



# **MOLECULAR APPROACHES TO HUMAN POLYGENIC DISEASE**

*A Wiley – Interscience Publication*

1987

**JOHN WILEY & SONS**

---

Chichester · New York · Brisbane · Toronto · Singapore



---

**MOLECULAR  
APPROACHES TO  
HUMAN POLYGENIC  
DISEASE**

---

The Ciba Foundation is an international scientific and educational charity. It was established in 1947 by the Swiss chemical and pharmaceutical company of CIBA Limited—now CIBA-GEIGY Limited. The Foundation operates independently in London under English trust law.

The Ciba Foundation exists to promote international cooperation in biological, medical and chemical research. It organizes about eight international multidisciplinary symposia each year on topics that seem ready for discussion by a small group of research workers. The papers and discussions are published in the Ciba Foundation symposium series. The Foundation also holds many shorter meetings (not published), organized by the Foundation itself or by outside scientific organizations. The staff always welcome suggestions for future meetings.

The Foundation's house at 41 Portland Place, London, W1N 4BN, provides facilities for meetings of all kinds. Its Media Resource Service supplies information to journalists on all scientific and technological topics. The library, open seven days a week to any graduate in science or medicine, also provides information on scientific meetings throughout the world and answers general enquiries on biomedical and chemical subjects. Scientists from any part of the world may stay in the house during working visits to London.

---



# **MOLECULAR APPROACHES TO HUMAN POLYGENIC DISEASE**

*A Wiley – Interscience Publication*

1987

JOHN WILEY & SONS

---

Chichester · New York · Brisbane · Toronto · Singapore

© Ciba Foundation 1987

Published in 1987 by John Wiley & Sons Ltd, Chichester, UK.

Suggested series entry for library catalogues:  
Ciba Foundation Symposia

Ciba Foundation Symposium 130  
× + 274 pages, 43 figures, 33 tables

**Library of Congress Cataloging in Publication Data:**

Molecular approaches to human polygenic disease.

(Ciba Foundation symposium; 130)

'Symposium on Molecular Approaches to Human Polygenic Disease, held at the Ciba Foundation, London, 25-27 November 1986'—Contents p. v

Editors: Gregory Bock (organizer) and GERALYN M. COLLINS.

'A Wiley-Interscience publication.'

Includes indexes.

1. Genetic disorders—Congresses. 2. Medical genetics—Congresses. 3. Molecular genetics—Congresses. I. Bock, Gregory. II. Collins, GERALYN M. III. Symposium on Molecular Approaches to Human Polygenic Disease (1986 : Ciba Foundation)

[DNLM: 1. Genetics—congresses. 2. Hereditary Diseases—congresses. 3. Linkage (Genetics)—congresses. W3 C161F v.130 / QZ 50 M718 1986] RB155.M625 1987 616'.042 87-13374

ISBN 0 471 91096 1

**British Library Cataloguing in Publication Data:**

Molecular approaches to human polygenic disease.—(CIBA Foundation Symposium; 130).

1. Medical genetics

I. Series

616'.042 RB155

ISBN 0 471 91096 1

Typeset by Inforum Ltd, Portsmouth  
Printed and bound in Great Britain.

# Contents

*Symposium on Molecular Approaches to Human Polygenic Disease, held at the Ciba Foundation, London, 25–27 November 1986*

*The topic for this symposium was proposed by Professor R. Williamson*

*Editors: Gregory Bock (Organizer) and GERALYN M. COLLINS*

**Sir David Weatherall** Introduction 1

**R. Williamson** Human gene mapping 3

*Discussion* 9

**K. Berg** Genetics of coronary heart disease and its risk factors 14

*Discussion* 28

**C.G. Davis** The LDL receptor: oligonucleotide-directed mutagenesis of the cytoplasmic domain 34

*Discussion* 46

**G. Utermann** Apolipoproteins, quantitative lipoprotein traits and multifactorial hyperlipidaemia 52

*Discussion* 63

**J.M. Taylor, S. Lauer, N. Elshourbagy, C. Reardon, E. Taxman, D. Walker, D. Chang and Y-K. Paik** Structure and evolution of human apolipoprotein genes: identification of regulatory elements of the human apolipoprotein E gene 70

*Discussion* 81

**General discussion** How well can DNA polymorphisms be correlated with clinical conditions? 87; Disease associations and 'heritability' 90; Family studies, population studies and linkage disequilibrium 92

**C.F. Sing and E.A. Boerwinkle** Genetic architecture of inter-individual variability in apolipoprotein, lipoprotein and lipid phenotypes 99

*Discussion* 122

- S.E. Humphries, P.J. Talmud and A.M. Kessling** Use of DNA polymorphisms of the apolipoprotein genes to study the role of genetic variation in the determination of serum lipid levels 128  
*Discussion* 145
- A.G. Motulsky, W. Burke, P.R. Billings and R.H. Ward** Hypertension and the genetics of red cell membrane abnormalities 150  
*Discussion* 161
- G.I. Bell, K. Xiang, S. Horita, N. Sanz and J.H. Karam** The molecular genetics of diabetes mellitus 167  
*Discussion* 179
- K.K. Kidd** Searching for major genes for psychiatric disorders 184  
*Discussion* 193
- J.I. Bell** The molecular genetics of HLA-related disorders 197  
*Discussion* 207
- W.F. Bodmer** The human genome sequence and the analysis of multifactorial traits 215  
*Discussion* 225
- M.A. Eglitis, P.W. Kantoff, J.R. McLachlin, A. Gillio, A.W. Flake, C. Bordignon, R.C. Moen, E.M. Karson, J.A. Zwiebel, D.B. Kohn, E. Gilboa, R.M. Blaese, M.R. Harrison, E.D. Zanjani, R. O'Reilly and W.F. Anderson** Gene therapy: efforts at developing large animal models for autologous bone marrow transplant and gene transfer with retroviral vectors 229  
*Discussion* 241
- G. Rose** Implications of genetic research for control measures 247  
*Discussion* 252
- Final general discussion** Genetic investigation of a disease with a complex pathogenesis 257; Relating genetic information to clinical phenotypes: ethical issues 261
- Sir David Weatherall** Chairman's summing-up 265
- Index of contributors 267
- Subject index 000



