

USE OF POPULATION ISOLATES FOR MAPPING COMPLEX TRAITS

Leena Peltonen‡, Aarno Palotie*§ and Kenneth Lange*||*

Geneticists have repeatedly turned to population isolates for mapping and cloning Mendelian disease genes. Population isolates possess many advantages in this regard. Foremost among these is the tendency for affected individuals to share ancestral haplotypes derived from a handful of founders. These haplotype signatures have guided scientists in the fine mapping of scores of rare disease genes. The past successes with Mendelian disorders using population isolates have prompted unprecedented interest among medical researchers in both the public and private sectors. Despite the obvious genetic and environmental complications, geneticists have targeted several population isolates for mapping genes for complex diseases.

MENDELIAN VARIANT
An individual phenotype that is due to a single gene.

The initial high expectations concerning the use of population isolates for mapping complex disease genes are now in danger of being eroded. Although clearly advantageous, this approach does not guarantee success. Furthermore, isolates vary in their history, genetic background and shared environment. Different populations warrant different study designs — especially for complex traits. Some of the pros and cons for the use of different populations are summarized in BOX 1 and are further discussed below. With the exception of MENDELIAN VARIANTS of complex diseases, such as **Hirschsprung disease** in the Amish or **nonsyndromic hearing loss** in Bedouins^{1–3}, population isolates have yielded few cloned genes and recently their benefits have been challenged.

Of course, mapping in large outbred populations has been just as disappointing. Many loci for common diseases have been mapped, but few have been pinned down to narrow chromosomal intervals. Several of the mapped loci are statistical ghosts that appear in some studies and disappear in others. There are several possible reasons for these inconsistencies: genetic heterogeneity, both at the allelic and locus levels; insufficient sample size; imperfect statistical analysis; diagnostic and genotyping errors; and pooling of diverse phenotypes into the same diagnostic classes.

Once mapping is accomplished, positional cloning of genes for common diseases is certainly much harder than for rare Mendelian diseases. Nonetheless, to para-

phrase the English mathematician Littlewood, it is too early to be prematurely pessimistic⁴. New genetic information and laboratory tools on the horizon will reinvigorate our mapping efforts in the twenty-first century. Our chances of success will be improved by a detailed catalogue of the entire human genome, high-throughput low-cost genotyping and microarray chips, hundreds of thousands of single nucleotide polymorphisms (SNPs), and by steady advances in basic biology, computing and statistics. Although we focus here on the utility of population isolates, we touch on these larger issues because they so strongly affect study designs and the analysis of collected data.

Features of isolated populations
Given the current size of the world's population, the human genome is less diverse than might be expected. This observation is explained in part by two phenomena — the recent divergence of humans from other primates and the relatively small size of the human population over most of its history. In prehistoric times, the human population expanded in size as it migrated into new territory. The early waves of migration of anatomically modern humans out of Africa about 100,000 years ago⁵ almost certainly involved small groups; the resulting genetic bottlenecks account for the low genetic diversity of the rest of the world compared with Africa⁶. From 50,000 to 30,000 years ago, mankind began migrating

*Department of Medical Genetics, University of Helsinki and National Public Health Institute, Finland. ‡Department of Human Genetics, §Department of Pathology and Laboratory Medicine and ||Department of Biomathematics, University of California at Los Angeles School of Medicine, California 90095, USA. Correspondence to: L.P.‡ e-mail: Lpeltonen@mednet.ucla.edu

Box 1 | Isolated versus outbred populations

Benefits of population isolates

- Higher prevalence for some diseases
- More inbreeding and the opportunity to map recessive genes
- More uniform genetic background
- Good genealogical records
- Easier to standardize phenotype definitions
- Wider intervals of linkage disequilibrium
- Closer to Hardy–Weinberg equilibrium
- Less migration and more intact families
- More uniform environment

Benefits of outbred populations

- More affected people
- More opportunity for replication
- Markers more polymorphic
- Genes mapped pertinent to more of humanity

into new regions, such as the Americas and Australia. The bottleneck in the American migration manifests itself in the limited genetic diversity of South American Indians. A third wave of migration and growth occurred about 10,000 years ago with the spread of agriculture after the last glacial period.

Genetic consequences of a bottleneck. Population isolates show significantly less genetic diversity than humanity as a whole on the basis of analyses of both chromosomal and mitochondrial DNA^{7,8}. By definition, all population isolates start with a small group of founders. Many isolates experience bottlenecks alternating with periods of rapid growth. Famine, war, environmental disruption and infectious disease epidemics can all create POPULATION BOTTLENECKS^{9–11}. The demographic history of Finland illustrates the formation of both a middle-aged population and several young population isolates within one geographical region (BOX 2). As an isolate rebounds from a crash, it experiences more inbreeding and GENETIC DRIFT. The longer the rebound takes, the more opportunity there is for genetic drift. Immigration counteracts the effects of isolation, and can restore diversity quickly if it occurs before the isolate expands markedly in size. Compared with immigration, mutation increases diversity on a much slower timescale.

Genetic drift has a profound effect on the genotypic makeup of a population isolate. Recessive disease alleles and neutral alleles are both subject to genetic drift in an isolate. Some disease alleles are lost through drift, and some are pushed to much higher frequencies. Every population isolate consequently has its own special set of recessive diseases. In contrast to recessive alleles, dominant disease alleles are weeded out more quickly unless they involve negligible reduction in fitness (for example, late onset). For this reason, dominant disease alleles tend to be more uniformly distributed across the world. Drift affects rare marker alleles and rare haplotypes in much the same way as it does recessive disease

alleles. Common marker alleles and common haplotypes are seldom totally lost from an isolate unless the initial number of founders is very small^{9,12,13}.

The importance of demographic history. Population isolates can have strikingly different demographic histories. Unfortunately, for most isolates we lack reliable information on their initial genetic makeup, their total number of founders, and the extent and duration of their isolation. Only a handful of isolates, such as Iceland, northern Sweden or Finland, have easily accessible genealogical records. We can roughly infer the extent of a population's isolation by its trademark genetic diseases. Older isolates tend to have fewer such diseases because selection has had a longer time to operate. More systematic analysis of DNA variation between and within populations is clearly needed⁶. Such research, which is in its infancy, is a stated goal of the **Human Genome Project**.

Some population isolates have proved more valuable than others in genetic studies^{14,15}. Isolates such as the Lapps of Scandinavia (Saami-populations) and the Basques of southern Europe are well established (200–400 generations old) and demographically stable¹². Because of their small size and paucity of recessive diseases, these groups have been largely ignored by geneticists. Of more value in mapping rare diseases are the Finns, Amish, Sardinians and Bedouins. These isolates also show a high frequency of certain Mendelian disorders and many disease genes have been identified in them. For studies of complex traits, geneticists have tended to target younger population isolates (10–20 generations old) that originated from a small number of founders and underwent rapid population expansion. Examples are certain populations in eastern Finland, Costa Rica, Quebec, Newfoundland and some regions of Holland. Recessive disease genes have been mapped using study samples of only 5–10 affected individuals and widely spread markers^{16,17}. The hope is that the reduced genetic diversity found in these isolates will simplify the study of complex traits. Even if the same genes come into play, it would be reasonable to expect fewer alleles and recognizable HAPLOTYPE SIGNATURES among the affected individuals in young population isolates.

