

Supporting Online Material and Tables S1 and S2.

Supporting Data

Characteristics of genetically complex traits that complicate fine mapping and positional cloning: The expression of complex trait phenotypes is dependent on the interaction between multiple genes (**epistasis**), and between genetic and environmental influences (**gene-environment interactions**). Different sets of genes may give rise to the same phenotype (**locus heterogeneity**), whilst the presence of a susceptibility gene can be found in phenotypically distinct individuals (**variable penetrance**). Environmental mimics of the trait (**phenocopies**) may be indistinguishable from cases with a more strongly inherited basis. These features result in broad likelihood confidence intervals in linkage studies and confound efforts to fine map the location of the underlying genes. Use of linkage disequilibrium mapping in humans (1), specialised inbred strains or genetically heterogeneous stocks in model organisms (2, 3, 4, 5) progeny testing in experimental crosses (6), and the study of very large numbers of meioses, are amongst the main strategies employed to fine map complex trait genes.

General limitations to fine mapping and positional cloning of complex trait genes: The greater degree of scientific control in experimental and model organisms compared to human genetics has allowed mapping of complex trait genes in experimental crosses to proceed faster, and with far greater statistical certainty, than in humans. This is in part because of the ease of increasing pedigree and sample size in experimental crosses, and because of the ability to breed and phenotype the offspring of individuals with critical recombinants, a procedure known as **progeny testing** (2, 6). In addition, specialised resources are available to experimental geneticists such as inbred, consomic and congenic strains in rodents (7, 8, 9, 10), nearly isogenic lines in plants (11, 12) and deficiency stocks in *Drosophila* (3) that reduce or eliminate allelic and locus heterogeneity. Finally, by using a method known as '**Mendelizing QTLs**', single gene mutations can be used to sensitize QTL analysis in humans and other species (13). These sensitising mutations convert QTLs from background genes with weak phenotypic effects to variants with strong effects that segregate as oligogenic if not Mendelian traits. As a result, they are more amenable to genetic and functional characterization and to positional cloning.

Using these resources, direct rather than indirect tests of linkage can be carried out, with markers that are 100% informative in all meioses. These approaches have yielded robust, highly statistically significant linkages to complex traits in large numbers of studies in rodents, plants and *Drosophila*. Statistically significant, replicable linkages in humans have been harder to obtain.

A crucial factor in proceeding from map location to gene identification is the relationship between genetic and physical distance, which varies markedly between species (table S2). Because the physical distance spanned in a given genetic interval is much greater in rodents than in plants and flies, a complex trait mapped to a 10cM interval in rats or mice will on average require a ten-fold longer segment of chromosome to be searched than, for example, in *Arabidopsis* or rice. Ultimately the ability to fine map a given complex trait gene down to a chromosomal region that can be screened by sequencing (perhaps less than 1-2Mbp) will rest on the size of the gene effect, the size of the study population, the isolation of the gene (where possible) in congenic or nearly isogenic lines, and the relationship between genetic and physical distance for the particular chromosomal segment under study.

Special strategies for fine mapping in humans: The low power of genetic linkage for mapping complex trait genes in humans (1) has led to increasing focus on whole genome and localised tests of linkage disequilibrium, with some early indications of success (14, 15, 16). Massive resources have been deployed for producing dense single nucleotide (SNP) maps of the human and mouse genomes, in preparation for whole genome tests of association, using either individual SNPs or more likely groups of SNPs representing the majority of common haplotypes (17, 18, 19, 20, 21). These study designs potentially require very large sample sizes and generation of huge numbers (tens of millions) of genotypes, for which cost-efficient methodologies are only just becoming available.

Table S1. Molecular basis of complex trait genes localized initially in genome-wide linkage studies.

Trait	Gene ^a	Relative risk	Phenotypic variance	Molecular basis	Formal complementation?	Ref.
<i>Human</i>						
Type 1 diabetes	<i>HLA-DQB</i>	107 ^b	-	Aspartic acid at position 57 is protective	No	(22, 23)
Type 1 diabetes	<i>HLA-DQA</i>	23 ^c	-	Multiple amino acid substitutions	No	(24)
Type 2 diabetes	<i>CAPN10</i>	2.6-5.0 ^d	-	Intronic haplotype	No	(14)
Crohn's Disease	<i>CARD15</i>	1.5-3 ^c 17.6-44 ^f	-	Multiple amino acid substitutions, single nucleotide insertion at nucleotide position 3020 causing frameshift and protein truncation	No	(15, 16)
Alzheimer's disease	<i>ApoE</i>	2.84 ^g	-	Arginine substituted for cysteine at position 112	No	(25, 26)
Asthma	<i>ADAM33</i>	na	-	Coding and non-coding haplotype pairs	No	(27)
Angiotensin I-converting enzyme levels	<i>ACE</i>	-	27%	Multiple intragenic SNPs	No	(28, 29)
<i>Cattle</i>						
Milk yield Milk composition	<i>DGATI</i>	-	3% ^h 31% ^h	Alanine substituted for lysine at position 232	No	(30)
<i>Mouse</i>						
Intestinal neoplasia, <i>Mom1</i>	<i>Pla2g2a</i>	-	21-43%	Insertion of thymidine at nucleotide position 208	Yes	(31, 32)
Type 1 diabetes	<i>I-Eα</i> <i>I-Aβ</i>	42	-	Promoter deletion Histidine substituted for proline at	Yes Yes	(33, 34, 35, 36)

