations (*45*).

In type 2 diabetes, a genome-wide screen identified a region on chromosome 2 that is strongly linked to disease (*46*). By looking for interaction with other linked regions, the region of interest was narrowed to 7 cM, which fortuitously spanned only 1700 kbp. The region contained multiple SNPs associated with diabetes, three of which, in the region of the gene encoding calpain-10 (*CAPN10*), formed a susceptibility haplotype in Mexican Americans and Northern Europeans that together with a second susceptibility haplotype were proposed to affect diabetes susceptibility (*47*). Because some (*48*) but not all (*49*) subsequent studies replicated the association between *CAPN10* and diabetes or plasma glucose, and because the mechanism by which *CAPN10* haplotypes cause diabetes susceptibility remains uncertain, it is likely either that *CAPN10*-mediated susceptibility is limited to certain populations or that other genes within the haplotype are involved.

An apolipoprotein E4 (*APOE4*) allele is associated in a dose-dependent manner with susceptibility to Alzheimer's disease (*50, 51*). This genetic association, together with the presence of the APOE4 protein in brain lesions and the role of ApoE in amyloid deposition (*52*), provides strong evidence for the direct role of the *APOE4* allele in susceptibility to Alzheimer's disease.

Two groups showed that *NOD2* (now known as *CARD15*) is a susceptibility gene for Crohn's disease (*53, 54*). Using a positional candidate approach based on linkage analysis and association studies, both groups identified frameshift and missense variants within the *NOD2* gene that were associated with Crohn's disease but not with ulcerative colitis, another inflammatory bowel disease. Highly localized linkage disequilibrium mapping in subsequent confirmatory reports (*55, 56*) and strong biological candidacy make it highly likely that *NOD2/CARD15* is a primary Crohn's disease gene, although confirmation with knock-in studies in mice is needed.

Nature of Molecular Variants

Although the number of complex traits for which proof is available is small, they provide the first glimpses into the DNA sequence variation that underlies these phenotypes (Table 1). Some phenotypes are caused

by single-nucleotide variants (e.g., ADH catalytic efficiency, *Cblb* in diabetes), others by multiple nucleotides in single genes (e.g., ADH protein level, *Mtap1a* in hearing loss) or by multiple nucleotides in closely linked genes (*MKT1*, *END3*, and *RHO2* in high-temperature growth). The causative lesions include small and large deletions (e.g., *C5* in allergic asthma, *Cd36* in fatty acid metabolism); they can be nucleotide variants in the coding region (e.g., *DGAT1* in milk composition) or in the noncoding regulatory regions (e.g., *Tb1* in apical dominance). It is striking that several of the identified QTLs (*Mom1*, *moth1*) were found in surveys involving modifier genes. Phenotype modification occurs when expression of one gene alters the phenotype normally conferred by another gene (*1*). Typically, the modifier has little if any detectable phenotypic effect on its own, but can cause subtle or profound changes in the expression of the phenotype caused by mutation at another gene locus. This supports the proposal that study of modifier genes is an effective means to simplify the analysis of complex traits (*57*). Obviously, as genes and variants that are responsible for other complex traits are identified in conventional, modifier, and regulatory surveys, a better sense will emerge of the variety of sequence variants and their relative frequencies.