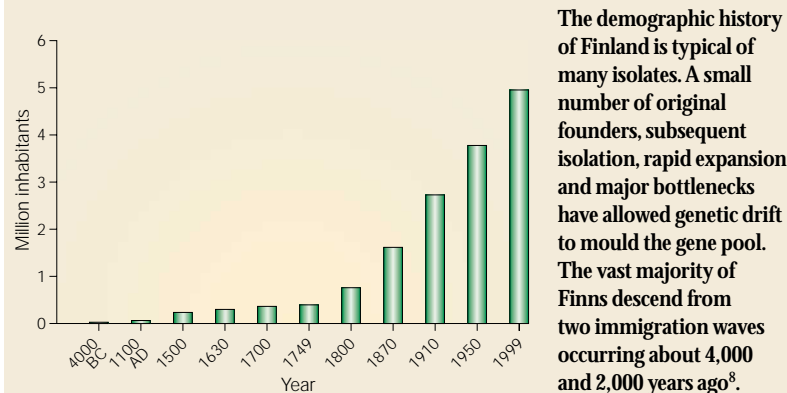
Environmental and cultural features. Population isolates offer other advantages in addition to reduced genetic complexity. The people in most isolates share a common environment and culture. Differences in diet, exercise, sanitary conditions and exposure to infectious diseases are minimized. A common language and religion usually promote social cohesion. Therefore, researchers studying groups such as specific Bedouin tribes or the Amish avoid some of the environmental noise surrounding complex diseases, which are determined by a combination of nature and nurture. Small isolated populations also offer unusual opportunities for the standardization of diagnostic and phenotypic criteria, which increases diagnostic reliability. One example is Finland, where five medical schools with shared academic traditions train all the clinicians in the country.

POPULATION BOTTLENECK
A marked reduction in population size followed by the survival and expansion of a small random sample of the original population.

GENETIC DRIFT
The random fluctuation in allele frequencies as genes are transmitted from one generation to the next.

HAPLOTYPE SIGNATURE
The haplotype surrounding a particular disease susceptibility allele. The haplotype signature can be identified among the affected individuals of an isolated population.

Box 2 | Finnish population history



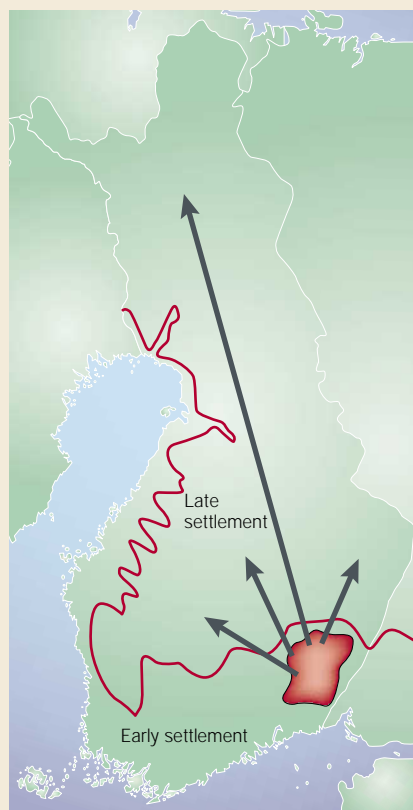
The demographic history of Finland is typical of many isolates. A small number of original founders, subsequent isolation, rapid expansion and major bottlenecks have allowed genetic drift to mould the gene pool. The vast majority of Finns descend from two immigration waves occurring about 4,000 and 2,000 years ago⁸. The earlier wave involved

eastern Uralic speakers, and the later wave, Indo-European speakers from the south. Both Y-chromosomal haplotypes and mitochondrial sequences show the low genetic diversity among Finns compared with other European populations and confirm the long-standing isolation of Finland. The size of the founding population(s) is unknown, but as late as the twelfth century, the population of Finland was only about 50,000, as shown in the top figure. It reached 400,000 by the mid-seventeenth century, only to experience the great famine of 1696–1698, where one third of the population perished. Since then, the Finnish population has grown relatively rapidly from 250,000 at the beginning of the eighteenth century to its present figure of 5,100,000.

Starting in the sixteenth century, during the reign of the Swedish King Gustavus of Vasa (1523–1560), internal migrations created regional subisolates (as shown in the bottom figure). The population spread from the early settlement region on the southern and western coastline towards the east and north. The subisolates in the late settlement region were established for the most part by groups of farmers originating from a small area of South Savo in southeastern Finland. They moved to the central, then western, and finally northern parts of the country, clearing the land by fire.

Within a century, the inhabited land area of Finland doubled. Until the Second World War, many of these northeastern settlements grew rapidly without further immigration to supplement the descendants of their 40–60 founding families⁶⁶. The reign of Gustavus of Vasa also saw the establishment of a national system of population records, an important resource for later genetic studies of Finns. Using these records, many deceased individuals can be traced to common ancestors, especially in the subisolates of the late settlement region.

Finland's demographic history has led to a unique catalogue of genetic diseases. Around 30, mostly recessive diseases, are highly enriched in Finland. Other diseases, such as phenylketonuria and cystic fibrosis, are almost non-existent. Molecular studies have exposed one major mutation (78–98% alleles) in most Finnish Mendelian diseases and have revealed long genetic intervals of linkage disequilibrium (LD) flanking the disease gene¹⁴, with the length of the LD interval reflecting the age of the mutation.



In addition to these advantages, national population registries are available for certain isolated populations. In countries such as Iceland, Finland, Denmark and Sweden, registries have been maintained over several centuries; the registries record births, deaths, marriages and migrations of individuals, and are of enormous value for the reconstruction of large pedigrees. The Nordic countries have also established cohorts for specific groups, such as twins. These cohorts contain an enormous amount of phenotypic information systematically collected over decades by well-trained epidemiologists. The free, high-quality health-care systems in Scandinavian countries result in high response rates (over 75%) to questionnaires, eager participation in epidemiological studies and a favourable attitude towards genetic research. The state-funded health-care systems also produce standardized, high-quality patient histories that contain valuable longitudinal information about disease processes. The most extensive example of the systematic use of population history, genealogical records and nationwide health-care registries in genetic research is the [deCode project](#) in Iceland (BOX 3).

Simple diseases versus complex diseases
Population isolates have been used with great success in identifying single-gene defects caused by rare alleles. The qualitative phenotypes typically encountered can be diagnosed reliably and unanimously by skilled clinicians. One of the standard Mendelian patterns of inheritance (autosomal recessive, autosomal dominant or X-linked) applies and penetrance is usually complete. Any disease having these features is an ideal target for linkage analysis, whether it is conducted in an isolate or in a large outbred population. However, the haplotype signatures found in disease chromosomes of affected individuals of a population isolate offer a shortcut for disease gene mapping (BOX 4). Monitoring just a few distantly related affected individuals can sometimes greatly narrow the chromosomal interval containing the disease locus^{16,17}. Whether isolates will provide a similar shortcut for complex disease genes remains to be seen. The outcome will depend strongly on the number and character of these genes.