# Participants

**G.I. Bell** The University of Chicago, Howard Hughes Medical Institute,  
920 East 58 Street, Chicago, Illinois 60637, USA

**J.I. Bell\*** Division of Immunology, Department of Medicine, Stanford  
University School of Medicine, Stanford, California 94305, USA

**K. Berg** Institute of Medical Genetics, University of Oslo, PO Box 1036,  
Blindern, Oslo 3, Norway

**M. Bobrow** Paediatric Research Unit, United Medical & Dental Schools of  
Guy's and St Thomas's Hospitals, The Prince Philip Research Laboratories,  
Guy's Tower, London Bridge, London SE1 9RT, UK

**Sir Walter Bodmer** Imperial Cancer Research Fund, PO Box 123,  
44 Lincoln's Inn Fields, London WC2A 3PX, UK

**E.A. Boerwinkle** (*Ciba Foundation Bursar*) Center for Demographic &  
Population Genetics, Graduate School Building, South 250, University of  
Texas at Houston, PO Box 20334, Houston, Texas 77225, USA

**C.G. Davis** Division of Allergy & Immunology, Room U, 426, Department  
of Medicine, University of California — San Francisco, School of Medicine,  
San Francisco, California 94143-0724, USA

**M.A. Eglitis** Laboratory of Molecular Hematology, National Heart, Lung  
and Blood Institute, Building 10, Room 7D18, National Institutes of Health,  
Bethesda, Maryland 20892, USA

**M.A. Ferguson-Smith** University Department of Medical Genetics, Duncan  
Guthrie Institute of Medical Genetics, Yorkhill, Glasgow G3 8SJ, UK

**S.E. Humphries** The Charing Cross Sunley Medical Research Centre,  
Lurgan Avenue, Hammersmith, London W6 8LW, UK

---

\* *Present address:* Nuffield Department of Clinical Medicine, John Radcliffe Hospital,  
Headington, Oxford OX3 9DU, UK

**J. Kaprio** (*Ciba Foundation Bursar*) The Finnish Twin Cohort Study, Dept of Public Health Science, University of Helsinki, Kallioliinatie 4, SF-00140 Helsinki, Finland

**K.K. Kidd** Department of Human Genetics, Yale University School of Medicine, I-310 SHM, PO Box 3333, New Haven, Connecticut 06510-8005, USA

**G.M. Lathrop** Howard Hughes Medical Institute Research Laboratories, University of Utah, 603 Wintrobe Building, Salt Lake City, Utah 84132, USA

**J.K. Lloyd** Department of Child Health, Institute of Child Health, 30 Guilford Street, London WC1N 1EH, UK

**T.W. Meade** MRC Epidemiology & Medical Care Unit, Northwick Park Hospital, Watford Road, Harrow, Middlesex HA1 3UJ, UK

**M. Mikkelsen** Department of Medical Genetics, The John F Kennedy Institute, 7GL Landevej, DK-2600 Glostrup, Denmark

**A.G. Motulsky** Department of Medicine & Genetics, Center for Inherited Diseases RG-25, University of Washington School of Medicine, Seattle, WA 98195, USA

**M.F. Oliver** Cardiovascular Research Unit, Department of Medicine, University of Edinburgh, Hugh Robson Bldg, George Square, Edinburgh EH8 9XF, UK

**G.A. Rose** Division of Medical Statistics & Epidemiology, London School of Hygiene & Tropical Medicine, Keppel Street, (Gower Street), London WC1E 7HT, UK

**J. Scott** Molecular Medicine Research Group, Division of Clinical Sciences, MRC Clinical Research Centre, Watford Road, Harrow, Middlesex HA1 3UJ, UK

**C.F. Sing** Department of Human Genetics, University of Michigan School of Medicine, 4708 Medical Science II, Box 015, Ann Arbor, Michigan 48109-0010, USA

**J.M. Taylor** Gladstone Foundation Laboratories, University of California at San Francisco, PO Box 40608, San Francisco, CA 94140, USA

**G. Utermann** Institute for Medical Biology & Genetics, University of  
Innsbruck, Schöpfstrasse 41, A-6020 Innsbruck, Austria

**Sir David Weatherall** (*Chairman*) Nuffield Department of Clinical Medicine,  
John Radcliffe Hospital, Headington, Oxford OX3 9DU, UK

**R. Williamson** Department of Biochemistry, St Mary's Hospital Medical  
School, Norfolk Place, London W2 1PG, UK



# Introduction

Sir David Weatherall

*Nuffield Department of Clinical Medicine, John Radcliffe Hospital, Headington, Oxford OX3 9DU, UK*

*1987 Molecular approaches to human polygenic disease. Wiley, Chichester (Ciba Foundation Symposium 130) p 1–2*

Over the past few years the application of the new methods of recombinant DNA technology has told us a great deal about the molecular pathology of single-gene disorders. Indeed, it is likely that we already have a good idea about the repertoire of the different mutations that underlie these diseases. Many human genes have been cloned and restriction fragment length polymorphisms (RFLPs) have been defined, and these markers, together with a variety of anonymous probes and probes for highly variable regions (HVRs), are allowing us to build up maps of many parts of the human genome. It seems likely that within the foreseeable future we shall have a map of large areas of the genome. These advances have already had valuable practical application for carrier detection and for prenatal diagnosis of genetic disease and it is probable that in the near future we shall be able to start gene therapy, at least for a few single-gene disorders.

