

The allelic structure of common disease

Desmond J. Smith^{1,2,*} and Aldons J. Lusis^{3,4,5}

¹Department of Molecular and Medical Pharmacology, ²Crump Institute for Molecular Imaging, ³Department of Microbiology, Immunology and Molecular Genetics, ⁴Department of Medicine and ⁵Department of Human Genetics, University of California, Los Angeles, CA 90095, USA

Received June 28, 2002; Accepted July 27, 2002

A better understanding of the allelic structure of common human disease loci may help identification of the responsible genes, and is thus a topic of considerable practical importance. If few alleles at each locus account for the majority of disease risk, then screening for these causative factors will be greatly simplified. In contrast, if large numbers of independent alleles are responsible, dramatic improvements in genotyping speed will be necessary, placing the dream of personalized medicine far in the future. In this review, the evidence for and against the optimistic and pessimistic viewpoints is discussed. It appears that neither position has been proved or disproved, but the available evidence indicates that common diseases are due at least in part to genes with a small number of disease-associated alleles.

One of the early dreams of the human genome project was the goal of personalized medicine. The vision was that each individual would have a compact disk containing their own unique DNA sequence, guiding myriad decisions on lifestyle, disease prevention, treatment and conceivably even career choice. At present, DNA sequencing technology is nowhere near cheap enough or fast enough to allow this dream. Nevertheless, genotyping technologies have made dramatic strides over the last decade, and these technologies may now be fast enough to allow interrogation of all relevant disease alleles in any given individual. The feasibility of this approach depends critically on the allelic structure of human disease genes. If all 30 000 genes make important contributions to disease susceptibility, and there are 1000 alleles for each gene, then a total of 3×10^7 genotypings will have to be reliably performed—a daunting prospect. However, if only 1000 genes make appreciable contributions to human disease, and each of these genes has only 4 relevant alleles, then the number of genotypings to be performed is only 4000—highly feasible. In this review, we look at the present knowledge on the allelic complexity of human disease genes, as well as providing a brief glance to the future.

PROPERTIES OF ALLELIC SPACE

A perhaps oversimplified aid to visualizing the properties of human disease genes is provided in Figure 1. This figure depicts a three-dimensional allelic space, with the axes of disease frequency, number of responsible loci and mean number of alleles at each locus. Regarding the second axis, number of

responsible loci, genetic traits are traditionally divided into monogenic (single gene or simple trait), polygenic (many genes), and multifactorial or complex (interactions between multiple genes and environment). Because a single locus is responsible for the monogenic disorders, identification of the relevant loci is usually straightforward, provided that a sufficient number of multiplex families are available. In contrast, the genetics of complex traits remains murky, and success in identifying the responsible loci has been elusive. This is because many genes and environment conspire together in a series of complicated epistatic and synergistic relationships to cause the final disease phenotype. In general, the number of loci identified through linkage for any given complex disease trait is less than 10, but linkage can be expected to identify only the major contributing loci. Since, with very few exceptions, the genes responsible for complex traits have yet to be identified, it remains an open question as to the number of alleles at each locus.

Two extreme models are usually envisaged for complex disease traits (Fig. 1) (1). One idea is that common alleles at a handful of loci interact to cause disease. This is referred to as the interaction model or the common disease/common variant (CD/CV) hypothesis. The competing notion is that rare alleles at numerous loci can each single-handedly cause the disease. This is referred to as the genetic heterogeneity model. Two other extreme models are also formally possible: complex disease traits are caused by a small number of loci each with a large number of alleles, or they are caused by a large number of loci with a large number of alleles (Fig. 1). One of these last two alternatives (a small number of loci each with a large number of alleles) is rendered unlikely by considerations of human population structure (see below).

*To whom correspondence should be addressed at: Pharmacology, UCLA School of Medicine, 23-120 CHS, Box 951735, Los Angeles, CA 90095-1735, USA. Tel: +1 3102060086; Fax: +1 3108256267. Email: dsmith@mednet.ucla.edu
Correspondence may also be addressed to A.J. Lusis at: Microbiology, Immunology and Molecular Genetics, UCLA, 3730 MRL, Box 951679, Los Angeles, CA 90095-1679, USA. Tel: +1 3108251359; Fax: +1 3107947345. Email: jlusis@mednet.ucla.edu