Complex traits do not conform to Mendelian transmission patterns and are less clear-cut diagnostically. Only rare pedigrees have several affected individuals. Multiple genes of varying influence presumably influence most complex traits, but two extreme models can help focus our thinking. For example, we can imagine that common alleles at a handful of loci interact to cause a disease. Alternatively, rare alleles at numerous loci can each single-handedly cause the disease. The interaction model and the genetic heterogeneity model are not the only plausible hypotheses, but it is clear that population isolates are likely to be more valuable in the latter model. Isolates offer an opportunity to limit the number of rare disease alleles.

In practice, most successes in mapping complex disease loci in population isolates have depended on large pedigrees with proven or predicted genealogical ties between affected individuals. Linkage analysis with such

Box 3 | The Icelandic experiment



(Bartholomew Press Ltd. Reproduced with permission from Harper Collins Publishers.)

The deCODE project, founded in 1996, uses the genetically isolated population of Iceland and, perhaps more importantly, the extensive Icelandic genealogical database. Patients with a particular disease are clustered into large extended families, sometimes comprising hundreds of individuals over more than ten generations.

Iceland was founded in the ninth and tenth centuries by a limited number of settlers from Norway, Ireland and Scotland. Because the country has experienced little immigration over the past eleven centuries, most of the 275,000 living Icelanders are descendants of these original settlers. Y-chromosomal haplotypes suggest that 20–25% of Icelandic founding males were of Gaelic origin and 75–80% of Scandinavian origin⁶⁷. A tradition of recording family trees has left a truly unique resource for tracking the heredity of many diseases over hundreds of years. The genealogy of most Icelanders can be traced back over 1,000 years.

On 22 January 2000, deCODE received a 12-year licence from the Icelandic Ministry of Health to build and operate an Icelandic Health Sector Database (IHD) incorporating anonymous patient records from Iceland's national health-care system. IHD contains information on most of Iceland's population; only 7% of the population has opted out. The licence also permits deCODE to cross-reference the IHD data with the genealogical registers and genotyping data obtained from sampled Icelanders. To capitalize on this, deCODE has established a high-throughput genotyping enterprise in Reykjavik, which is rapidly producing large amounts of genetic information from sampled individuals. In addition to its value in mapping of complex disease genes, this cohort should facilitate studies of gene–environment interactions, correlation of genetic background with disease progression and variability in treatment response. Iceland's action in providing exclusive access to health-care information to a single private company has generated considerable debate, both in Iceland itself and in the wider genetics community. A special target of criticism has been the presumption that citizens consent to inclusion in the study unless they actively opt out.

unusual pedigrees has exposed rare disease genes for **nonpolyptic colon cancer** and familial **combined hyperlipidaemia** in Finnish families, for Hirschsprung disease in Mennonites, and for nonsyndromic hearing loss in Bedouins^{1–3,18,19}. Some of the complex disease loci identified in various population isolates are summarized in TABLE 1. (For an extended version of this Table, see **supplementary information** online.) Genome scans based on strategies to monitor intrafamilial association and linkage disequilibrium in population isolates have so far been less successful and/or less frequently applied than linkage analysis.

Unfortunately, many complex diseases are simple affected-versus-normal dichotomies at the clinical level. Rigorous quantification of intermediate disease states or disease liability is desirable because it increases the information available for mapping studies. Some diseases, such as lipid disorders, offer better opportunities for quantification than other diseases, such as psychiatric disorders. Nonetheless, most attempts to refine disease phenotypes will be well worth the effort. Short of

full quantification, subdivision of patient populations by qualitative clinical criteria can be extremely helpful in minimizing genetic heterogeneity. The benefits of good statistical treatment and careful dissection of disease phenotypes can be seen in the identification of type II diabetes loci on chromosomes 1q and 11q in Pima Indians. The significance for the 11q linkage increased when disease phenotypes were adjusted for obesity²⁰.

Haplotype mapping is often used as a follow-up to linkage analysis for locus restriction in isolated populations, but it is more successful with monogenic recessive diseases than it is with monogenic dominant diseases. An affected person almost always possesses a chromosome segment surrounding the disease locus that is identical by descent to similar segments in other affected individuals in the population. This makes it trivial to construct the common ancestral haplotype. For a dominant disease, the situation can be quite different. Any dominant disease mutation starts a clan of affected carriers. If the dominant disease is subject to strong selection, then the clan goes extinct in a few generations. Subsequent affected individuals arise from new mutations. This leaves little opportunity for the rearrangement of localized haplotype signatures by recombination. For example, in Finland, **Marfan syndrome** is caused by distinct mutations showing different haplotype signatures. For dominant diseases with negligible effects on fitness, such as **Huntington disease** and breast cancer associated with mutations in **BRCA1** and **BRCA2**, haplotype mapping has a much better chance of success. Fortunately, complex diseases tend to afflict the old and so have limited effects on fitness. Therefore, under the genetic heterogeneity hypothesis for such a disease, many large pedigrees with dominant transmission will exist. In population isolates, such pedigrees can be identified and are more likely to be affected by a single mutation. If multiple disease genes are segregating in an isolate, then regional subisolates can be investigated for disease clusters.

Of course, this is the most optimistic hypothesis. When multiple disease genes segregate, it is harder to discern a single haplotype signature, and it becomes necessary to resort to statistical measures of association or TRANSMISSION DISTORTION. If a disease allele is old, its haplotype signature will be short. For this reason, haplotype signatures tend to be shorter in older outbred populations than in young population isolates. Short haplotype signatures can be compensated for by increasing the density of the marker map. Under the genetic heterogeneity model, the number of involved alleles will be large. Under the genetic interaction model, the involved alleles are apt to be old. With intermediate models, we potentially face both problems.

The creation of the **Single Nucleotide Polymorphism (SNP) Consortium** and the discovery of tens of thousands of intragenic SNPs have provided a tremendous boost to complex disease genetics²¹. These markers provide a set of gene-specific tags that can be typed by robots and scored more reliably than tandem-repeat markers. The density of SNPs needed for linkage disequilibrium mapping is a matter of debate. In outbred populations, the most pessimistic view is that one SNP

TRANSMISSION DISTORTION
Over or under transmission
of certain alleles to
affected individuals.

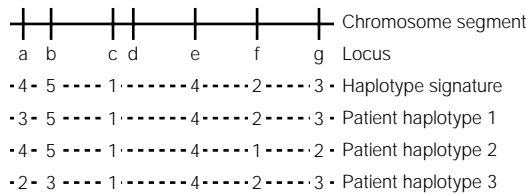
Box 4 | Haplotype mapping

Haplotype mapping is often carried out as part of a genome scan. In a population isolate, the appearance of a rare Mendelian disease is almost always attributable to a single founder gene or mutation.