When it comes to the genetic analysis of common conditions like coronary artery disease, diabetes, autoimmune disease, the major psychoses, and other important disorders of western societies, the position is much more complicated. Epidemiological evidence suggests that many of these conditions have a strong environmental component in their aetiology and, although genetic factors are undoubtedly involved, the conditions do not follow any clear-cut pattern of inheritance. Pedigree analyses are bedevilled by difficulties of assignment and although complex statistical methods are available for the analysis of polygenic inheritance, in practice they are often difficult to apply. However, a number of putative 'candidate genes' exist for many of these conditions and, hence, it has been suggested that the new techniques of molecular biology will also be applicable to the analysis of these very complex disorders.

The objective of this symposium is to try to define better the problems of multifactorial inheritance and its analysis by recombinant DNA technology and, hence, to determine which directions this field might follow in the future.

This will not be easy. My own belief is that it may be a very long time before we obtain enough information to be able to make useful predictions about high risk groups of individuals for any of these common diseases. However, what these studies might do is to teach us more about the basic pathogenesis of this important group of diseases; such information is badly needed, because we know so little about their aetiology. Our current approaches to management are almost entirely symptomatic and are many steps removed from disease processes at the cellular and molecular level. It is still something of an act of faith to believe that such information will change clinical practice. Nevertheless, because it is equally unlikely that the complete removal of environmental 'risk' factors is possible, it is important that we examine the basic pathophysiology of the conditions. We need to attack them from *both* directions.

# Human gene mapping

Robert Williamson

*Department of Biochemistry, St Mary's Hospital Medical School, University of London, Norfolk Place, London W2 1PG, UK*

*Abstract.* It is now possible to map the human genome completely with a set of closely linked markers. Over 500 coding genes have been cloned and localized, as have approximately 2000 anonymous DNA fragments, most of which recognize two-allele polymorphisms that are caused by single base changes which alter the recognition site for a restriction enzyme (restriction fragment length polymorphisms). Most human chromosomes have been mapped, with markers in defined order placed approximately 10 map units apart. Chromosomes X and 21 are particularly well mapped, with over 200 probes ordered on X. The strategy during the next few years will encompass moving from a linkage map to a set of overlapping cosmid or phage clones, and finally to a complete sequence of regions of chromosomes and entire chromosomes. A complete sequence of the human genome should transform our understanding of development, the control of gene expression, and the parameters of genetic disease.

*1987 Molecular approaches to human polygenic disease. Wiley, Chichester (Ciba Foundation Symposium 130) p 3-13*

Less than 10 years ago, Kan & Dozy (1978) identified a variation in DNA sequence adjacent to the human  $\beta$ -globin gene. This sequence change was remarkable in several ways. It occurred not in a coding sequence for a protein, but several thousand base pairs downstream from the structural genes for the  $\beta$ -globin chains. It gave two sequences, only one of which could be recognized by a restriction endonuclease, and it therefore gave two DNA fragments of different sizes after digestion. Finally, it was linked to the gene that codes for sickle cell  $\beta$ -globin; because the  $\beta^S$ -globin gene is selected for in heterozygotes, one allele was found preferentially in association with the mutant gene.

This is a classical case of linkage disequilibrium, when an allelic marker is found to be close enough to another gene to be co-inherited at greater than random frequency, and also when one of the two alleles is found to be associated with one of two possible neighbouring phenotypes (or genotypes). Of course, both linkage and disequilibrium had been recognized for many years for protein variants and phenotypes, but this finding was of monumental significance, for two reasons, as recognized immediately by

Solomon & Bodmer (1979). First, single base changes in the DNA sequence are far from rare; Jeffreys (1979) estimated that they occur once in every hundred or so base pairs, and while this estimate may be on the high side (since it was determined for a population rather than for individuals) there is little doubt that each person has several million single base-change differences between the two corresponding haploid genome sets found in each cell. Second, most of these differences occur in DNA that we assume to be neutral, between genes rather than in coding sequences. Therefore, unlike protein differences which are often deleterious or selective, DNA differences may be passed from generation to generation, apparently making little or no difference to the individual.

Chromosomes are jumbled during meiosis by recombination. Cross-overs occur at least once for each chromosome and, at most, four or five times for the largest chromosomes. There are perhaps 50 meiotic exchanges in all, per chromosome set per generation – a very small number of randomizations compared to the very large number of potential markers. It is the linking together of these markers, and their linkage in turn to phenotypes (whether a normal variant or a pathological condition) that has revolutionized human genetics. And more is to come, for within the next few years the human genome will be sequenced in its entirety, leading to further advances in understanding of gene organization and expression.

