

# Basic Molecular and Cell Biology

## Molecular genetics of common diseases

JAMES SCOTT

Coronary heart disease, essential hypertension, diabetes mellitus, senile dementia, manic depressive psychosis, schizophrenia, and other common diseases of adults tend to cluster in families. The familial aggregation of these disorders is rarely caused by a single gene defect; rather, it results from the cumulative interaction of a number of genes with environmental factors. These disorders are therefore said to show multifactorial or polygenic inheritance. The risk of polygenic disease in first degree relatives is generally less than the 1 in 4 risk for mendelian recessive disorders, being about 5-10% (table I). The risk of multifactorial disease, however, varies from one disease to another and from family to family. Within a family the risk depends on the severity of the disorder in the proband, the number of affected family members, and the contribution from environmental risk factors.

TABLE I—Risks for common polygenic diseases of adults<sup>22</sup>

Disorder in proband	Risk for first degree relatives (%)
Coronary heart disease	{ 8 for male relatives 3 for female relatives
Diabetes mellitus	5-10
Epilepsy	5-10
Hypertension	10
Manic depressive psychosis	10-15
Psoriasis	10-15
Schizophrenia	15
Thyroid disease	10

### Number of gene loci implicated in polygenic disease

For any particular disease we need to ascertain the genes that may participate. For a complex disorder such as coronary heart disease, in which plasma lipoproteins, the coagulation system, and the cellular elements of the blood and arterial wall all play a part, the number of genes may be large. I have taken one component of this problem as an example; table II lists genes that may be associated with the regulation of plasma cholesterol concentration. The list is not comprehensive because there are proteins that affect plasma cholesterol concentrations which have not even been identified. The issues are, firstly, to establish which of these genes has a major effect on the genetic variance in plasma cholesterol concentration—the extent to which each gene contributes, if at all, has still to be established—and, secondly, to identify new loci that contribute. For other disorders such as diabetes and manic depression the list of candidate genes is short since few of the gene loci have been identified.

TABLE II—Genes associated with lipid metabolism

Class	Gene
Apolipoproteins	{ Apo-AI Apo-AII Apo-AIV Apo-B Apo-CI Apo-CII Apo-CIII Apo-E
Enzymes	{ HMG CoA reductase Lecithin cholesterol acyl transferase Fatty acyl-CoA cholesterol acyl transferase Endothelial lipoprotein lipase Hepatic triglyceride lipase Fatty acid synthetase Phosphatidic acid phosphohydrolase Cholesterol ester hydrolase Cholesterol 7- $\alpha$ hydrolase
Transfer proteins	Lipid transfer proteins
Receptors	{ Low density lipoprotein Apo-E High density lipoprotein

### CANDIDATE GENES

How can we establish whether a specific gene contributes to a disorder? As an example, let us consider manic depressive psychosis, which may be due to abnormal dopamine metabolism. The enzyme tyrosine hydroxylase has a key regulatory role in dopamine biosynthesis. Perhaps a mutation of the tyrosine hydroxylase gene on chromosome 11 contributes to this disorder. This suggestion can be tested using probes for the tyrosine hydroxylase gene or nearby flanking genes. If a mutation of the tyrosine hydroxylase gene causes the disease cosegregation of a specific allele of the gene with the disease within affected families could be tracked with restriction fragment length polymorphisms. If cosegregation does not exist the tyrosine hydroxylase gene can be eliminated as a suspect in those families. On the other hand, if consistent cosegregation does exist a causal role for the gene in the disease is strongly supported. Remarkably, in a Pennsylvania Amish kindred a locus identical with or adjacent to the tyrosine hydroxylase gene has been associated with this disease.<sup>1</sup>

The tyrosine hydroxylase locus does not, however, provide the whole answer to manic depressive illness. Genetic heterogeneity exists and at least two other loci are also implicated. One has been localised near the fragile site on the X chromosome, and there is good evidence for at least one other locus in the genome.<sup>2,5</sup>

### CLASSIC AND REVERSE GENETICS

Classic genetic studies of chromosome deletions and translocations may provide clues to help identify occult loci responsible for genetic disease. Alzheimer's disease occurs as early as the fourth decade in individuals with trisomy of the short arm of chromosome 21

Division of Molecular Medicine, MRC Clinical Research Centre, Harrow, Middlesex HA1 3UJ