				position 56 Serine substituted for aspartic acid at position 57		
Hearing loss, <i>moth1</i>	<i>Mtap1a</i>	-	43-57%	Multiple amino acid substitutions	Yes	(6)
Type 1 diabetes	<i>Il2</i>	2.7	-	Proline substituted for serine at position 6	No	(37, 38)
Experimental allergic asthma	<i>C5</i>	-	8%	Deletion of 2 nucleotides causing frameshift in coding sequence and protein truncation	No	(39, 40, 41)
Saccharin preference	<i>Tas1r3</i>	-	31.2%	Alanine substituted for threonine at position 55 Threonine substituted for isoleucine at position 60	No	(42, 43)
Plasmacytoma susceptibility	<i>P16^{INK4a}</i>	4.9 ⁱ	-	Proline substituted for histidine at position 18 Isoleucine substituted for valine at position 51	No	(44, 45)
<i>Rat</i>						
Insulin action Fatty acid metabolism	<i>Cd36</i>	-	11-15% 47%	Chromosomal deletion creating chimeric non-functional protein	Yes	(46, 47, 48)
Type 1 diabetes	<i>Cblb</i>	4.4	-	Stop codon substituted for Arginine at position 455	Yes	(49, 50)
Blood pressure and 18- hydroxylase activity	<i>Cyp11b1</i>	30 ^j	-	Leucine substituted for valine at position 381 Leucine substituted for isoleucine at position 384	No	(51, 52, 53)
<i>Drosophila</i>						
ADH activity	<i>Adh</i>	-	67-80%	Threonine substituted for lysine at position 192	Yes	(54)

ADH concentration	<i>Adh</i>	-	variable	Intronic, 3' UTR and silent exonic substitutions	Yes	(55)
Bristle number	<i>scabrous</i> , <i>Delta</i> , <i>Achaete-scute</i>	-	21%-32% ^k 6%-12% ^k 5% ^k	Insertion/deletion or SNP variants within putative regulatory regions	No	(56) (57) (58)
Tomato						
Fruit size <i>fw2.2</i>	<i>OFRX</i>	-	5-30%	Promoter and/or missense coding sequence variants	Yes	(11)
Sugar content <i>Brix9-2-5</i>	<i>Lin5</i>	-	na	Missense coding and/or regulatory intronic variants	No	(59)
Rice						
Photoperiod sensitivity <i>Hd6</i>	<i>CK2α</i>	-	58.7%	Stop codon substituted for lysine at position 91	Yes	(60, 61)
Photoperiod sensitivity <i>Hd1</i>	<i>Se1</i>	-	67%	Deletion of 43 nucleotides within exon 1 Insertion of 433 nucleotides within intron 1 Deletion of 2 nucleotides within exon 2	Yes	(62)
Maize						
Apical dominance	<i>Tb1</i>	-	20-25%	Gene regulation	No	(63, 64, 65, 66)
Arabidopsis						
Flowering time <i>EDI</i>	<i>CRY2</i>	-	27.5%-56.2%	Methionine substituted for valine at position 367	Yes	(12)
S. cerevisiae						
High-temperature growth (Htg)	<i>MKT1</i> , <i>END3</i> , <i>RHO2</i>	30.6	-	Multiple amino acid substitutions	Yes	(67)

^a The genes listed were identified according to the criteria of proof described in the text but may not include every complex trait gene identified.