The disease allele can be identified by searching for a common haplotype signature shared among patients. As the ancestral haplotype signature is passed from generation to generation, it is disrupted by recombination. Partial conservation of the haplotype signature in a patient strongly suggests that the disease locus resides in the conserved region of the haplotype.

As an example of haplotype mapping, the figure shows a chromosome segment that bears a disease locus (designated d) and six marker loci (designated a, b, c, e, f and g). Each of the four haplotypes shown is a sequence of numbered alleles at these six markers. The disease allele enters the population on the background of the ancestral haplotype signature. The three patient haplotypes allow us to reconstruct the signature and map the disease locus d. Patient haplotype 1 has experienced recombination between markers a and b, patient haplotype 2 has experienced recombination between markers e and f, and patient haplotype 3 has experienced recombination between markers b and c. The smallest region of overlap lies between markers b and f and consequently contains the disease locus d.

The smallest region of overlap inferred from multiple, partially conserved haplotypes is often much smaller than the smallest region of overlap inferred from linkage mapping of contemporary pedigrees. In population isolates, haplotype signatures can extend over several centiMorgans¹⁴. For some rare diseases, haplotype mapping has been successful using only four or five patients and a sparse genome scan of only 200–300 markers^{16,17}.



in every 3 kb will be necessary²², but more optimistic figures, such as one SNP in every 100 kb^{21,23}, have also been proposed. The current evidence of great variation of linkage disequilibrium across the human genome needs to be extended^{6,24}. Whatever the outcome of the genome-wide search for linkage disequilibrium, it is clear that SNPs will vastly improve ASSOCIATION TESTING and the resolution of haplotype mapping. But there is an important hurdle to overcome — harvesting the full potential of SNPs requires a more efficient, more flexible and cheaper method of SNP genotyping.

As part of the debate surrounding SNPs, recent research has indicated that linkage disequilibrium in random chromosomes of population isolates such as Finland (about 2,000 years old) and Sardinia (about 10,000 years old) may only be slightly elevated relative to neighbouring outbred populations^{25,26}. In contrast, Mohlke²⁷ found the marker-to-marker linkage disequilibrium reached over 1 cM in Finns. Because these studies have focused on non-disease alleles and common marker polymorphisms that probably predate the formation of isolates, they are not necessarily relevant to disease alleles. This uncertainty raises once again the issue of whether the genetic interaction or the genetic heterogeneity model is more pertinent to complex disease. It would be interesting to determine the overall level of linkage disequilibrium in the vicinity of the complex disease genes highlighted in TABLE 1, both within the relevant population isolate and near-by outbred populations.

Statistical strategies in isolated populations. Regardless of whether they are applied to population isolates or large outbred populations, statistical methods for gene mapping fall into several well-defined categories. For the coarse mapping of Mendelian traits, linkage analysis has proved remarkably successful¹⁵. Linkage analysis evaluates the recombination events occurring within a sample of pedigrees under a precise model of disease inheritance. The evidence in favour of linkage is summarized by the lod (logarithm of the odds) score plotted against the recombination fraction separating the disease locus and a nearby marker locus. The position where this curve peaks is usually taken as an estimate of the recombination fraction between the two loci.

With common diseases, the situation is less favourable to linkage analysis. The required genetic model at the trait locus is almost always unclear, and genetic heterogeneity quickly erodes lod scores²⁸. Some statistical geneticists recommend conducting linkage analysis assuming a monogenic disease model with reduced penetrance²⁹. The extremes of dominant and recessive inheritance are thought to cover most practical cases. More intellectually satisfying are nonparametric allele-sharing methods; these avoid explicit assumptions about disease transmission^{30,31}. If the affected people within a collection of pedigrees or sibships show an unusual propensity to share marker alleles in a particular chromosomal region, then that region is likely to harbour a disease gene. The power of parametric versus nonparametric methods depends strongly on how the disease in question is actually inherited. Although simulation studies indicate that parametric methods may possess greater power than nonparametric methods, the space of possible disease models is enormous and little explored by simulations³². This question is intertwined with the debate over population isolates because isolates limit genetic heterogeneity.

Association studies can be divided into case-control studies and familial tests of transmission distortion. In the case of an anonymous marker, association is the byproduct of linkage disequilibrium. By contrast, the alleles at a candidate locus may directly influence the risk of disease. In a genetically homogeneous population, case-control studies are the method of choice and involve straightforward comparison of allele or haplotype frequencies in random samples of cases and controls³³. When POPULATION STRATIFICATION is an issue, as it usually is in the United States, association studies rely heavily on the transmission-disequilibrium test (TDT)^{34,35}. The simplest version of the TDT uses a CONTINGENCY TABLE with columns corresponding to alleles and rows corresponding to parental genes transmitted or not transmitted to affected children. Whether one or more alleles are preferentially transmitted to the affected children is evaluated given the marker genotypes of the parents.

Although the TDT was originally recommended for independent trios of both parents and an affected child, versions of the TDT exist for sibships and even pedigrees^{36–40}. With parametric versions, such as the GAMETE COMPETITION MODEL, it should be possible to combine case-control data and familial transmission data into a

ASSOCIATION TESTING

A statistical approach that tests for association between marker or candidate gene alleles and diseases.

POPULATION STRATIFICATION

Subdivision of a population into different ethnic groups with potentially different marker allele frequencies and different disease prevalences.

CONTINGENCY TABLE

A table or matrix to count the numbers of observations falling into various categories. Each category is classified on the basis of several factors.

GAMETE COMPETITION MODEL

A statistical model that views transmission of marker alleles to affected children as a contest between the alleles. Each allele is ranked much as competing teams are ranked in a sports league.

Table 1 | Genome scans in isolated populations

Population	Age of population	Reported genome scans	Study sample	Loci showing linkage	Evidence from other populations (gene/locus identified)
Amish*	About 250 years	Bipolar disease	1 extended pedigree, 207 individuals	Chr 6,13, 15 (REF. 68)	Chr 13 (REF. 69)
Hutterites*	About 100 years	Allergic asthma	653 individuals	NS, suggestive: Chr 1, 3p, 5q, 13q (REFS 70,71)	Chr 1, 5q (REF. 72)
Mennonite*	About 200 years	Hirschsprung disease	1 family for linkage, 28 families for association	Chr 13q22 (REF. 1)	Endothelin B receptor ²
Pima Indians*	>10,000 years	Diabetes and body mass index (BMI)	264 nuclear families, 966 siblings	BMI and diabetes Chr 11 q Diabetes Chr 1q (REF. 73)	Chr 11q (NS), 1q (REFS 74,75)
Bedouins [†]	200 years	Nonsyndromic deafness	1 extended family, 55 individuals	Chr 13q12 (REF. 76)	Connexin 26 (REFS 3,77,78)
Finland, late settlement [‡]	330 years	Schizophrenia	21 families, 233 individuals	Chr 1q (REF. 79)	Chr 1q 21–22 (REF. 80)
Finland, old settlement [‡]	2,000 years	Multiple sclerosis	21 families, 191 individuals	Chr 6p, 17q22 (REFS 54,81)	Chr 6p, 17q (REFS 55,82)
Whole Finland [‡]		Familial combined hyperlipidemia	35 families, 168 individuals (stage 2: 206 individuals)	Chr 10p 1q 21 (REF. 58)	Chr 1q21, locus also in mouse syntenic region (REFS 83,84)
Iceland [§]	About 1,000 years	Schizophrenia	5 families, 91 individuals	Chr 6p Significant after adding samples from other populations ⁸⁵	Chr 6p (REFS 86,87)
Northern Sweden [§]	350 years	Familial prostate cancer	28 families, 366 individuals	Chr 1q 21 (REF. 88)	Chr 1q21 (REFS 89,90)