JAMES SCOTT, MSc, FRCP, head of division, consultant physician

(Down's syndrome). Interestingly, the gene that codes for the Alzheimer's amyloid protein has been localised to the short arm of chromosome 21.<sup>6</sup> This gene codes for a membrane protein, possibly a neurotransmitter receptor.<sup>7</sup> In hereditary Alzheimer's disease the gene is duplicated. The similarity between this condition and Down's syndrome is remarkable. With a common disorder such as senile dementia (prevalence 1 in 20 in people over the age of 65) we may expect that common polymorphisms of the Alzheimer's amyloid protein gene might be found to predispose to the disease. Congenital heart disease is also common in patients with Down's syndrome; this may provide a clue to the localisation of one of the genes that lead to this common birth defect. Classic genetics has recently provided a candidate locus on chromosome 5 for a gene that may contribute to schizophrenia.<sup>8</sup> Also the study of X chromosome linked cleft palate and tied tongue has provided a locus for this disorder.<sup>9</sup>

In many genetic diseases the phenotype provides no information about the locus or the biochemical abnormality responsible for the disease. In monogenic disorders the logical extension of the candidate gene approach is to use random deoxyribonucleic acid (DNA) probes and restriction fragment length polymorphism segregation analysis to pinpoint the chromosomal locus responsible for the disease. This approach has been successful in identifying loci for cystic fibrosis<sup>10</sup> (and a candidate gene has been discovered) and for adult polycystic kidney disease.<sup>11</sup> For polygenic disease, where a common phenotype is produced by the interaction of a number of genes, the use of this approach is much more difficult. If clinical material is chosen carefully to maximise single gene effects, however, there is no reason why this approach should not work. The study of large families such as the Pennsylvania Amish with manic depression is a clear example.<sup>1</sup> For the other loci implicated in manic depression a search of the genome using random probes is entirely reasonable.<sup>4,5</sup> For many diseases this may be the only approach.

#### MOUSE MODELS

The study of recombinant inbred strains of mice provides another valuable tool for analysing the complex patterns of inheritance found in polygenic disease and for identifying occult loci. The genetics of insulin dependent diabetes mellitus have been the subject of intensive study in man. Genetic predisposition combined with other (presumably environmental) factors leads to islet destruction and clinical diabetes. The presence of specific alleles in the HLA gene complex (HLA DR3 and DR4) is important for the development of the disease. Hints about the other gene loci that participate have come from the study of non-obese diabetic mice.<sup>12</sup> In these animals three recessive loci on different chromosomes have been implicated. One of these loci is the MHC complex (H-2), another is the Thy-1 locus, and the third has not yet been localised. The THY1 locus in man is on the long arm of chromosome 11. In addition, a dominant locus associated with the control of T cell hyperplasia has been implicated. This model is close to human insulin dependent diabetes and is likely to point to the important loci concerned.

Studies of mouse lipoprotein metabolism have also provided information about the regulation of human plasma lipoprotein concentrations.<sup>13,14</sup> A mouse locus has been identified which regulates the concentration of plasma high density lipoprotein. It is distinct from genes coding for the apolipoprotein, structural components of high density lipoprotein. These studies indicate the presence of a new locus that may be important for determining the risk of atherosclerosis.

#### GENETIC EPIDEMIOLOGY

Genetic epidemiology has provided statistical methods for establishing the contribution of a particular locus to a trait, such as plasma cholesterol concentration.<sup>15</sup> Variation at the apo-E and apo-B loci has been shown to contribute 40% of the total genetic variance which affects plasma cholesterol concentrations.<sup>15,16</sup>

Another risk factor for coronary heart disease is plasma fibrinogen concentration. Genetic variation at the fibrinogen locus has recently been shown to account for 15% of the total phenotypic variance in plasma fibrinogen concentrations.<sup>17</sup> If these observations can be extended to other polygenic traits we can be optimistic that the detailed study of a few loci is likely to allow the refinement of diagnostic tools for predicting the risk of a particular disorder.