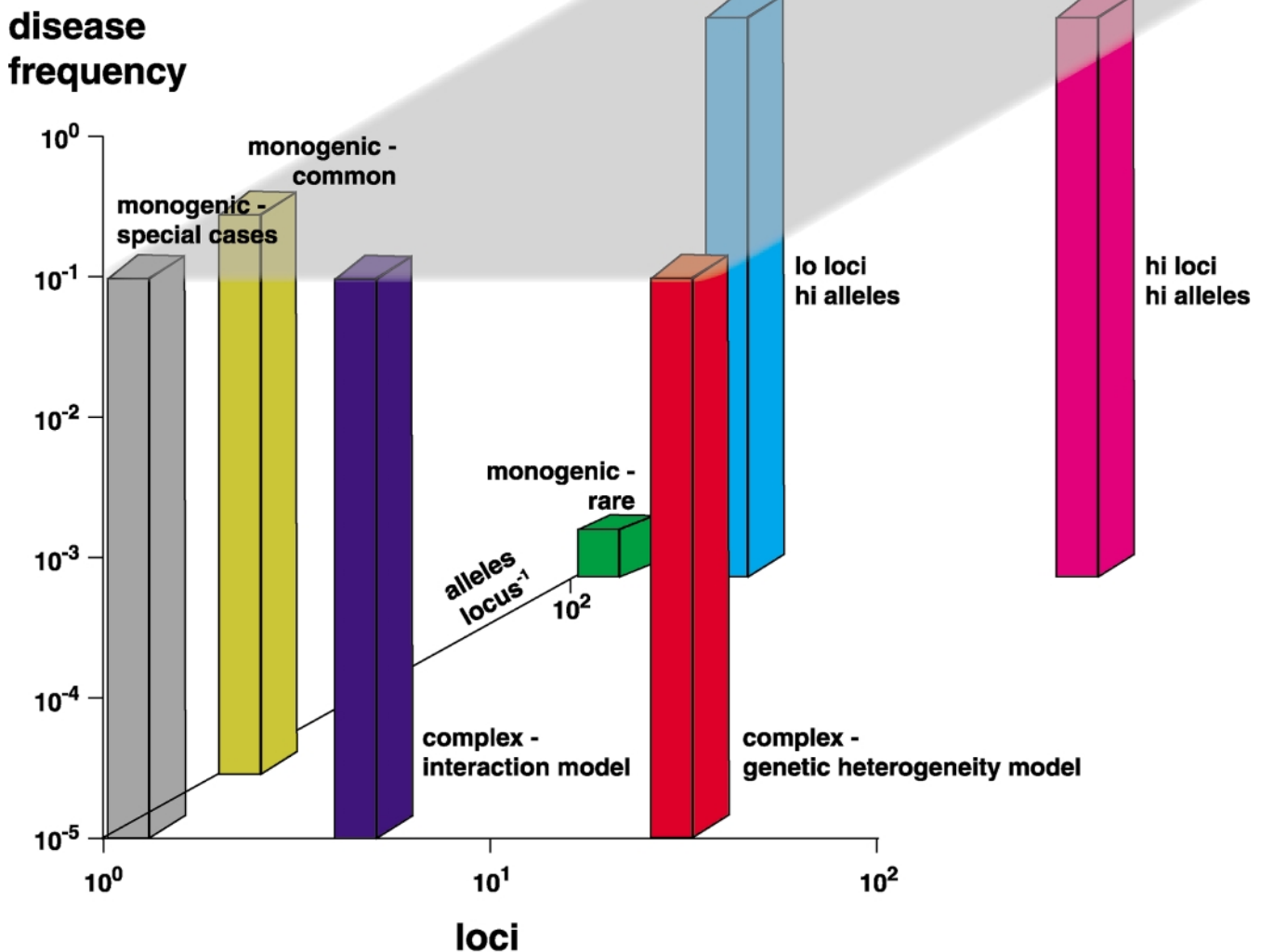


Figure 1. Allelic space. This three-dimensional space provides an aid to visualizing possible structures of polymorphism for human disease genes. For monogenic disorders, rare diseases (green bar) have complex allelic spectra (high number of alleles per locus), while common diseases (yellow bar) have simple spectra (low number of alleles per locus). Special cases (gray bar) exist in situations such as heterozygous advantage and population bottlenecks. For common complex disease traits, the situation is less clear. Two extreme models are the interaction model (CD/CV hypothesis, dark blue bar) and the genetic heterogeneity model (red bar). However, two other extreme models are also formally possible: a small number of loci each with a large number of alleles (cyan bar), and a large number of loci each with a large number of alleles (magenta bar). Considerations of human population genetics (2) suggest that one of the four extreme models, a small number of loci each with a large number of alleles (cyan bar), is unlikely. However, present data are insufficient to confidently place complex disease traits between the remaining three extremes, indicated by the two-dimensional gray surface. In this figure, the following examples are employed (2, 10): monogenic (rare)—retinoblastoma; monogenic (common)—glucose-6-phosphate dehydrogenase deficiency (South China); monogenic (special case)— β -thalassemia (Sardinia); complex (interaction model)—Alzheimer's disease (speculative); complex (genetic heterogeneity model)—colon cancer (speculative). No known examples yet exist for the two hypothetical cases of common complex diseases caused by a small number of loci each with a large number of alleles or a large number of loci each with a large number of alleles. For clarity, the disease frequencies for both the monogenic and common complex disorders have been rounded to 10^{-1} , but all are within one order of magnitude of the known frequency. In the cases of the complex diseases, Alzheimer's disease and colon cancer, the assigned values for the number of loci and the number of alleles per locus are highly speculative. For the purposes of the figure, it has been assumed that Alzheimer's disease is an example of the interaction model, whereas colon cancer is an example of the genetic heterogeneity model. Although both assumptions are conjectural, there is some evidence in favor of the genetic heterogeneity model for cancer (10).