References and Notes
1. J. H. Nadeau, *Nature Rev. Genet.* **2**, 165 (2001).
2. E. S. Lander, N. J. Schork, *Science* **265**, 2037 (1994).
3. N. J. Risch, *Nature* **405**, 847 (2000).
4. See supporting data on *Science* Online.
5. E. Lander, L. Kruglyak, *Nature Genet.* **11**, 241 (1995).
6. N. Risch, K. Merikangas, *Science* **273**, 1516 (1996).
7. I. Dahlman *et al.*, *Nature Genet.* **30**, 149 (2002).
8. R. W. Doerge, G. A. Churchill, *Genetics* **142**, 285 (1996).
9. A. Ikeda *et al.*, *Nature Genet.* **30**, 401 (2002).
10. A. Frary *et al.*, *Science* **289**, 85 (2000).
11. M. Yano *et al.*, *Plant Cell* **12**, 2473 (2000).
12. Y. Takahashi, A. Shomura, T. Sasaki, M. Yano, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 7922 (2001).
13. S. B. Gabriel *et al.*, *Science* **296**, 2225 (2002).
14. L. F. Stam, C. C. Laurie, *Genetics* **144**, 1559 (1996).
15. T. J. Aitman *et al.*, *Nature Genet.* **21**, 76 (1999).
16. M. E. Legare, F. S. Bartlett, W. N. Frankel, *Genome Res.* **10**, 42 (2000).
17. L. M. Steinmetz *et al.*, *Nature* **416**, 326 (2002).
18. D. W. Nebert, T. P. Dalton, G. W. Stuart, M. J. I. Carvan, *Ann. N.Y. Acad. Sci.* **919**, 148 (2000).
19. T. F. Mackay, *Nature Rev. Genet.* **2**, 11 (2001).
20. M.-C. King, A. C. Wilson, *Science* **188**, 107 (1975).
21. R. Korstanje, B. Paigen, *Nature Genet.* **31**, 235 (2002).
22. D. J. Symula *et al.*, *Nature Genet.* **23**, 241 (1999).
23. C. R. Cowles, J. N. Hirschhorn, D. Altshuler, E. S. Lander, *Nature Genet.* **32**, 432 (2002).
24. M. Choudhary, C. C. Laurie, *Genetics* **129**, 481 (1991).
25. W. Coppieters *et al.*, *Mamm. Genome* **9**, 540 (1998).
26. B. Grisart *et al.*, *Genome Res.* **12**, 222 (2002).
27. H. Nishimoto, H. Kikutani, K. Yamamura, T. Kishimoto, *Nature* **328**, 432 (1987).
28. T. Lund *et al.*, *Nature* **345**, 727 (1990).
29. T. Miyazaki *et al.*, *Nature* **345**, 722 (1990).
30. M. MacPhee *et al.*, *Cell* **81**, 957 (1995).
31. R. T. Cormier *et al.*, *Nature Genet.* **17**, 88 (1997).
32. K. H. Hong, J. C. Bonventre, E. O'Leary, J. V. Bonventre, E. S. Lander, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 3935 (2001).
33. K. Noben-Trauth, J. K. Naggert, M. A. North, P. M. Nishina, *Nature* **380**, 534 (1996).

34. P. W. Kleyn et al., Cell 85, 281 (1996).
35. S. Ikeda et al., Invest. Ophthalmol. Vis. Sci. 40, 2706 (1999).
36. T. J. Aitman et al., Nature Genet. 16, 197 (1997).
37. A. M. Glazier, J. Scott, T. J. Aitman, Mamm. Genome 13, 108 (2002).
38. M. Pravenec et al., Nature Genet. 27, 156 (2001).
39. N. Yokoi et al., J. Clin. Invest. 100, 2015 (1997).
40. N. Yokoi et al., Nature Genet. 31, 391 (2002).
41. K. Bachmaier et al., Nature 403, 211 (2000).
42. J. A. Todd, J. I. Bell, H. O. McDevitt, Nature 329, 599 (1987).
43. J. A. Todd et al., Science 240, 1003 (1988).
44. J. A. Todd, L. A. Wicker, Immunity 15, 387 (2001).
45. J. A. Todd et al., Nature 338, 587 (1989).
46. C. L. Hanis et al., Nature Genet. 13, 161 (1996).
47. Y. Horikawa et al., Nature Genet. 26, 163 (2000).
48. P. G. Cassell et al., Diabetes 51, 1622 (2002).
49. T. E. Fingerlin et al., Diabetes 51, 1644 (2002).
50. M. A. Pericak-Vance et al., Am. J. Hum. Genet. 48, 1034 (1991).
51. E. H. Corder et al., Science 261, 921 (1993).
52. K. R. Bales et al., Proc. Natl. Acad. Sci. U.S.A. 96, 15233 (1999).
53. J. P. Hugot et al., Nature 411, 599 (2001).
54. Y. Ogura et al., Nature 411, 603 (2001).
55. J. Hampe et al., Lancet 359, 1661 (2002).
56. L. Murillo, J. B. A. Crusius, A. A. van Bodegraven, B. Z. Alizadeh, A. S. Pena, Immunogenetics 54, 59 (2002).
57. A. Matin, J. H. Nadeau, Trends Genet. 17, 727 (2001).
58. V. McKusick, personal communication.
59. L. Peltonen, V. A. McKusick, Science 291, 1224 (2001).
60. Supported by the MRC Clinical Sciences Centre and Wellcome Trust Cardiovascular Functional Genomics initiative (T.J.A.) and the Charles B. Wang Foundation (J.H.N.). We thank T. Mackay, M. Burmeister, and C. Wallace for careful reading and constructive criticism of the manuscript.