Every effort has been taken to select those loci that provide most evidence for linkage. However, the definition for significance is ambiguous. A lod (logarithm of odds) value >3.2 for parametric analyses and a $p < 0.001$ for nonparametric analyses has been considered significant. In most of these publications, values have not been corrected for multiple testing, adding a further level of uncertainty. *North American populations. †Populations from the Middle East and North Africa. The locus was originally detected in a consanguineous Tunisian pedigree. ‡Northern European populations. (Chr, chromosome; NS, nonsignificant.) For an extended version of this Table, see supplementary information online.

single analysis that checks both differences in allele frequencies and distorted transmission to children⁴¹.

Just as with linkage mapping, population isolates have certain advantages over outbred populations in association studies. The uniform genetic and environmental background of a population isolate makes case-control studies less suspect. Population isolates and outbred populations depart from genetic equilibrium in different ways. Hardy–Weinberg equilibrium, which is more likely in a population isolate than in a population with considerable racial admixture, helps in testing differences in allele frequencies between cases and controls. Linkage disequilibrium with rare disease alleles extends over greater distances in young population isolates and enables one to use fewer markers in an association study.

Haplotype mapping has the advantage of revealing ancient recombination events in addition to the contemporary recombination events detectable by linkage mapping. Linkage mapping operates best on a coarse scale where haplotype signatures are lost. So, it is sensible to view linkage and haplotype mapping as complementary techniques. The usual strategy is to use linkage mapping to identify a candidate region and then to follow up with haplotype mapping to pinpoint the disease locus within that region^{10,13}. If a good candidate gene is suspected in advance, it is possible to reverse the standard procedure and test first for association between the disease and alleles at the candidate gene or

nearby markers. A positive association can then be confirmed by linkage analysis. In practice, most geneticists hedge their candidate gene bets by including these genes as markers in a genome scan.

Experience and common sense dictate that any positive mapping claims must be replicated. Such replication can be done in either outbred populations or different population isolates. For common diseases, outbred populations clearly contain the bulk of affected people. Genes originally mapped or confirmed in an outbred population are likely to be pertinent to more of humanity. Mapping genes within multiple populations not only holds the promise of confirmation but also the chance of further restricting the mapped interval that contains the gene. It is worth stressing here the advantages of older outbred populations, with their shorter range of linkage disequilibrium.

There is great interest in quantifying the extent of linkage disequilibrium around a disease gene and using this as a guide to positional cloning. Inference from any single marker tends to be unreliable. Because of differences in levels of polymorphism and the chance association of disease mutations with marker alleles, single-marker measures of disequilibrium do not consistently increase or decrease as the location of a disease gene is approached. This fact has prompted the development of multimarker measures of disequilibrium and model-based methods of analy-

sis that exploit these measures^{42–50}. The jury is still out on how well these methods work in practice, but it is clear that they are intended for outbred populations, as well as isolates.

Rare genes for common diseases

Rare genes causing complex diseases should be treasured, not dismissed as epidemiologically irrelevant. Such genes provide wedges of understanding to crack open whole metabolic pathways and uncover new candidate genes for further genetic study. For instance, the discovery of the non-polytopic colon cancer gene has led to a new molecular mechanism for malignancy^{51,52}.

Mapping Mendelian forms of complex diseases helps clinicians sharpen diagnostic categories. Medical genetics has repeatedly advanced by splitting what was previously grouped together. Before genes were discovered in large pedigrees for **Alzheimer disease**, colon cancer and Hirschsprung disease, these were considered 'complex traits'. Only in retrospect were the mapped genes transformed into 'rare Mendelian forms of complex diseases'. Complex traits are complex because they represent umbrella diagnoses — similar end-state phenotypes shaped by a combination of the environment and a diverse set of major genes, each with a significant impact. Every rare Mendelian form of a complex disease mapped in a population isolate gives more insight into genetics and disease pathogenesis. This clarifying effect is illustrated by single-gene diseases such as **Bardet-Biedl syndrome**, for which different loci were identified in unrelated Bedouin tribes⁵³. The rewards of a divide-and-conquer strategy will be even greater in diseases such as epilepsy, cancer and cardiovascular disorders.

Even in the presence of large pedigrees and careful diagnostic splitting, geneticists will have to contend with the segregation of multiple disease genes in most data sets. This problem must be faced even in population isolates. The same disease genes present in an isolate will most probably also be present in larger outbred populations. For example, linkage analyses of 21 multiple sclerosis families (MS) from a late settlement subisolate of Finland indicated four possible loci involved in disease pathogenesis⁵⁴. Consistent with earlier studies, the locus of greatest significance was identified in the MHC region on chromosome 6. Furthermore, the Finnish findings of other MS-predisposing loci on chromosomes 5p, 17q and 18q are also supported by reports from other populations⁵⁵. Another interesting example is familial combined hyperlipidaemia. In Finnish families, a major locus on chromosome 1q has been identified in a region orthologous to a mouse locus causing hypertriglyceridaemia^{19,56}. This finding has also been replicated in other populations⁵⁷. Again, the same set of families also provided evidence for further predisposing loci⁵⁸.

The multiple sclerosis and hyperlipidaemia loci noted above probably represent major genes that explain complex phenotypes in special families from a special population. In other populations, the same loci most probably also exist, either as major loci or minor loci modifying the effects of major loci. Further genetic study and more exhaustive statistical analysis are war-

ranted. The disease examples above emphasize the need for extensive, worldwide collaboration and pooled data analyses. They also call for statistical strategies that are flexible enough to handle multivariate traits, correlated environments, gene–environment interaction, gene–gene interaction (epistasis) and pleiotropy. Variance component methods seem to be the only statistical tools that can deal with such complexity. These methods have proved their worth in mapping quantitative trait loci (QTL) and can be applied to many complex human diseases^{20,59–64}.

Although the focus of this review has been on human studies, it is clear that knowledge gained from other species will have a profound impact on medical genetic studies of complex traits. Inbred strains of mice have short lifespans and permit exquisite control of the genetic and environmental backgrounds of most traits. If experimental crosses are used to map complex traits in mice or other species, then conserved regions and homologous genes can be checked in humans. This mouse–human crosstalk strategy has led to the search for human disease loci in regions showing conserved synteny between man and mouse. Besides providing candidate regions and candidate genes to be tested in human study samples, experimental opportunities in other species will be essential to demonstrate the functional consequences of identified disease-associated gene variants. The greatest drawback of experimental species is the lack of appropriate animal models for some common human diseases, particularly for psychiatric disorders and behavioural traits. Research findings from plants and other model systems will also contribute to our understanding of human biology. The recent triumph in cloning the QTL gene determining fruit size in tomatoes provides a model for what can be done with quantitative traits⁶⁵.