However, between the remaining three extremes, complex disease traits could occupy any middle ground.

INFERENCES FROM HUMAN POPULATION STRUCTURE

Reich and Lander (2) have provided a simple framework for understanding the allelic spectra (number and frequency of

disease-predisposing alleles) of human disease genes. In addition, this work provides some justification for the CD/CV hypothesis.

There is a large amount of evidence suggesting a dramatic expansion of the human population $\sim 18\,000$ – $150\,000$ years or 700–6000 generations ago. This resulted in an increased effective population size from 10^4 individuals to the present size of 6×10^9 . Assume expansion of one allele of a rare

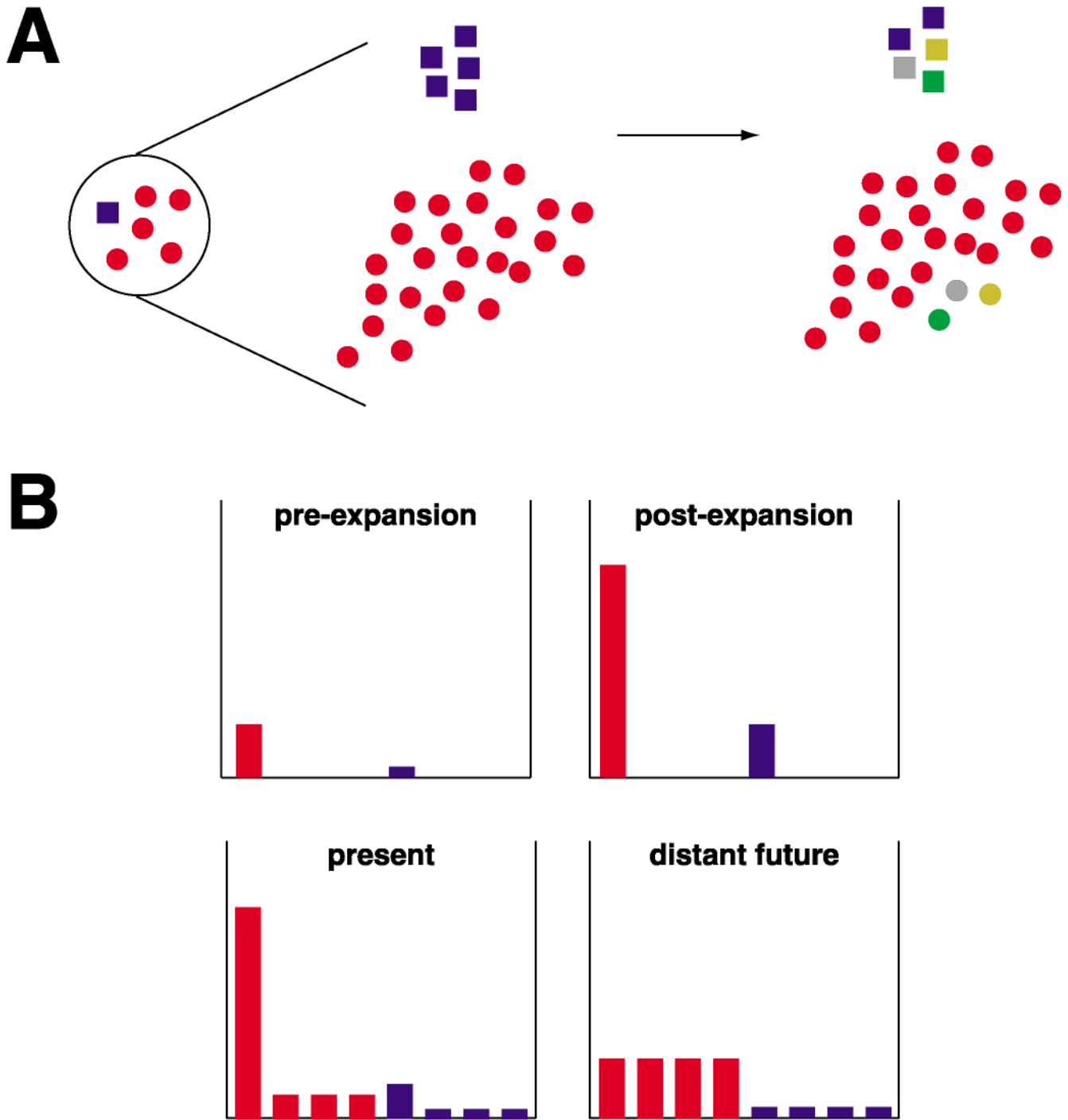


Figure 2. Human population structure. **(A)** A dramatic population expansion leads to increased numbers of individuals with a single allele of a rare (5 squares) and common (25 circles) disease gene. Because the rare disease gene is under strong negative selection, three alleles disappear in the population, and are replaced by three new alleles. Thus, only 2/5 (40%) of alleles represent the ancestral allele. In contrast, the common disease gene is under much less strong negative selection, and only one allele is lost to be replaced by three new alleles. The proportion of ancestral alleles is hence 24/27 (89%). Consequently, the common disease gene has a much simpler allelic structure in the present population than the rare gene. **(B)** Graphical depiction of **(A)**. The population expansion leads to proportionate increases of a single allele of a common disease gene (red) and a rare disease gene (blue). In the present human population, the rapid loss of alleles and smaller population reservoir of the rare disease gene has led to the allelic spectrum of this gene becoming relatively much more diverse than the common gene. However, in the far distant future (at equilibrium), the common disease gene can be expected to catch up with the rare gene, and the allelic spectra of both genes will be equally diverse.

disease gene, and one allele of a common gene (Fig. 2A). With some notable exceptions, the available evidence suggests fairly similar mutation rates for most genes, $\sim 10^{-6}$ – 10^{-5} per gene per generation (3–7). For the rare disease, the number of individuals in the expanded population is still small, and hence in relative terms they are quickly swamped by newly mutated alleles. In contrast, the population reservoir for the common disease gene is large, and it takes a long time for these alleles to be taken over by new mutations. In addition, in the face of similar mutation rates, the lower prevalence of rare diseases implies a higher allele disappearance rate (negative selection) for these diseases than for common diseases. Both of these effects (reservoir effect and selection pressure) imply that after the population expansion, allelic spectra became diverse much more quickly for the rare