It is still unclear how much of the DNA of humans and other mammals codes for protein; probably no more than 5%, although much of the rest is interspersed as intervening sequences (or introns) between blocks that specify amino acids. The DNA sequence is specific (more or less) when it is coding for a protein; it must be, since alterations in the amino acid sequence would change the properties of the polypeptide and would have phenotypic consequences. The DNA sequence is less specific in introns, and shows most variation from person to person in sequences that separate one gene from another, or where there are stretches of short repeats that probably fulfil a structural role or are sites for recombination.

Two main methods are used to visualize sequence differences between two homologous chromosomes. The first is to determine the order of the DNA bases directly, and then to synthesize a short single-strand sequence that is homologous to the region where the change occurs. Even a single base mismatch is sufficient to cause destabilization of the double helix and, if the oligonucleotide probe is labelled, it will remain hybridized to the perfectly complementary strand at a higher temperature than to the mutated strand (Thein & Wallace 1986). The alternative, and more traditional, method is to follow the inheritance of restriction sites by the size of the DNA fragments that are generated. The Southern blot technique reveals the size of a hybridizing fragment by its rate of migration through an agarose gel; characteristic band sizes are seen for each polymorphic variant, all of which are inherited in a Mendelian fashion (White et al 1985).



If any sequence, whether it is a coding gene or a marker for a chromosome region, is to be followed through a family to reveal its function, one should be able to recognize it uniquely. For a majority of the human DNA sequence, hybridization (sequence pairing) is so specific that a complementary DNA molecule (a 'gene-specific probe') will recognize the sequence perfectly when one uses defined salt concentrations and temperature. It is usual to clone such gene probes in plasmids or phage, to obtain biological replication that gives large amounts of pure sequence, although chemical synthesis is now an alternative.

Over 500 genes have been cloned to date. These include most of the genes that code for major structural proteins, and many that code for enzymes. If a protein can be identified as a 'spot' on a two-dimensional electrophoresis pattern, the corresponding DNA sequence can usually be isolated by using 'reverse genetics' (Glover 1985). Each of these DNA sequences is a marker, both for the protein that it encodes and for the chromosomal region where it resides. If the gene also specifies a pathological condition when mutated, the level of the defect can easily be determined by gene analysis (Cooper & Schmidtke 1986).

What of the vast majority of the 3000 or so single-gene defects, for which no causative protein defect is known? For the X-linked diseases such as Duchenne muscular dystrophy and chronic granulomatous disease, at least the chromosome is known, and linkage studies to determine the region of the mutation are possible (Davies et al 1983). It is then possible to 'walk' from the linked gene to the defect itself, which is usually a coding sequence, by using what is rapidly becoming a standard armamentarium of molecular procedures. Among these are pulse-field gel electrophoresis, directional walking vectors, *NotI* junction libraries for ordering chromosome fragments, methods for selecting small regions of human chromosomes in mouse cells ('chromosome-mediated gene transfer'), cross-screening of cDNA and genomic libraries, and the construction of cosmid overlap maps. My objective is not to discuss these techniques in detail, but merely to catalogue them to demonstrate that an *entirely new* range of techniques is coming to the fore in molecular biology, to add to the conventional cloning and sequencing strategies of the past five years. One might realistically describe these as a far more sophisticated set of technological procedures for studying complex genomes.

These techniques are not only being applied to the X chromosome but also to both dominant and recessive autosomal diseases, such as Huntington's chorea and cystic fibrosis (Gusella et al 1984, Williamson 1987, Estivill et al 1987). Each has been localized precisely to a small region of a specific human chromosome, and attempts are now being made to 'walk' to the defective gene. It is only in this way that population-based carrier testing and new developments in treatment will take place.

However, rather than discuss single-gene defects at great length, I would like to outline some work we have been doing at St Mary's, and to speculate

on where it might lead. We have been attempting to determine which genes lead to a high risk of coronary artery disease (CAD) and of cleft palate (CP), in part for their own sake, and partly as a paradigm of polygenetic and multifactorial inheritance.

*Most* common diseases are partly of genetic origin. The balance between genetic and environmental causation can be estimated by studying the independent occurrence of disease in first-degree relatives. This approach is even more conclusive if the disease strikes relatives who do not share a common environment, as for twins or siblings who are separated during childhood. In this way, it has been shown that coronary artery disease, hypertension, some forms of cancer (particularly cancers of the breast and colon), diabetes, manic-depressive psychosis, Alzheimer's disease and schizophrenia are all, in part, inherited. However, in every case environment also plays a part, as shown most conclusively by the fact that both twins in an identical pair (who *must* share the same genotype) do not always develop the disease.

Because DNA is so complex, it might seem that an infinite number of possible genes might affect (for instance) the level of blood cholesterol, artery wall structure, enzymes of lipid metabolism and the like, each of which might play a part in CAD risk. However, a surprisingly small number of allelic genes that determine a trait is sufficient to give a gaussian distribution of a variable in a population. In some diseases the co-inheritance of only two genes can dramatically alter the clinical picture, as for thalassaemia, a disease that we understand well (Weatherall & Clegg 1981). In a mild form,  $\beta$ -thalassaemia intermedia, the co-inheritance of a defective  $\alpha$ -globin gene and a pair of defective  $\beta$ -globin genes leads to a *less* severe disease than the classical  $\beta$ -thalassaemia. The patient has a less marked chain imbalance; the more 'serious' is the compensating  $\alpha$ -thalassaemia (within limits), then the more likely is the 'patient' to be healthy.