Conclusions

Population isolates with stable cultures, universal health-care and demographic registries provide the closest human approximations to the ideal research conditions possible with inbred animal strains. Families or cases collected from population isolates show fewer environmental differences and less genetic heterogeneity than those collected from outbred populations. Good genealogical records, standardized health care, and reliable medical information all make it easier to assemble study samples, identify cases with shared ancestors and quantify phenotypes. However, not all isolates are alike. Differences in demographic history and the characteristic genetic diseases of isolated populations necessitate different research strategies incorporating optimized marker densities for genome scans and the most appropriate statistical tools for analysis. Whether or not the genetic features of population isolates turn out to be the key to unlocking the genetics of complex disease, the centralized registries and health-care records available in some isolated populations create an infrastructure for epidemiological and genetic research that is hard to match in more diverse countries, such as the United States.

UNASCERTAINED SAMPLE
A sample selected without
regard to disease status.

More effort should be put into constructing unascertained population cohorts for the longitudinal study of normal variation in genetic isolates. The continuing genetic studies in Iceland already combine many of the strengths of the ideal human study sample noted in BOX 1. Other largely untapped resources are the population-wide twin cohorts established 30 to 40 years ago in Scandinavian countries, which contain longitudinal data on life style, environment and medical history. Altogether, the registries represent an UNASCERTAINED SAMPLE of over 60,000 sib pairs from a population of 18 million people. Two-thirds of these twins are dizygotic, and in contrast to ordinary siblings, they share prenatal and early postnatal life with their twins. By using population registers, many twin pairs can be linked to large families in subisolates. Although information collected in the twin cohorts does not cover all complex traits, it represents a significant

resource for understanding complex diseases and for disentangling the combined influences of genes and environment in the pathogenesis of these diseases. If these cohorts are used in a manner acceptable by participants and communities, they would be a great asset in genetic epidemiological studies of complex traits.

Links

DATABASE LINKS [Hirschsprung disease](#) | [nonsyndromic hearing loss](#) | [nonpolytopic colon cancer](#) | [combined hyperlipidaemia](#) | [Marfan syndrome](#) | [Huntington disease](#) | [BRCA1](#) | [BRCA2](#) | [Bardet-Biedl syndrome](#) | [Alzheimer disease](#)

FURTHER INFORMATION [Human Genome Project](#) | [deCODE project](#) | [The SNP consortium](#)