diseases than the common. In the long run, however, it is expected that the spectra for both rare and common diseases will become equally diverse (Fig. 2B). Reich and Lander estimate that an intermediate stage presently exists in the population, where rare diseases have very diverse allelic structure but the common disorders have simple structure. A prediction of this theory is that rare disorders will largely consist of recent alleles whereas common disorders will consist of old (ancestral) alleles.

ALLELIC STRUCTURE OF SIMPLE DISEASES

Since >1000 Mendelian (monogenic) disease genes have been identified, our understanding of their complexity is considerable. In general, rare diseases with strong negative selection exhibit very large allelic diversity (8). An exception to the pattern of high allelic diversity for rare disorders is when these alleles also provide heterozygous advantage. One example is provided by the thalassemias, which confer resistance to malaria. Once arisen, the strong positive selective pressure conferred by these single alleles ensures their relatively rapid spread through the population. Another exception to the high allelic diversity of rare disorders occurs if a population bottleneck results in greatly simplified spectra. For example, among Finnish and Ashkenazi Jewish populations, there are numerous examples of disease where individual alleles show greatly elevated frequencies compared with other populations. Overall then, for monogenic disorders there is good agreement between the theoretical predictions of Reich and Lander (2) and the available data.

ALLELIC STRUCTURE OF COMPLEX DISEASES

For common complex disorders, there are relatively few data on the diversity of allelic spectra due to the difficulty in identifying the responsible genes. The few known examples, however, tend to support the interaction model and the CD/CV hypothesis, since most of these genes have a small number of common, disease-associated alleles (Table 1). With a couple of possible exceptions, the genes/loci listed in Table 1 have been confirmed in multiple studies and are unlikely to represent type 1 errors. It should be noted, however, that the listed genes may not be fully representative of common disease genes, since genes with a small number of common alleles would be easier to identify by association than genes with a large number of rare alleles.

Another valid test of the CD/CV hypothesis is to examine the genetic complexity of quantitative traits related to common disease, such as blood pressure, body mass index and plasma lipids. One well-studied group of phenotypes comprises the levels and compositions of plasma lipoproteins, which are strongly associated with coronary heart disease. With the exception of mutations of the LDL receptor that underlie familial hypercholesterolemia (a relatively common monogenic disorder), most of the known variations associated with the traits are due to a small number of common alleles (Table 1).