#### Genetic polymorphism and susceptibility to disease

The differences between humans in traits such as skin colour, height, intelligence, and blood pressure extend to the ability to handle environmental insults such as exposure to infectious agents and chemical carcinogens and excessive consumption of saturated fat and cholesterol. Individual differences are produced by the DNA sequence variation that exists throughout the genome. The DNA between two individuals varies every 200-500 base pairs.<sup>18</sup> DNA sequence variation is more frequent in the DNA between genes because natural selection conserves sequences that regulate gene expression and code for proteins. Protein polymorphism can be detected by isoelectric focusing, which identifies charged amino acid variants. Charge variants affect one fifth of all proteins (hence genes).<sup>19</sup> Isoelectric focusing certainly underestimates protein polymorphism because it does not detect uncharged amino acid changes; only four of the 20 essential amino acids are charged. Moreover, variation in the DNA sequences concerned with regulation of gene expression will not be detected.

Common alleles (variants at a single locus) or polymorphisms form the basis of human diversity, including the ability to handle environmental challenge. Such alleles are likely to have increased in prevalence owing to positive selection acting on variants that confer selective advantage in the heterozygous state. For example, the genetic polymorphisms that contribute to the common polygenic diseases such as coronary heart disease, essential hypertension, and diabetes mellitus almost certainly have a high prevalence in the population. Perhaps the variants which once conferred selective advantage for our hunter-gatherer ancestors by maintaining blood pressure and blood glucose and plasma cholesterol concentrations in a hostile environment now respond to overnutrition by predisposing to the major diseases that affect modern man.

One of the best researched examples of polymorphism that affects susceptibility to disease occurs at the apo-E locus.<sup>15</sup> The presence of the E4 allele (which occurs in 15% of people) increases plasma cholesterol concentration by about 8% compared with the E3 wild type allele (which occurs in 77% of people), whereas the E2 allele (which occurs in 8%) decreases cholesterol concentration by 13%. These small changes in plasma cholesterol concentration are sufficient to produce a substantial enrichment of the E2 allele with advancing age and decrease of the E4 allele in the aging population, presumably as a result of an increased number of deaths from myocardial infarction. Alleles at the apo-B locus similarly affect plasma cholesterol concentration and the risk of myocardial infarction.<sup>16,20</sup> A most important example of genetic polymorphism which leads to susceptibility to infection with the human immunodeficiency virus and progression to the acquired immune deficiency syndrome (AIDS) has recently been described at the  $\alpha^2$  globulin locus (vitamin D binding factor).<sup>21</sup> Selection acting on these alleles could act rapidly in the absence of a cure for AIDS.

Rare alleles may also contribute to genetic diversity and susceptibility to disease. Homozygosity for a defect of the gene cystathionine synthetase causes collagen malformation and vascular thrombosis. Heterozygosity (prevalence 1 in 70) for the same rare allele has now been associated with increased risk of atherosclerotic peripheral vascular disease and cerebrovascular disease.<sup>22</sup>

#### IDENTIFICATION OF POLYMORPHIC ALLELES THAT PREDISPOSE TO POLYGENIC DISEASE

Charge variants may be identified by isoelectric focusing and direct protein sequencing. Occasionally, close association (linkage

disequilibrium) can be shown between a restriction fragment length polymorphism and a mutation causing disease. The first such association was found between an HpaI restriction fragment length polymorphism and the mutation causing sickle cell disease.<sup>23</sup> Further examples of such polymorphisms associated with predisposition to polygenic disease have now been described. These include those associated with increased plasma concentrations of lipoproteins<sup>16,24</sup> and fibrinogen,<sup>17</sup> increased risk of cardiovascular disease,<sup>20</sup> and the risk of amyloidosis complicating juvenile arthritis.<sup>25</sup> An alternative and in many cases more searching approach is to identify restriction fragment length polymorphism haplotypes (restriction fragment length polymorphisms in linkage disequilibrium at the same locus) that are associated with disease traits. Ultimately, there is no substitute for identifying precise mutation by DNA sequence analysis. Once a mutation has been identified its detection by amplification of genomic DNA and the use of allele specific oligonucleotide probes makes diagnosis a relatively trivial matter.<sup>26</sup>