Thalassaemia intermedia is a polygenic disease. The genes that code for  $\alpha$ - and  $\beta$ -globins are on different chromosomes, and do not interact at the DNA level, but they do so only when the proteins have been synthesized in the cell. Therefore, some forms of the 'simple' disease  $\beta$ -thalassaemia are just as polygenic as more complex conditions, such as coronary artery disease, since the genes that determine the clinical phenotype in each case can interact and compensate (in this case, to the 'patient's' benefit) only at the cellular level.

In most multifactorial diseases, there are clues to some candidate genes. We use the term 'candidate gene' for a disease to designate a DNA sequence for which there is evidence that, at least in some cases, the gene is involved in increasing or reducing the risk of the disease developing. Such an inference may be made because in some family or other there is a demonstrable and major gene defect (seen, perhaps, as the absence of a protein, or as a functional change) which causes a related (but not identical) clinical syndrome. Alternatively, there may be epidemiological evidence that risk varies

with the amount of a protein or enzyme, or with the presence or absence of a structural variant of it.

For coronary artery disease, the candidate genes include those coding for: the apolipoproteins, which carry cholesterol and lipids from food via the hepatic portal circulation to the tissues of the body; the receptors (and, in particular, the LDL receptor) that recognize the circulating lipoproteins and cause them to enter the cells that line the arterial walls; the enzymes, such as HMGCoA reductase and lecithin-cholesterol acyltransferase, which regulate cholesterol biosynthesis; and proteins such as fibrinogen, high levels of which are associated with high risks in the population. These genes can now be followed, singly and in cohorts, through families to see if they really are the arbiters of an increased risk of heart attack.

Therefore, it is already possible to generalize about the scientific prerequisites for accurate genetic analysis of the inherited contribution to multifactorial diseases. First, it is important to have access to a set of gene markers, preferably located accurately on chromosomes, and either linked to one another or coding for known proteins. Large families, not necessarily with any inherited disease, are required to determine linkage between probes, and cosmid libraries, preferably specific to human chromosomes, are needed for gene 'walking'. Such families have become available through the resources of the Centre d'Etude du Polymorphisme Humain, in Paris, which provides DNA samples in return for the resulting linkage data. Therefore, both the gene map and the reference families can be considered as available to molecular geneticists for human studies.

Secondly, it is essential to have sufficient families who have been well characterized clinically for a given disease, and by the same set of criteria. This is relatively simple for, say, hypertension; obviously it is a great deal more difficult for schizophrenia. It is often valuable to have a 'core population' of patients who have been seen by one major clinical department, as this tends to lead to more uniform diagnostic criteria (but does not guarantee it). It is particularly useful to have access to a few very large families from geographically (or culturally) isolated groups, such as the Old-Order Amish sect. In such families it is more likely that a particular gene will show 'mendelian' inheritance because of the homozygosity of other genes that, in outbred families, act as variable modifiers.

Thirdly, it is necessary to have clues. These may be sequences that are clearly candidate genes (as the LDL receptor in coronary artery disease), or genes that are identified functionally (one must regard the gene for insulin as a candidate gene for diabetes in this context). Even if a protein is known to have a normal structure, the gene may still be mutated so as to cause reduced expression, as in  $\beta$ -thalassaemia. Other inspired guesses about candidate genes may rely on chromosomal localization (as for atrioventricular septum defects and chromosome 21q, because of the high incidence of the defects in

Down's syndrome) or on linkage (as for haemochromatosis and HLA).

A particularly valuable kind of candidate gene is a structural one which, when mutated, produces a pathological condition in a few well studied families, from which generalizations can be made. It is for this reason that we are particularly excited to be able to follow a family of 220 persons, in which cleft palate is segregating as a single-gene mendelian trait. Using this very powerful family has enabled us to find a linkage which locates the defect (Moore et al 1987).

Cleft palate, in general, is neither sex-linked (in fact, there are more women than men affected) nor caused by a defect in a single gene. Why, then, is this family important, when the identification of the gene that causes this rare form of sex-linked cleft palate will help only a minute proportion of the total number of cases? It seems reasonable to assume that the far more common environmental causes of cleft palate, at present poorly understood, are via similar cellular mechanisms during embryonic development. Therefore, not only does the rare mendelian family suggest candidate genes which may be involved in the more common sporadic cases (determined genetically and environmentally), but it even indicates the route through which a purely environmental factor can cause a malformation. Moreover, because of the current advances in our ability to study gene expression during early development, it should be possible to determine not only which genes are defective but also the mechanism through which they act during embryogenesis (Akhurst 1986).

With the new generation of molecular biology tools outlined above, we already have a total human gene map, and will have complete chromosome-by-chromosome sequences for the human genome by the end of the century. This development neither can nor should be resisted. It will happen in any case, in order to help in the identification of mutations that cause single-gene diseases such as cystic fibrosis, and acquired conditions that involve a set of single mutations, as for cancer. However, the most exciting prospect is the possibility of understanding the interactions between several genes, and between genes and the environment, which lead to complex phenotypes. While this paper has been couched in terms of pathology, it will be equally relevant to any human variable. This will have implications for our understanding of the environment, the prevention of handicap, and the meaning and constraints upon variability – in total, a range of positive implications quite beyond what was imagined when gene isolation first became possible a decade ago.

### *Acknowledgements*

The work of the molecular genetics group at St Mary's Hospital Medical School has been supported for the past 10 years by the Medical Research Council, the Cystic Fibrosis Research Trust and several other generous grants from medical charities. This paper is based in part upon a lecture given by the author in Tokyo in May 1986.