- Puffenberger, E. G. *et al.* Identity-by-descent and association mapping of a recessive gene for Hirschsprung disease on human chromosome 13q22. *Hum. Mol. Genet.* **3**, 1217–1225 (1994).
- Puffenberger, E. G. *et al.* A missense mutation of the Endothelin-B receptor gene in multigenic Hirschsprung's disease. *Cell* **79**, 1257–1266 (1994).
- Scott D. A. *et al.* Nonsyndromic autosomal recessive deafness is linked to the DFNB1 locus in a large inbred Bedouin family from Israel. *Am. J. Hum. Genet.* **57**, 965–968 (1995).
- Bollobas, B. *Littlewood's Miscellany* (Cambridge Univ. Press, Cambridge, 1988).
- Cavalli-Sforza, L. The DNA revolution in population genetics. *Trends Genet.* **14**, 60–65 (1998).
- Kidd, J. R. *et al.* Haplotypes and linkage disequilibrium at the phenylalanine hydroxylase locus, PAH, in a global, MH representation of populations. *Am. J. Hum. Genet.* **66**, 1882–1899 (2000).
- Sajantila, A. *et al.* Paternal and maternal linkages reveal a bottleneck in the founding of the Finnish population. *Proc. Natl Acad. Sci. USA* **93**, 12035–12039 (1996).
- Kittles, R. A. *et al.* Dual origins of Finns revealed by Y chromosome haplotype variation. *Am. J. Hum. Genet.* **62**, 1171–1179 (1998).
- de la Chapelle, A & Wright F. A. Linkage disequilibrium mapping in isolate populations: The example of Finland revisited. *Proc. Natl Acad. Sci. USA* **95**, 12416–12423 (1998).
- Jorde, L. B. *et al.* Gene mapping in isolated populations: New roles for old friends? *Hum. Hered.* **50**, 57–65 (2000).
- Varilo, T. *et al.* Linkage disequilibrium in isolated populations: Finland and a young sub-population of Kuusamo. *Eur. J. Hum. Genet.* (in the press).
- Terwilliger, J. D., Zöllner, S., Laan, M. & Pääbo, S. Mapping in small populations with no demographic expansion. *Hum. Hered.* **48**, 138–154 (1998).
- Peltonen, L. Positional cloning of disease genes: Advantages of genetic isolates. *Hum. Hered.* **50**, 66–75 (2000).
- Peltonen, L., Jalanko, A. & Varilo, T. Molecular genetics of the Finnish disease heritage. *Hum. Mol. Genet.* **8**, 1913–1923 (1999).
- Wright, A. F., Carothers, A. D. & Pirastu, M. Population choice in mapping genes for complex diseases. *Nature Genet.* **23**, 397–404 (1999).
- Considers in some detail the advantages and disadvantages of various populations in mapping common disease genes.
- Houwen, R. H. *et al.* Genome screening by searching for shared segments: Mapping a gene for benign recurrent intrahepatic cholestasis. *Nature Genet.* **8**, 380–386 (1994).
- Shows the power of the founder effect and population isolation in the search for a disease gene. Scanning for a haplotype signature in affected individuals can be used to localize the gene.
- Nikali, K. *et al.* Random search for shared chromosomal regions: the assignment of a new hereditary ataxia locus. *Am. J. Hum. Genet.* **56**, 1088–1095 (1995).
- Peltomaki, P. *et al.* Genetic mapping of a locus predisposing to human colorectal cancer. *Science* **260**, 810–812 (1993).
- Pajukanta, P. *et al.* Linkage of familial combined hyperlipidaemia to chromosome 1q21–q23. *Nature Genet.* **18**, 369–373 (1998).
- Hanson, R. L. *et al.* An autosomal genomic scan for loci linked to type II diabetes mellitus and body-mass index in Pima Indians. *Am. J. Hum. Genet.* **63**, 1130–1138 (1998).
- Collins, F. S., Guyer, M. S. & Charkravarti, A. Variations on a theme: cataloguing human DNA sequence variation. *Science* **278**, 1580–1581 (1997).
- Lays out the rationale for the discovery and application of single nucleotide polymorphisms.
- Kruglyak, L. Prospect for whole-genome linkage disequilibrium mapping of common disease genes. *Nature Genet.* **22**, 139–144 (1999).
- Zwick, M. E. *et al.* Characterizing human genomic variation and linkage disequilibrium in multiple 100kb genomic segments using large-scale, microarray-based SNP detection. *Am. J. Hum. Genet.* **67**, S22 (2000).
- Nickerson, D. A. *et al.* DNA sequence diversity in a 9.7-kb region of the human lipoprotein lipase gene. *Nature Genet.* **19**, 233–240 (1998).
- Explores one human gene intensely for polymorphisms and linkage disequilibrium. This type of study is a forerunner of what will be done on a genome-wide basis.
- Eaves, I. A. *et al.* The genetically isolated populations of Finland and Sardinia may not be a panacea for linkage disequilibrium mapping of common disease genes. *Nature Genet.* **25**, 320–323 (2000).
- Taillon-Miller, P. *et al.* Juxtaposed regions of extensive and minimal linkage disequilibrium in human Xq25 and Xq28. *Nature Genet.* **25**, 324–328 (2000).
- Moehlke, K. L. *et al.* Marker-marker linkage disequilibrium extends beyond 1 cM on chromosome 20 in Finns. *Am. J. Hum. Genet.* **67**, S25 (2000).
- Greenberg, D. A., Abreu, P. & Hodge, S. E. The power to detect linkage in complex disease by means of simple LOD-score analyses. *Am. J. Hum. Genet.* **63**, 870–879 (1998).
- Clerget-Darpoux, F., Bonaiti-Pellie, C. & Hochez, J. Effects of misspecifying genetic parameters in lod score analysis. *Biometrics* **42**, 393–399 (1986).
- Kruglyak, L., Daly, M. J., Reeve-Daly, M. P. & Lander, E. S. Parametric and nonparametric linkage analysis: A unified multipoint approach. *Am. J. Hum. Genet.* **58**, 1347–1363 (1996).
- Sobel, E. & Lange, K. Descent graphs in pedigree analysis: applications to haplotyping, location scores, and marker sharing statistics. *Am. J. Hum. Genet.* **58**, 1323–1337 (1996).
- Abreu, P. C., Greenberg, D. A. & Hodge, S. E. Direct power comparisons between simple lod scores and NPL scores for linkage analysis in complex diseases. *Am. J. Hum. Genet.* **65**, 847–857 (1999).
- Risch, N. & Merikangas, K. The future of genetic studies of complex human diseases. *Science* **273**, 1516–1517 (1996).
- Argues the case for association studies over linkage studies.
- Terwilliger, J. D. & Ott, J. A haplotype-based 'haplotype relative risk' approach to detecting allelic associations. *Hum. Hered.* **42**, 337–346 (1992).
- Spielman, R. S., McGinnis, R. E. & Ewens, W. J. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am. J. Hum. Genet.* **52**, 506–516 (1993).
- Allison, D. B., Heo, M., Kaplan, N. & Martin, E. R. Sibling-based tests of linkage and association for quantitative traits. *Am. J. Hum. Genet.* **64**, 1754–1763 (1999).
- Boehnke, M. & Langefeld, C. D. Genetic association mapping based on discordant sib pairs: the discordant-alleles test. *Am. J. Hum. Genet.* **62**, 950–961 (1998).
- Spielman, R. S. & Ewens, W. J. A sibship test for linkage in the presence of association: the sib transmission/disequilibrium test. *Am. J. Hum. Genet.* **62**, 450–458 (1998).
- Schaid, D. J. & Rowland, C. Use of parents, sibs, and unrelated controls for detection of associations between genetic markers and disease. *Am. J. Hum. Genet.* **63**, 1492–1506 (1998).
- Teng, J. & Risch, N. The relative power of family-based and case-control designs for linkage disequilibrium studies of complex human diseases. II. Individual genotyping. *Genome Res.* **9**, 234–241 (1999).
- Sinsheimer, J. S., Blangero, J. & Lange, K. Gamete competition models. *Am. J. Hum. Genet.* **66**, 1168–1172 (2000).
- Collins, A. & Morton, N. E. Mapping a disease locus by allelic association. *Proc. Natl Acad. Sci. USA* **95**, 1741–1745 (1998).
- Devlin, B., Risch, N. & Roeder, K. Disequilibrium mapping: composite likelihood for pairwise disequilibrium. *Genomics* **36**, 1–16 (1996).
- Kaplan, N. L., Hill, W. G. & Weir, B. S. Likelihood methods for locating disease genes in non-equilibrium populations. *Am. J. Hum. Genet.* **56**, 18–32 (1995).
- Lazzeroni, L. C. & Lange, K. A conditional inference framework for extending the transmission/disequilibrium test. *Hum. Hered.* **48**, 67–81 (1998).
- McPeck, M. S. & Strahs, A. Assessment of linkage disequilibrium by the decay of haplotype sharing, with application to fine scale genetic mapping. *Am. J. Hum. Genet.* **65**, 858–875 (1999).
- Rannala, B. & Slatkin, M. Likelihood analysis of disequilibrium mapping, and related problems. *Am. J. Hum. Genet.* **62**, 459–473 (1998).
- Terwilliger, J. D. A powerful likelihood method for the analysis of linkage disequilibrium between trait loci and one or more polymorphic marker loci. *Am. J. Hum. Genet.* **56**, 777–787 (1995).
- Xiong, M. & Guo, S. W. Fine-scale genetic mapping based on linkage disequilibrium: theory and applications. *Am. J. Hum. Genet.* **60**, 1513–1531 (1997).