### Linkage and linkage disequilibrium

Two or more gene loci that exist in close proximity on the same chromosome and cosegregate in family studies are said to show genetic linkage. If linked genes confer susceptibility to a disease this will have implications for the level of risk, particularly if disease susceptibility alleles are in linkage disequilibrium. Linkage disequilibrium is best illustrated by the HLA complex.<sup>27</sup> The HLA genes are located on the short arm of chromosome 6 and are made up of four closely linked loci (A, B, C, and D). The products of these genes are proteins concerned with the recognition of foreign antigens. Each HLA locus consists of multiple alleles, of which an individual may inherit any two of possibly 20. Certain HLA genes occur together more often than would be predicted by chance alone. HLA A1 and B8 in north European white populations are at least four times as common as would be expected by chance and are described as being in linkage disequilibrium. It is highly likely that this association was selected during the evolutionary history of the north European white population; it may have conferred protection against the plague. An example of selection may be seen among the ancestors of Dutch settlers in Surinam: most of the original settlers succumbed to a typhoid epidemic, and only those with particular HLA antigens survived.<sup>28</sup> New examples of linkage disequilibrium between disease susceptibility alleles are likely to emerge, especially where selection has favoured their close association. The example of polymorphisms which produce a relatively high blood pressure and high blood glucose and serum cholesterol concentrations in order to combat a hostile environment has already been given.

### The future

Before the year 2000 linkage and physical maps of the entire human genome will have been established and much of the genome will have been sequenced.<sup>29</sup> Most of the genes implicated in common polygenic disease of adults and common birth defects of children will have been characterised, and the mutant alleles which predispose to disease will have been fully identified. By use of procedures for amplifying DNA and allele specific oligonucleotide probes it should be possible to screen for polymorphisms associated with a variety of common diseases in a single reaction.

Screening in early life for genes which predispose to the common ailments later in life has clear advantages but poses ethical and practical problems.<sup>30</sup> For a disorder which can be prevented such as coronary heart disease it is a remarkable bonus to know and to treat

early. Predictive screening may lead to modification of lifestyle or the introduction of specific treatment. Difficulties can be envisaged, however: termination of pregnancy, lifestyle, health insurance, or occupational decisions could be based on relatively minor and ill judged genetic risks. And will it pay? For monogenic disorders it is certainly cheaper to screen antenatally than to care for the chronically handicapped.<sup>31</sup> For common diseases of adults the answer is also likely to be yes—for example, in terms of health care alone coronary heart disease has been estimated to cost £1 billion a year, and this does not count the even greater cost to industry and to commerce. The bill could probably be halved by appropriate preventive measures.

I thank Miss Lesley Sargeant for typing the manuscript.