## References

- Akhurst RJ 1986 The use of gene probes in studying human reproduction and embryology. *Hum Reprod* 1:213–219
- Cooper DN, Schmidtke J 1986 Diagnosis of genetic disease using recombinant DNA. *Hum Genet* 73:1–11
- Davies KE, Pearson PL, Harper PS, Murray JM, O'Brien T, Sarfarazi M, Williamson R 1983 Linkage analysis of two cloned DNA sequences flanking the Duchenne muscular dystrophy locus on the short arm of the human X chromosome. *Nucl Acids Res* 11:2303–2312
- Estivill X, Farrall M, Scambler PJ, Bell GM, Hawley KM, Lench NJ, Bates GP, Kruyer HC, Frederick PA, Stanier P, Watson EK, Williamson R, Wainwright BJ 1987 A candidate for the cystic fibrosis locus isolated by selection for methylation-free islands. *Nature (Lond)* 326:840–845
- Glover DM 1985 DNA cloning, a practical approach, Vol 1. IRL Press, Oxford
- Gusella JF, Tanzi RE, Anderson MA et al 1984 DNA markers for nervous system diseases. *Science (Wash DC)* 225:1320–1326
- Jeffreys AJ 1979 DNA sequence variants in the  $\alpha$ - $\gamma$ -,  $\beta$ - $\gamma$ -,  $\delta$ - and  $\beta$ -globin genes of man. *Cell* 18:1–10
- Kan YW, Dozy AM 1978 Polymorphism of DNA sequence adjacent to the human  $\beta$ -globin structural gene: relation to sickle mutation. *Proc Natl Acad Sci USA* 75:5631–5635
- Moore GE, Ivens A, Chambers J, Farrall M, Williamson R, Page DC, Bjornsson A, Arnason A, Jenson O 1987 Linkage of a cleft palate gene: a model for multifactorial genetic disorders. *Nature (Lond)* 326:91–92
- Solomon E, Bodmer WF 1979 Evolution of a sickle variant gene. *Lancet* 1:923
- Thein SL, Wallace RB 1986 The use of synthetic oligonucleotides as specific hybridisation probes in the diagnosis of genetic disorders. In: Davies KE (ed) *Human genetic disease*. IRL Press, Oxford, p 33–50
- Weatherall DJ, Clegg JB 1981 *The thalassaemia syndromes*, 3rd edn. Blackwell Scientific Publications, Oxford
- White R, Leppert M, Bishop DT et al 1985 Construction of linkage maps with DNA markers for human chromosomes. *Nature (Lond)* 313:101–105
- Williamson R 1987 The cystic fibrosis locus – a progress report. *Dis Markers* 5:59–63

## DISCUSSION

*Weatherall:* Could you outline your strategy for starting, *de novo*, to analyse a single-gene disorder with an unknown biochemical defect? What is the value, in such an analysis, of hypervariable (HVR) probes, or the minisatellite probes developed by Jeffreys et al (1985)?

*Williamson:* If there are known candidate genes for the disease, whether because of biochemical alterations or chromosomal aberrations, one would obviously start with those. However, if we assume there are no candidate genes or chromosomal regions, the first essential requirement is a set of families with

multiple affected members for linkage studies. Dominant and X-linked diseases are easier to study than recessives, since unaffected members contribute to linkage. For a recessive disorder, we normally require a minimum of 20 families, each with three affected sibs; unaffected sibs do not contribute to the analysis since they may be unaffected homozygotes or carriers.

We would then contact a protein chemist, such as Hans Eiberg of Copenhagen, who has facilities for analysing 60 different serum and red-cell protein polymorphisms in families. Like DNA probes, these markers are now mostly localized chromosomally. It was Eiberg et al (1985) who obtained the first linkage to cystic fibrosis. These protein markers are still of great value; the analysis is for the most part automated, and at least one obtains a lot of exclusion data.

I would then try to use DNA probes that are highly informative, and to exclude the genome, chromosome by chromosome. Many probes now recognize restriction fragment length polymorphisms (RFLPs), but a single probe, however informative, cannot exclude a large area of a chromosome because of statistical limitations. For this, one needs to use a set of linked probes of known order and interprobe distance, and to conduct the multipoint linkage by using the computer programs devised by Lathrop et al (1984). However informative a single probe, it cannot be as valuable as a set of ordered probes in chromosome exclusion. The hypervariable probes recognize many sequences on different chromosomes. It is sometimes possible to recognize segregation with one of these alleles, particularly for dominant inheritance. However, unlike the use of the single copy probe, this does not in itself give a chromosomal location; the particular band recognized by the hypervariable probe must now be cloned, and an adjacent single copy sequence must be used for mapping. Unfortunately, it is possible to find a linkage to one sequence recognized by Alec Jeffreys' probe (Jeffreys et al 1985), and then to spend many months trying to locate that band on the chromosome map.

*Lathrop:* CEPH (Centre d'Etude du Polymorphisme Humain) distributes DNA for linkage studies on a panel of 40 reference families. A large number of laboratories are now participating in typing genetic markers in these families. The CEPH panel is used because it allows the study of linkage between loci typed in different laboratories and, consequently, detailed genetic maps can be constructed much faster. Ray White's laboratory, at the Howard Hughes Medical Institute in Salt Lake City, has now characterized more than 200 DNA polymorphisms and classical genetic markers in the CEPH panel, and in an additional 20 reference families. We are currently constructing linkage maps of several chromosomes by using that database of genotypes. Dr Williamson has discussed the potential of hypervariable loci with a high degree of polymorphism for linkage studies. Y. Nakamura has been interested in developing such a set of genetic markers in Ray White's laboratory. Over 50 highly polymorphic probes that they have isolated are now being distributed through the ATCC

(American Type Culture Collection). In due course, many more highly polymorphic probes that they are currently characterizing will be made available to the scientific community. J-M. Lalouel and I have been collaborating with Ray White in an effort to use the Salt Lake City database to map new marker loci. We have found that approximately 70% of unmapped loci can be given a chromosomal assignment from linkage relationships with other markers in the database. Of course, this percentage will increase as the database expands. The highly polymorphic markers developed in Salt Lake City are being studied in the reference families. We hope that not only will the probes be available but so, too, will their localizations.