^b Genotype relative risk

^c Allele relative risk

^d Relative risk of haplotype combination

^e Heterozygous risk

^f Homozygous and compound heterozygous risk

^g Increase in risk per ApoE 4 allele

^h Effect of K232A variant

ⁱ Relative risk calculated from (45)

^j Relative risk calculated from (51)

^k Percentage of total genetic variance

na, not available

Table S2. Physical and genetic lengths of genomes commonly used for complex trait analysis

Species	Mbp in Genome	cM in Genome	Mean Mbp/cM	References
<i>S. cerevisiae</i>	12	4100	0.003	(68, 69)
<i>Arabidopsis</i>	125	500	0.25	(70, 71)
Rice	466	1522	0.31	(72, 73, 74)
<i>Drosophila</i>	120	281	0.43	(75, 76)
Tomato	900	1200	0.75	(77)
Human	2910	3699	0.79	(17, 78, 79)
Maize	2500	2033	1.23	(80, 81)
Rat	2800	1749	1.60	http://public.bcm.tmc.edu/pa/rgsc-genome.htm , (82)
Mouse	2500	1398	1.78	(83, 84)

References

1. N. Risch, K. Merikangas, *Science* 273, 1516-1517 (1996).
2. A. Darvasi, *Nature Genet* 18, 19-24 (1998).
3. T. F. Mackay, *Nature Rev Genet* 2, 11-20 (2001).
4. R. Mauricio, *Nature Rev Genet* 2, 370-381 (2001).
5. J. Flint, R. Mott, *Nature Rev Genet* 2, 437-445 (2001).
6. A. Ikeda, et al., *Nature Genet* 30, 401-405 (2002).
7. A. Darvasi, M. Soller, *Behav Genet* 27, 125-132 (1997).
8. A. Matin, G. B. Collin, Y. Asada, D. Varnum, J. H. Nadeau, *Nature Genet* 23, 237-240 (1999).
9. J. H. Nadeau, J. B. Singer, A. Matin, E. S. Lander, *Nature Genet* 24, 221-225 (2000).
10. H. J. Jacob, A. E. Kwitek, *Nature Rev Genet* 3, 33-42 (2002).
11. A. Frary, et al., *Science* 289, 85-88 (2000).
12. S. El-Din El-Assal, C. Alonso-Blanco, A. J. Peeters, V. Raz, M. Koornneef, *Nature Genet* 29, 435-440 (2001).
13. A. Matin, J. H. Nadeau, *Trends Genet* 17, 727-731 (2001).
14. Y. Horikawa, et al., *Nature Genet* 26, 163-175 (2000).
15. J. P. Hugot, et al., *Nature* 411, 599-603 (2001).
16. Y. Ogura, et al., *Nature* 411, 603-606 (2001).
17. J. C. Venter, et al., *Science* 291, 1304-1351 (2001).
18. M. J. Daly, J. D. Rioux, S. F. Schaffner, T. J. Hudson, E. S. Lander, *Nature Genet* 29, 229-232 (2001).
19. G. C. Johnson, et al., *Nature Genet* 29, 233-237 (2001).
20. J. Couzin, *Science* 296, 1391-1393 (2002).
21. S. B. Gabriel, et al., *Science* 296, 2225-2229 (2002).
22. J. A. Todd, J. I. Bell, H. O. McDevitt, *Nature* 329, 599-604 (1987).
23. P. M. Morel, J. S. Dorman, J. A. Todd, H. O. McDevitt, M. Trucco, *Proc Natl Acad Sci* 85, 8111-8115 (1988).
24. J. A. Todd, et al., *Nature* 338, 587-589 (1989).
25. E. H. Corder, et al., *Science* 261, 921-923 (1993).
26. R. L. Raffai, L. M. Dong, R. V. J. Farese, K. H. Weisgraber, *Proc Natl Acad Sci U S A* 98, 11587-11591 (2001).
27. P. Van Eerdewegh, et al., *Nature* 418, 426-430 (2002).
28. C. A. McKenzie, et al., *Am J Hum Genet* 57, 1426-1435 (1995).
29. R. Cox, et al., *Hum Mol Genet* 11, 2969-2977 (2002).
30. B. Grisart, et al., *Genome Res* 12, 222-231 (2002).
31. M. MacPhee, et al., *Cell* 81, 957-966 (1995).