A corollary of the CD/CV hypothesis is that each individual will carry susceptibility genes for numerous disorders. This point is difficult to examine systematically in humans, although, clearly, older individuals frequently suffer from a host of common maladies. A more convincing example is provided by surveying inbred mouse strains for differences in disease-related traits. Table 1 lists some common diseases and related traits differing between the two inbred strains DBA/2J and C57BL/6J. Many of these traits have been analyzed in genetic crosses between the two strains, and are known to be due to multiple quantitative trait loci (QTL). One might argue that, since the mice have been maintained by continuous brother–sister mating for many generations, numerous mutations have been fixed, and these are responsible for the trait differences. A counter argument to this, however, is the fact that a number of the loci observed in the DBA/2J \times C57BL/6J cross are also observed in crosses with other strains, and, in some cases, these loci are syntenic with human loci for the corresponding traits (9).

Apart from these examples, the rapid fall-off in disease rates in relatives of affected observed for most common diseases argues for the interaction model and against the heterogeneity model. In addition, many common diseases exhibit a high ratio of relative risk for monozygotic twins to that for dizygotic twins, also consistent with the genetics of the multifactorial threshold model (2,10). However, there are exceptions. For example, the common disease familial combined hyperlipidemia (prevalence 1–2%) exhibits an apparent dominant mode of inheritance through many generations, yet is clearly very complex (11). Furthermore, the inheritance of cancers is, in some ways, more consistent with a heterogeneity model (see below).

Other challenges to the CD/CV hypothesis, also in the area of cancer, are presented by the breast cancer genes *BRCA1* and *BRCA2* (12). These loci contribute to a relatively common genetic disorder, but have many different rare mutations, even in the Ashkenazi population (see above). In addition, even when the CD/CV hypothesis is true, disease genes may be dominated by a few alleles, but have large numbers of minor alleles. In the case of cystic fibrosis, 67% of disease alleles in the Caucasian population are accounted for by a single mutation. Nevertheless, the total number of known disease alleles exceeds 100 (8).

DO ALL COMMON DISEASES HAVE A SIMILAR ALLELIC STRUCTURE?

The general allelic architecture of various diseases or classes of genes may differ depending on the biological mechanisms

Table 1. Three lines of evidence supporting the common disease/common variant hypothesis

1. Common diseases with known genes or loci				
Disease	Gene		No. of common disease-associated alleles	Refs
Affective illness	Serotonin receptor		2	(14)
Alzheimer's disease	Apolipoprotein E		3	(8)
Asthma	5-Lipoxygenase		Several	(15,16)
Autoimmune disease	Major histocompatibility complex		Several	(8)
Deep vein thrombosis	Factor V, Leiden allele		2	(8)
Hemochromatosis	HLA-H		2	(2)
Inflammatory bowel disease	Chromosome 5q31 cytokine cluster		Unknown	(20)
Inflammatory bowel disease	NOD 2		4	(22,23)
Myocardial infarction	Paraoxonase-1		2	(24)
Narcolepsy	Major histocompatibility complex		Several	(21)
Type 1 diabetes	Insulin variable number of tandem repeats		Several	(25)
Type 1 diabetes	Major histocompatibility complex		Several	(25)
Type 2 diabetes	Calpain 10		2	(26,27)
Type 2 diabetes	PPAR γ Pro12Ala		2	(2)
2. Common variations of plasma lipoproteins				
Trait	Frequency	Gene	No. of common alleles associated with variation	Refs
Familial Hypercholesterolemia	1/500	LDL receptor	Many, all rare	(28)
Familial defective ApoB100	1/800	Apolipoprotein B	2	(28)
Low density lipoprotein cholesterol levels	Quantitative	Apolipoprotein E	3	(28)
Lp(a) levels	Quantitative	Apolipoprotein (a)	>20	(28)
High-density lipoprotein levels	Quantitative	Hepatic lipase	2	(28)
Triglyceride levels	Quantitative	Apolipoprotein A5	2	(29)
Familial combined hyperlipidemia	1/100	ApoAI-C3-A4-A5 cluster	Several	(28)
Serum paraoxonase activity	Quantitative	Paraoxonase 1	2	(30)
Apolipoprotein all levels	Quantitative	Apolipoprotein A2	Multiple	(28)
3. Genetic variations relevant to common diseases differing between the two common inbred mouse strains DBA/2J and C57BL/6J (9)				
Atherosclerosis		Morphine response		
Plasma lipoprotein levels		Pain		
Cardiomyopathy susceptibility		Startle response		
Body fat		Aggression		
Insulin metabolism		Sperm motility		
Bone density		Susceptibility to bacterial infection (numerous)		
Tissue calcification		Aging		
Growth rate		Sleep patterns		
Muscle mass		Iron metabolism		
Seizures		Brain plaques		
Alcohol preference		Cancer susceptibility		