*Bobrow:* The problem of ordering probes, if one has a number of them placed together, has turned out to be a non-trivial exercise. Is linkage disequilibrium a practicable tool to use for helping to order the loci? Are any alternatives at our disposal?

*Kidd:* There are, indeed, problems with ordering loci, especially when different polymorphic loci are fairly close together on the chromosome. The problem can be reduced to one of sample size. Ordering loci along the chromosome requires identifying events that separate the loci—either crossovers in a linkage study or breakpoints in a somatic cell hybrid study. For loci close together, these events are, by definition, rare and large samples are required to find them. In addition, ordering loci in a linkage study requires that the crossover between two loci occur when both of those loci and a nearby third locus are all heterozygous. For many RFLPs the heterozygosity and polymorphism information content are so low that triply heterozygous individuals are very rare. This is why large collaborative projects are very important: they provide a large set of families and, ultimately, every crossover event in that set of pedigrees will be pinned down.

Linkage disequilibrium is a statistical phenomenon of a population, and is distinct from the genetic linkage that allows transmission of characters together in a family. In the population sample, disequilibrium exists when particular alleles at two or more separate loci are found together on the same chromosome more frequently than would be expected by chance alone. This phenomenon can have several causes, perhaps because there has been recent admixture between a population that was 'homozygous (+)' at both loci and one that was 'homozygous (-)': this would provide only  $+/+$  and  $-/-$  chromosomes. This disequilibrium will decay over generations, and will move towards equilibrium as a function of the recombination rate. If the loci are far apart on the chromosome, then after only two or three generations the loci will be in equilibrium. If the loci are very close together it may take hundreds of generations for equilibrium to be reached. Finding disequilibrium in the population is not proof, but tends to mean that the loci are very close: disequilibrium from past admixture or past mutational events has not been eliminated because the recombination rate is so low. Unfortunately, when one is dealing with very

small regions, chance events, such as individual mutational histories and individual recombination events, become almost as likely to increase as to decrease disequilibrium. One must also consider random genetic drift and historical accidents of admixture, both of which confound the relationship between genetic distance and disequilibrium. In such cases higher levels of disequilibrium do not necessarily correlate with the loci being closer; for very close loci one cannot order sites simply by saying that those loci with the highest disequilibrium are closest and those with less disequilibrium are less close.

*Bobrow:* Are there ways in which one can use disequilibrium as a crude tool for attempting to do relative ordering? Do you think that the problems of not being able to define mutational and immigrational histories are such that, for practical purposes, one can never say more than that disequilibrium suggests that the loci are fairly close? Could it ever be a finer tool than that?

*Kidd:* It may be a more useful tool in specific cases. In the albumin and  $\alpha$ -fetoprotein gene complex the disequilibrium does not follow the known order of the restriction site polymorphism (Murray et al 1984). In the  $\beta$ -haemoglobin cluster, because of the 'hot-spot' of recombination, there is a clear division into two groups which is quite valid, but within each group there is not such a clear ordering (Chakravarti et al 1984).

*Sing:* Linkage disequilibrium could well become a finer tool, Professor Bobrow. Eric Boerwinkle, Alan Templeton and I have been considering strategies that use information about linkage disequilibrium. We begin with the concept that Professor Kidd has just explained: the closer the sites are located to each other, the more important the mutation rate becomes. We have developed (for unrelated samples, not pedigree data) a method that is introduced in our paper, to be presented later at this symposium (p 99).

*Motulsky:* We have been interested in the apolipoprotein A-I-C-III locus where, within a distance of a few kilobases, two RFLP sites were in disequilibrium and one RFLP between these two appeared to be in equilibrium. Elizabeth Thompson (unpublished results) has done some statistical work on this phenomenon. The usual statistical parameters for the description of genetic equilibrium and disequilibrium do not have confidence limits. She found that in order to prove disequilibrium for the locus that seemingly appeared to be in equilibrium, one would need sample sizes that are so large that one can never obtain them in humans. The reports that show random order for equilibrium and disequilibrium of RFLPs within a few kilobases of each other need to be re-evaluated (Barker et al 1984, Litt & Jorde 1986).

*Rose:* As an epidemiologist I would be glad of help on a basic question concerning polygenic diseases. How do we know which those are? As I understand it, if we see a continuous distribution of phenotypes and if we see that inheritance is graded, it is often argued that the genetic determinants must be complicated. It has already been pointed out that one gene may undergo multiple mutations, producing different effects which may, in turn, be modifi-



able by an environmental factor that is graded over a wide range, and perhaps also by the whole genetic context. That would seem to produce quite sufficient opportunity for continuous phenotypic distributions and for graded inheritance on a monogenic basis. By what criteria, then, do we reject a simple single-gene explanation for any particular disease?