50. Göring, H. H. H. & Terwilliger, J. D. Linkage analysis in the presence of errors IV: Joint pseudomarker analysis of linkage and/or linkage disequilibrium on a mixture of pedigrees and singletons when the mode of inheritance cannot be accurately specified. *Am. J. Hum. Genet.* **66**, 1310–1327 (2000).
51. Leach, F. S. *et al.* Mutations of a *mutS* homolog in hereditary nonpolyposis colorectal cancer. *Cell* **75**, 1215–1225 (1993).
52. Bronner, C. E. *et al.* Mutation in the DNA mismatch pair gene homologue hMLH1 is associated with hereditary non-polyposis colon cancer. *Nature* **368**, 258–261 (1994).
53. Sheffield, V. C., Stone, E. M. & Carmi, R. Use of isolated inbred human populations for identification of disease genes. *Trends Genet.* **14**, 391–396 (1998).
54. Kuokkanen, S. *et al.* Genome wide scan of multiple sclerosis in Finnish multiplex families. *Am. J. Hum. Genet.* **61**, 1379–1387 (1997).
55. Chataway, J. *et al.* The genetics of multiple sclerosis: principles, background and updated results of the United Kingdom systematic genome screen. *Brain* **121**, 1869–1887 (1998).
56. Castellani, L. W. *et al.* Mapping a gene for combined hyperlipidemia in a mutant mouse strain. *Nature Genet.* **18**, 374–377 (1998).
57. Pei, W. *et al.* Support for linkage of familial combined hyperlipidemia to chromosome 1q21–q23 in Chinese and German families. *Clin. Genet.* **57**, 29–34 (2000).
58. Pajukanta, P. *et al.* Genomewide scan for familial combined hyperlipidemia genes in Finnish families, suggesting multiple susceptibility loci influencing triglyceride, cholesterol, and apolipoprotein B levels. *Am. J. Hum. Genet.* **64**, 1453–1463 (1999).
59. Almasy, L. & Blangero, J. Multipoint quantitative-trait linkage analysis in general pedigrees. *Am. J. Hum. Genet.* **62**, 1198–1211 (1998).
60. Amos, C. I. Robust variance-components approach for assessing genetic linkage in pedigrees. *Am. J. Hum. Genet.* **54**, 535–543 (1994).
61. Blangero, J. & Almasy, L. Multipoint oligenic linkage analysis of quantitative traits. *Genet. Epidemiol.* **14**, 959–964 (1997).
62. Goldar, D. E. Multipoint analysis of human quantitative genetic variation. *Am. J. Hum. Genet.* **47**, 957–967 (1990).
63. Hoppers, J. L. & Mathews, J. D. Extensions to multivariate normal models for pedigree analysis. *Ann. Hum. Genet.* **46**, 373–383 (1982).
64. Schork, N. J. Extended multipoint identity-by-descent analysis of human quantitative traits: efficiency, power, and modeling considerations. *Am. J. Hum. Genet.* **53**, 1306–1319 (1993).
65. Frary, A. *et al.* fw2.2: A quantitative trait locus key to the evolution of tomato fruit size. *Science* **289**, 85–88 (2000). **A landmark publication reporting the isolation of a quantitative trait locus gene by positional cloning.**
66. Varilo, T. *et al.* Linkage disequilibrium in isolated populations: Finland and a young subpopulation of Kuusamo. *Eur. J. Hum. Genet.* (in the press).
67. Helgason, A. *et al.* Estimating Scandinavian and Gaelic ancestry in the male settlers of Iceland. *Am. J. Hum. Genet.* **67**, 697–717 (2000).
68. Ginns, E. I. *et al.* A genome-wide search for chromosomal loci linked to bipolar affective disorder in the Old Order Amish. *Nature Genet.* **12**, 431–435 (1996).
69. Detera-Wadleigh, S. *et al.* A high-density genome scan detects evidence for a bipolar-disorder susceptibility locus on 13q32 and other potential loci on 1q32 and 18p11.2. *Proc. Natl Acad. Sci. USA* **96**, 5604–5609 (1999).
70. Ober, C. *et al.* Collaborative study on the genetics of asthma. Genome-wide search for asthma susceptibility loci in a founder population. *Hum. Mol. Genet.* **7**, 1393–1398 (1998).
71. Ober, C., Tsalenko, A., Parry, R. & Cox, N. J. A second-generation genomewide screen for asthma-susceptibility alleles in a founder population. *Am. J. Hum. Genet.* **67**, 1154–1162 (2000).
72. Marsh, D. G. *et al.* A genome-wide search for asthma susceptibility loci in ethnically diverse populations. *Nature Genet.* **15**, 389–392 (1997).
73. Hanson, R. L. *et al.* An autosomal genomic scan for loci linked to type II diabetes mellitus and body-mass index in Pima Indians. *Am. J. Hum. Genet.* **63**, 1130–1138 (1998). **Shows the power of the quantitative trait locus strategy and the quantitation of phenotypes for the identification of disease susceptibility loci.**
74. Elbein, S., Hoffman, M., Teng, K., Leppert, M. & Hasstect, S. A genome-wide search for type 2 diabetes susceptibility genes in Utah Caucasians. *Diabetes* **48**, 1175–1182 (1999).
75. Ghosh, S. *et al.* The Finland–United States investigation of non-insulin-dependent diabetes mellitus genetics (FUSION) study. I. An autosomal genome scan for genes that predispose to Type 2 diabetes. *Am. J. Hum. Genet.* **67**, 1174–1185 (2000).
76. Guilford, P. *et al.* A non-syndromic form of neurosensory, recessive deafness maps to the pericentromeric region of chromosome 13q. *Nature Genet.* **6**, 24–28 (1994).
77. Scott, D. A. *et al.* Nonsyndromic autosomal recessive deafness is linked to the DFNB1 locus in a large inbred Bedouin family from Israel. *Am. J. Hum. Genet.* **57**, 965–968 (1995).
78. Kelsell, D.P. *et al.* Connexin 26 mutations in hereditary nonsyndromic sensorineural deafness. *Nature* **387**, 80–83 (1997).
79. Hovatta, I. *et al.* A genomewide screen for schizophrenia genes in an isolated Finnish subpopulation suggesting multiple susceptibility loci. *Am. J. Hum. Genet.* **65**, 1114–1124 (1999).
80. Brzustowicz, L. M., Hodgkinson, K. A., Chow, E. W. C., Honer, W. G. & Bassett, A. S. Location of a major susceptibility locus for familial schizophrenia on chromosome 1q21–q22. *Science* **288**, 678–682 (2000).
81. Kuokkanen, S. *et al.* A putative vulnerability locus to multiple sclerosis maps to 5p14–p12 in a region syntenic to the murine locus *Eae2*. *Nature Genet.* **13**, 477–480 (1996).
82. Sawcer, S. *et al.* A genome screen in multiple sclerosis reveals susceptibility loci on chromosome 6p21 and 17q22. *Nature Genet.* **13**, 444–468 (1996).
83. Castellani, L. W. *et al.* Mapping a gene for combined hyperlipidemia in a mutant mouse strain. *Nature Genet.* **18**, 374–377 (1998).
84. Pei, W. *et al.* Support for linkage of familial combined hyperlipidemia to chromosome 1q21–q23 in Chinese and German families. *Clin. Genet.* **57**, 29–34 (2000).
85. Moises, H. W. *et al.* An international two-stage genome-wide search for schizophrenia susceptibility genes. *Nature Genet.* **11**, 321–324 (1995).
86. Straub, R. E. *et al.* A potential vulnerability locus for schizophrenia on chromosome 6p24–22: evidence for genetic heterogeneity. *Nature Genet.* **11**, 287–293 (1995).
87. Levinson, D. F. *et al.* Genome scan of schizophrenia. *Am. J. Psychiatry* **155**, 741–750 (1998).
88. Smith, J. R. *et al.* Major susceptibility locus for prostate cancer on chromosome 1 suggested by a genome wide search. *Science* **274**, 1371–1374 (1996).
89. Berthon, P. Predisposing gene for early-onset prostate cancer, localized on chromosome 1q42.2–43. *Am. J. Hum. Genet.* **62**, 1416–1424 (1998).
90. Xu, J. *et al.* Combined analysis of hereditary prostate cancer linkage to 1q24–25. Results from 772 hereditary prostate cancer families from the International Consortium for Prostate Cancer Genetics. *Am. J. Hum. Genet.* **66**, 945–957 (2000).

Acknowledgements

Our research has been supported by grants from the National Institutes of Health and an award to the Centre of Excellence in Disease Genetics by the Academy of Finland.