involved. Mitchison (13) has made the analogy between mutations in coding regions (hardware) and mutations in regulatory regions (software). He further divides genes into extrovert, concerned with the outside world (e.g. immunoglobulins and olfactory receptors) and the more numerous introvert genes, concerned with cellular homeostasis (e.g. transcription, signal transduction and neurotransmission). Mitchison suggests that for extrovert genes, hardware variation (coding regions) is more common, while for introvert genes, most genetic variation has mild effects and occurs in the software (regulatory regions). Many instances of promoter variation (software) are found in pharmacogenomics and psychiatry (14). One example is provided by a 44 bp

insertion/deletion polymorphism in the promoter region of the human serotonin transporter (5-HTT) gene. This polymorphism has been associated with affective illness and anxiety-related traits, and also alters 5-HTT gene expression, since the short promoter is less active than the long promoter. Another example of promoter polymorphism is found in the 5-lipoxygenase (5-LO) gene. This gene displays variations in the number of (G + C)-rich Sp1 and Egr-1 transcription factor-binding sites, which influence promoter efficiency through methylation (15). The resulting variations in 5-LO levels may be relevant to the inflammatory response in asthma (16).

It is pertinent to note that whereas our understanding of the kinds of variations that occur in Mendelian diseases is

extensive, our knowledge of the variations that explain common disease is very limited. Relevant to transcriptional regulation, one approach to the identification of common disease gene candidates is to survey cases and controls using expression array analysis. In this regard, it is noteworthy that about one-third of the known common disease genes (Table 1) affect expression levels (8).

Interesting evidence that the inheritance of cancers is distinct from that of most common diseases has been provided by Risch (10). With few exceptions, different cancers have roughly the same familial relative risk (FRR), although the FRR increases with early age of onset. Traditionally, most studies attribute a large environmental component to cancer. The relevant evidence includes such observations as dramatic shifts in cancer incidence due to population migrations. However, careful evaluation of twin and family studies suggest that genetics also plays a large part in cancer risk (10). A key quantity is R_{MD} , the FRR ratio between monozygotic (MZ) and dizygotic (DZ) twins. The idea is that MZ and DZ twins have a comparable sharing of environmental factors, so any difference in concordance rates between MZ and DZ twins must reflect genetic factors. For disease due to a single rare dominant gene, $R_{MD} = 2$, and for a recessive rare gene, $R_{MD} = 4$ but diminishes toward 2 if the allele is very common. For multiple loci, R_{MD} can become very large. For most cancers, $R_{MD} \sim 2$, compared with much higher ratios for many complex diseases such as schizophrenia, multiple sclerosis, and autism. These data are consistent with either rare dominant alleles or additive gene effects for cancer, rather than an interactive, multigenic model.

LINKAGE DISEQUILIBRIUM

Related to the CV/CD hypothesis is the idea that diagnostic costs can be reduced to manageable levels using linkage disequilibrium. The idea here is that the great population expansion is sufficiently recent that large blocks of haplotypes are likely to be conserved between genomes. If true, this would greatly reduce the density of single-nucleotide polymorphisms (SNPs) required for association studies and the expense of genotyping. An implicit assumption of this approach is that common haplotypes are responsible for common diseases.

Estimates of the extent of LD have been obtained by sampling various restricted genomic regions (17), as well as over 51 autosomal regions spanning a total of 13 megabases (18). In addition, somatic cell genetics has been used to analyze LD over the entire chromosome 21 (19). From LD sampling, the average distance over which useful LD was preserved was ~ 50 –60 kb. However, there was considerable variability, and in some loci, LD was found as far away as 500 kb, while in other regions, no useful disequilibrium was found. The chromosome 21 study found considerable linkage disequilibrium, with long blocks of conserved haplotypes. Interestingly, greater than 80% of the haplotype structure could be defined by less than 10% of the SNPs. Based on such studies, it appears that the extent of linkage disequilibrium in the genome will be region- and population-specific.

The dark side of linkage disequilibrium has been revealed by a recent study to identify genes on chromosome 5 involved in Crohn's disease (20). Although a region of linkage disequilibrium associated with the disease was identified, a lack of recombination events within the relevant haplotype block precluded identification of the responsible gene. Thus, although the existence of large haplotype blocks is good news from the point-of-view of cheap diagnostics, it may hinder definitive identification of the responsible genes and any associated biological insights. In this situation, the use of alternate populations may come to the rescue, allowing subdivision of large regions of linkage disequilibrium. For narcolepsy, it was possible to identify a recombination event in a Japanese family that broke up a haplotype block in the HLA region, allowing finer mapping of the responsible gene (21). Alternatively, scientists may have to rely on animal models, such as the mouse, to ultimately identify the causative gene.