### References

- Egeland JA, Gerhard DS, Pauls DL, *et al*. Bipolar affective disorders linked to DNA markers on chromosome 11. *Nature* 1987;325:783-7.
- Baron M, Risch N, Hamburger R, *et al*. Genetic linkage between X-chromosome markers and bipolar affective illness. *Nature* 1987;326:289-92.
- Deterra-Wadleigh SD, Berrettini WH, Goldin LR, Boorman D, Anderson S, Gershon ES. Close linkage of c-Harvey-ras-1 and the insulin gene to affective disorder is ruled out in three North American pedigrees. *Nature* 1987;325:806-8.
- Hodgkinson S, Sherrington R, Gurling H, *et al*. Molecular genetic evidence of heterogeneity in manic depression. *Nature* 1987;325:805-6.
- Mendlewicz J, Simon P, Sevy S, *et al*. Polymorphic DNA marker on X chromosome and manic depression. *Lancet* 1987;ii:1230-4.
- Delabar J-M, Goldgaber D, Lamour Y, *et al*. B amyloid gene duplication in Alzheimer's disease and karyotypically normal Down syndrome. *Science* 1987;235:1390-2.
- Kang J, Lemaire H-G, Unterbeck A, *et al*. The precursor of Alzheimer's disease amyloid A4 protein resembles a cell-surface receptor. *Nature* 1987;325:733-6.
- Bassett AS, Jones BD, McGillivray BC, Pantzar JT. Autosomal abnormality linked to schizophrenia. In: *New research. Programme and abstracts of the American Psychiatric Association's 140th annual general meeting, May 1987, Chicago, Illinois*. Washington: American Psychiatric Association, 1987:61.
- Moore GE, Ivens A, Chalmers J, *et al*. Linkage of an X-chromosome cleft palate gene. *Nature* 1987;326:91-2.
- Estivill X, Farrall M, Scambler PJ, *et al*. A candidate for the cystic fibrosis locus isolated by selection for methylation-free islands. *Nature* 1987;326:840-5.
- Reeders ST, Breuning MH, Davies KE, *et al*. A highly polymorphic DNA marker linked to adult polycystic kidney disease on chromosome 16. *Nature* 1985;317:542-4.
- Prochazka M, Leiter EH, Serreze DV, Coleman DL. Three recessive loci required for insulin-dependent diabetes in non-obese diabetic mice. *Science* 1987;237:286-9.
- Lusis AJ, Taylor BA, Quon D, Zolman S, LeBoeuf RC. Genetic factors controlling structure and expression of apolipoproteins B and E in mice. *J Biol Chem* 1987;262:7594-604.
- Paigen B, Mitchell D, Reue K, Morrow A, Lusis AJ, LeBoeuf RC. *Ath-1*, a gene determining atherosclerosis susceptibility and high density lipoprotein levels in mice. *Proc Natl Acad Sci USA* 1987;84:3763-7.
- Sing CF, Davignon J. Role of apolipoprotein E polymorphism in determining normal plasma lipid and lipoprotein variation. *Am J Hum Genet* 1985;37:268-85.
- Law A, Wallis SC, Powell LM, *et al*. Common DNA polymorphisms within coding sequence of apolipoprotein B gene associated with altered lipid levels. *Lancet* 1986;ii:1301-3.
- Humphries SE, Cook M, Dubowitz M, Stirling Y, Meade TW. Role of genetic variation at the fibrinogen locus in determination of plasma fibrinogen concentrations. *Lancet* 1987;ii:1452-4.
- Jeffreys AJ. DNA sequence variants in the G $\gamma$ -A $\gamma$ - $\delta$ - and B-globin genes of man. *Lancet* 1987;ii:1121-2.
- Rosenblum BB, Neel JV, Hanash SM. Two-dimensional electrophoresis of plasma polypeptides reveals "high" heterozygosity indices. *Proc Natl Acad Sci USA* 1983;80:5002-6.
- Hegele RA, Huang L-S, Herbert PN, *et al*. Apolipoprotein B—gene DNA polymorphisms associated with myocardial infarction. *N Engl J Med* 1986;315:1509-15.
- Eales L-J, Nye KE, Parkin JM, *et al*. Association of different allelic forms of group specific component with susceptibility to and clinical manifestation of human immunodeficiency virus infection. *Lancet* 1987;ii:999-1002.
- Boers GHJ, Smals AGH, Trijbels FJM, *et al*. Heterozygosity for homocystinuria in premature peripheral and cerebral occlusive arterial disease. *N Engl J Med* 1985;313:709-15.
- Kan YW, Dozy AM. Polymorphism of DNA sequence adjacent to human  $\beta$ -globin structural gene: relationship to sickle mutation. *Proc Natl Acad Sci USA* 1978;75:5631-5.
- Scott J, Knott TJ, Priestley LM, *et al*. High density lipoprotein composition is altered by a common DNA polymorphism adjacent to apoprotein AII gene in man. *Lancet* 1985;ii:771-3.
- Woo P, O'Brien J, Robson M, Ansell BM. A genetic marker for systemic amyloidosis in juvenile arthritis. *Lancet* (in press).
- Saiki RK, Scharf S, Faloona F, *et al*. Enzymatic amplification of B-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* 1985;230:1350-4.
- Harris H. *The principles of human biochemical genetics*. New York: Elsevier, 1980;316-406.
- de Vries RRP, Khan PM, Bernini LF, van Loghem E, van Rood JJ. Genetic control of survival in epidemics. *J Immunogenet* 1979;6:271-87.
- Anonymous. Mapping the human genome [Editorial]. *Lancet* 1987;ii:1121-2.
- Rowley PT. Genetic screening: marvel or menace. *Science* 1984;225:138-44.
- Chapple JC, Dale R, Evans BG. The new genetics: will it pay its way? *Lancet* 1987;ii:1189-92.
- Goldstein JL, Brown MS. Biological considerations in the approach to clinical medicine. In: Petersdorf RG, Adams RD, Braunwald E, Isselbacher KJ, Martin JB, Wilson JD, eds. *Principles of internal medicine*. New York: McGraw-Hill, 1983:311-23.