32. W. F. Dietrich, et al., *Cell* 75, 631-639 (1993).
33. H. Nishimoto, H. Kikutani, K. Yamamura, T. Kishimoto, *Nature* 328, 432-434 (1987).
34. T. Lund, et al., *Nature* 345, 727-729 (1990).
35. T. Miyazaki, et al., *Nature* 345, 722-724 (1990).
36. L. S. Wicker, et al., *J Exp Med* 165, 1639-1654 (1987).
37. P. L. Podolin, et al., *Cytokine* 12, 477-482 (2000).
38. S. Ghosh, et al., *Nature Genet* 4, 404-409 (1993).
39. S. L. Ewart, et al., *Am. J. Respir. Cell Mol. Biol.* 23, 537-545 (2000).
40. C. L. Karp, et al., *Nature Immunology* 1, 221-226 (2000).
41. R. A. Wetzel, D. T. Fleischer, D. L. Haviland, *J Biol Chem* 265, 2435-2440 (1990).
42. D. A. Blizard, B. Kotlus, M. E. Frank, *Chem Senses* 24, 373-385 (1999).
43. M. Max, et al., *Nature Genet* 28, 58-63 (2001).
44. S.-L. Zhang, et al., *Mol Cell Biol* 21, 310-318 (2001).
45. B. A. Mock, M. M. Krall, J. D. Dosik, *Proc Natl Acad Sci USA* 90, 9499-9503 (1993).
46. T. J. Aitman, et al., *Nature Genet* 16, 197-201 (1997).
47. T. J. Aitman, et al., *Nature Genet* 21, 76-83 (1999).
48. M. Pravenec, et al., *Nature Genet* 27, 156-158 (2001).
49. N. Yokoi, et al., *J Clin Invest* 100, 2015-2021 (1997).
50. N. Yokoi, et al., *Nature Genet* 31, 391-394 (2002).
51. G. T. Cicila, et al., *Nature Genetics* 3, 346-353 (1993).
52. G. T. Cicila, et al., *Genomics* 72, 51-60 (2001).
53. Y. Nonaka, et al., *Eur J Biochem* 258, 869-878 (1998).
54. M. Choudhary, C. C. Laurie, *Genetics* 129, 481-488 (1991).
55. L. F. Stam, C. C. Laurie, *Genetics* 144, 1559-1564 (1996).
56. C. Lai, R. F. Lyman, A. D. Long, C. H. Langley, T. F. C. Mackay, *Science* 266, 1697-1702 (1994).
57. A. D. Long, R. F. Lyman, C. H. Langley, T. F. C. Mackay, *Genetics* 149, 999-1017 (1998).
58. T. F. Mackay, C. H. Langley, *Nature* 348, 64-66 (1990).
59. E. Fridman, T. Pleban, D. Zamir, *Proc Natl Acad Sci U S A* 97, 4718-4723 (2000).
60. Y. Takahashi, A. Shomura, T. Sasaki, M. Yano, *Proc Natl Acad Sci U S A* 98, 7922-7927 (2001).
61. T. Yamamoto, H. Lin, T. Sasaki, M. Yano, *Genetics* 154, 885-891 (2000).
62. M. Yano, et al., *Plant Cell* 12, 2473-2484 (2000).
63. J. Doebley, A. Stec, C. Gustus, *Genetics* 141, 333-346 (1995).
64. R. L. Wang, A. Stec, J. Hey, L. Lukens, J. Doebley, *Nature* 398, 236-239 (1999).
65. J. Doebley, A. Stec, *Genetics* 134, 559-570 (1993).
66. J. Doebley, A. Stec, L. Hubbard, *Nature* 386, 485-488 (1997).
67. L. M. Steinmetz, et al., *Nature* 416, 326-330 (2002).
68. H. W. Mewes, et al., *Nature* 387, 7-65 (1997).
69. J. M. Cherry, et al., *Nature* 387, 67-73 (1997).
70. The Arabidopsis Genome Initiative, *Nature* 408, 796-815 (2000).
71. C. Chang, J. L. Bowman, A. W. DeJohn, E. S. Lander, E. M. Meyerowitz, *Proc Natl Acad Sci USA* 85, 6856-6860 (1988).
72. S. A. Goff, *Science* 296, 92-100 (2002).
73. J. Yu, *Science* 296, 79-92 (2002).
74. Y. Harushima, et al., *Genetics* 148, 479-494 (1998).
75. M. D. Adams, *Science* 287, 2185-2195 (2000).
76. D. L. Lindsley, G. G. Zimm, *The genome of Drosophila melanogaster* (Academic Press, 1992).
77. S. D. Tanksley, et al., *Genetics* 132, 1141-1160 (1992).
78. E. S. Lander, *Nature* 409, 860-921 (2001).
79. C. Dib, et al., *Nature* 380, 152-154 (1996).
80. H. Fu, Z. Zheng, H. K. Dooner, *Proc Natl Acad Sci U S A* 99, 1082-1087 (2002).
81. D. Q. Xiang, et al., *Yi Chuan Xue Bao* 28, 778-784 (2001).
82. A. M. Brown, et al., *Mamm Genome* 9, 521-530 (1998).
83. R. H. Waterston, et al., *Nature* 420, 520-562 (2002).
84. W. F. Dietrich, et al., *Nature* 380, 149-152 (1996).