Despite continued improvements in genotyping speed, the power of these technologies is not unlimited. At present, hopes are focused on exploiting the structure of the human population to realize the dream of individualized genotyping. This population structure, which results from its recent dramatic expansion, has been theorized to result in simple allelic spectra for common disease genes and considerable linkage disequilibrium. There is presently insufficient data to know whether these expectations will be realized, although there are promising hints.

THE FUTURE OF HUMAN GENETICS

Although particular alleles and haplotypes may be numerically rare in the population as a whole, it is nevertheless likely that each individual will harbor a significant number of such variants. In addition, estimates of gene mutation rates imply that each individual has about three *de novo* gene mutations within their genome. Thus, even if the interaction model proves true for most loci, personalized medicine will still benefit from a more individually tailored genotyping.

The distinctions between the interaction and genetic heterogeneity models can be restated using the following question: 'In the human population, would we rather know all rare 'severe' genetic mutations (e.g. nonsense, frameshift and non-synonymous), or would we rather know all common variations?' With present limitations on the speed of genotyping and lack of knowledge of the genetic causation of common human disease, the best choice between these two alternatives is unclear. Of course, the optimal situation would be to know the answer to both questions. Thus, the need for ever more rapid genotyping and sequencing technologies is likely to continue for the foreseeable future.

The distinctions between the interaction and genetic heterogeneity models can be restated using the following question: 'In the human population, would we rather know all rare 'severe' genetic mutations (e.g. nonsense, frameshift and non-synonymous), or would we rather know all common variations?' With present limitations on the speed of genotyping and lack of knowledge of the genetic causation of common human disease, the best choice between these two alternatives is unclear. Of course, the optimal situation would be to know the answer to both questions. Thus, the need for ever more rapid genotyping and sequencing technologies is likely to continue for the foreseeable future.

ACKNOWLEDGEMENTS

This work was supported by grants from the US NIH (HL028481, HL030568 and DA015802).

REFERENCES

1. Peltonen, L., Palotie, A. and Lange, K. (2000) Use of population isolates for mapping complex traits. *Nat. Rev. Genet.*, **1**, 182–190.
2. Reich, D.A. and Lander, E.S. (2001) On the allelic spectrum of human disease. *Trends Genet.*, **17**, 502–510.

3. Bellus, G.A., Hefferon, T.W., Ortiz de Luna, R.I., Hecht, J.T., Horton, W.A., Machado, M., Kaitila, I., McIntosh, I. and Francomano, C.A. (1995) Achondroplasia is defined by recurrent G380R mutations of FGFR3. *Am. J. Hum. Genet.*, **56**, 368–373.
4. Crow, J.F. (2000) The origins, patterns and implications of human spontaneous mutation. *Nat. Rev. Genet.*, **1**, 40–47.
5. Eyre-Walker, A. and Keightley, P.D. (1999) High genomic deleterious mutation rates in hominids. *Nature*, **397**, 344–347.
6. Peltomaki, P. (2001) Deficient DNA mismatch repair: a common etiologic risk factor for colon cancer. *Hum. Mol. Genet.*, **10**, 735–740.
7. Sakanaranarayanan, K. (1998) Ionizing radiation and genetic risks IX. Estimates of the frequencies of mendelian diseases and spontaneous mutation rates in human populations: a 1998 perspective. *Mutat. Res.*, **411**, 129–178.
8. Online Mendelian Inheritance in Man. <http://www.ncbi.nlm.nih.gov/omim/>
9. Jackson Laboratory Website. <http://www.jax.org>
10. Risch, N. (2001) The genetic epidemiology of cancer: interpreting family and twin studies and their implications for molecular genetic approaches. *Cancer Epidemiol. Biomarkers Prev.*, **10**, 733–741.
11. Aouizerat, B.E., Allayee, H., Cantor, R.M., Davis, R.C., Lanning, C.D., Wen, P.-Z., Dallinga-Thie, G.M., deBruin, T.W.A., Rotter, J.I. and Lusic, A.J. (1999) A genome scan for familial combined hyperlipidemia reveals evidence for a locus on chromosome 11. *Am. J. Hum. Genet.*, **65**, 397–412.
12. Roa, B.B., Boyd, A.A., Volcik, K. and Richards, C.S. (1996) Ashkenazi Jewish population frequencies for common mutations in BRCA1 and BRCA2. *Nat. Genet.*, **14**, 185–187.
13. Mitchison, A. (1997) Partitioning of genetic variation between regulatory and coding gene segments: the predominance of software variation in genes encoding introvert proteins. *Immunogenetics*, **46**, 46–52.
14. Catalano, M. (2001) Functionally gene-linked polymorphic regions and genetically controlled neurotransmitters metabolism. *Eur. Neuropsychopharmacol.*, **11**, 431–439.
15. Uhl, J., Klan, N., Rose, M., Entian, K.D., Werz, O. and Steinhilber, D. (2002) The 5-lipoxygenase promoter is regulated by DNA methylation. *J. Biol. Chem.*, **277**, 4374–4379.
16. In, K.H., Silverman, E.S., Asano, K., Beier, D., Fischer, A.R., Keith, T.P., Serino, K., Yandava, C., De Sanctis, G.T. and Drazen, J.M. (1999) Mutations in the human 5-lipoxygenase gene. *Clin. Rev. Allergy Immunol.*, **17**, 59–69.
17. Doris, P.A. (2002) Hypertension genetics, single nucleotide polymorphisms, and the common disease: common variant hypothesis. *Hypertension*, **39**, 323–331.
18. Gabriel, S.B., Schaffner, S.F., Nguyen, H., Moore, J.M., Roy, J., Blumenstiel, B., Higgins, J., DeFelice, M., Lochner, A., Faggart, M. *et al.* (2002) The structure of haplotype blocks in the human genome. *Science*, **296**, 2225–2229.
19. Patil, N., Berno, A.J., Hinds, D.A., Barrett, W.A., Doshi, J.M., Hacker, C.R., Kautzer, C.R., Lee, D.H., Marjoribanks, C., McDonough, D.P. *et al.* (2001) Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. *Science*, **294**, 1719–1723.
20. Rioux, J.D., Daly, M.J., Silverberg, M.S., Lindblad, K., Steinhart, H., Cohen, Z., Delmonte, T., Kocher, K., Miller, K., Guschwan, S. *et al.* (2001) Genetic variation in the 5q31 cytokine gene cluster confers susceptibility to Crohn disease. *Nat. Genet.*, **29**, 223–228.
21. Miyagawa, T., Hohjoh, H., Honda, Y., Juji, T. and Tokunaga, K. (2000) Identification of a telomeric boundary of the HLA region with potential for predisposition to human narcolepsy. *Immunogenetics*, **52**, 12–18.
22. Ogura, Y., Bonen, D.K., Inohara, N., Nicolae, D.L., Chen, F.F., Ramos, R., Britton, H., Moran, T., Karaliuskas, R., Duerr, R.H. *et al.* (2001) A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. *Nature*, **411**, 603–606.
23. Hugot, J.P., Chamaillard, M., Zouali, H., Lesage, S., Cezard, J.P., Belaiche, J., Almer, S., Tysk, C., O'Morain, C.A., Gassull, M. *et al.* (2001) Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature*, **411**, 599–603.
24. Mackness, M.I., Durrington, P.N., Ayub, A. and Mackness, B. (1999) Low serum paraoxonase: a risk factor for atherosclerotic disease? *Chem. Biol. Interact.*, **119/120**, 389–397.
25. Todd, J.A. (1999) From genome to aetiology in a multifactorial disease, type 1 diabetes. *BioEssays*, **21**, 164–174.
26. Cox, N.J. (2001) Challenges in identifying genetic variation affecting susceptibility to type 2 diabetes: examples from studies of the calpain-10 gene. *Hum. Mol. Genet.*, **10**, 2301–2305.
27. Sreenan, S.K., Zhou, Y.P., Otani, K., Hansen, P.A., Currie, K.P., Pan, C.Y., Lee, J.P., Ostrega, D.M., Pugh, W., Horikawa, Y. *et al.* (2001) Calpains play a role in insulin secretion and action. *Diabetes*, **50**, 2013–2020.
28. Lusic, A.J., Ivandic, B. and Castellani, L.W. (2002) Lipoprotein and lipid metabolism. In Rimoin, D. L., Conner, J. M. Pyeritz, R. E. and Emery, A. E. H. (eds), *Emery and Rimoin's Principles and Practice of Medical Genetics*, 4th edn. Churchill Livingstone, London, pp. 2500–2537.
29. Pennacchio, L.A., Olivier, M., Hubacek, J.A., Cohen, J.C., Cox, D.R., Fruchart, J.C., Krauss, R.M. and Rubin, E.M. (2001) An apolipoprotein influencing triglycerides in humans and mice revealed by comparative sequencing. *Science*, **294**, 169–173.
30. Mackness, B., Durrington, P.N. and Mackness, M.I. (1998) Human serum paraoxonase. *Gen. Pharmacol.*, **31**, 329–336.