*Williamson:* Several diseases, such as thalassaemia intermedia, are known to be polygenic, as both the  $\alpha$ - and  $\beta$ -globin genes are involved. For many conditions, for example, coronary heart disease, we know of severe single-gene defects in *different* genes which lead to phenotypes that prompt us to investigate whether a set of minor defects, involving the same genes, would add together to give the sort of inheritance pattern that one sees in the population. Nevertheless, the term 'polygenic', or 'polygenetic', or 'multifactorial' disease is a useful blanket term for describing those diseases for which we know there is a genetic component that cannot be described in a simple mendelian way. It is interesting to consider the value of using pure reductionism to the extent that you propose. Our consideration of a particular family with cleft palate comes into that category (Moore et al 1987). We do not yet know whether Alzheimer's disease and psychotic depression will be best described by a model involving different mutations to one gene, rather than by the more traditional model of a multifactorial disease.

## References

- Barker D, Holme T, White R 1984 A locus on chromosome 11p with multiple restriction site polymorphisms. *Am J Hum Genet* 36:1159–1171
- Chakravarti A, Buetow KH, Antonarakis SE, Waber PG, Boehm CD 1984 Nonuniform recombination within the human beta-globulin gene cluster. *Am J Hum Genet* 36:1239–1258
- Eiberg H, Mohr J, Schmiegelow K, Nielsen LS, Williamson R 1985 Linkage relationships of paraoxonase (PON) with other markers: indication of PON–cystic fibrosis synteny. *Clin Genet* 28:265–271
- Jeffreys A, Wilson V, Thein SL 1985 Hypervariable 'minisatellite' regions in human DNA. *Nature (Lond)* 314:67–73
- Lathrop GM, Lalouel J-M, Julier C, Ott J 1984 Strategies for multilocus linkage analysis in humans. *Proc Natl Acad Sci USA* 81:3443–3446
- Litt M, Jorde LB 1986 Linkage disequilibria between pairs of loci within a highly polymorphic region of chromosome 2Q. *Am J Hum Genet* 39:166–178
- Moore GE, Ivens A, Chambers J, Farrall M, Williamson R, Page DC, Bjornsson A, Arnason A, Jensson O 1987 Linkage of a cleft palate gene: a model for multifactorial genetic disorders. *Nature (Lond)* 326: 91–92
- Murray JC, Mills KA, Demopoulos CM, Hornung S, Motulsky AG 1984 Linkage disequilibrium and evolutionary relationships of DNA variants (restriction enzyme fragment length polymorphisms) at the serum albumin locus. *Proc Natl Acad Sci USA* 81:3486–3490

# Genetics of coronary heart disease and its risk factors

Kåre Berg

*Institute of Medical Genetics, University of Oslo (and Department of Medical Genetics, City of Oslo), P.O. Box 1036, Blindern, 0315 Oslo 3, Norway\**

*Abstract.* Lipoprotein parameters related to coronary heart disease (CHD) exhibit impressive heritability, and several traditional genetic marker systems are associated with lipid levels or CHD. Recent studies indicate a population-attributable risk of 28% for myocardial infarction for men below age 60 in the top quartile of Lp(a) lipoprotein levels. Thus, a high level of Lp(a) lipoprotein emerges as a major genetic risk factor for premature CHD. Studies of DNA polymorphisms at apolipoprotein loci have uncovered associations with lipid levels and genetic linkage between DNA polymorphisms at the apolipoprotein B (apoB) locus and the Ag(x) antigenic polymorphism of low density lipoprotein. This finding proves that the Ag(x) antigenic variation resides in apoB and co-assigns its locus to chromosome 2. Lipid associations of the Ag(x) polymorphism and of DNA polymorphisms at the apoB locus are internally consistent and consistent with association between Ag(x) and DNA variants. A new approach to the study of gene-environment interactions, using monozygotic twin pairs makes it possible to uncover genes that contribute to the frame within which lifestyle factors can cause changes in a clinically relevant quantitative parameter such as serum cholesterol concentration. A new concept of interaction between 'level genes' and 'variability genes' in the aetiology of atherosclerosis emerges from these studies.

*1987 Molecular approaches to human polygenic disease. Wiley, Chichester (Ciba Foundation Symposium 130) p 14-33*

Several sets of data point to an important role for genetic factors in the aetiology of atherosclerotic disease, in particular premature coronary heart disease (CHD) (for review, see Berg 1983, 1985, 1986a). Clustering of relatively young CHD cases in families has been found by many workers. Studies in Finland have, in many families, revealed a pattern of aggregation of such cases that would agree with autosomal dominant inheritance. In an

---

\*This paper was prepared when the author was a Fogarty Scholar-in-Residence at the Fogarty International Center for Advanced Studies in the Health Sciences, National Institutes of Health, Bethesda, Maryland, USA.

American study a heritability of 0.56 was observed for CHD prior to the age of 55, even after exclusion of cases of monogenic hyperlipidaemia. Aggregation of young cases of myocardial infarction in families cannot be explained adequately by similarity within families with respect to traditional risk factors. Having had a first-degree relative who contracted premature CHD is by itself a significant risk factor.

### Heritability of lipoprotein parameters

Several lipoprotein parameters are known to be related to susceptibility or resistance to atherosclerotic disease. Some of them exhibit significant heritability. We have previously reported the results of studying 198 like-sexed twin pairs with respect to the serum levels of cholesterol, triglycerides, apolipoprotein B (apoB), apolipoprotein A-I (apoA-I) and apolipoprotein A-II (apoA-II) (Berg 1983). We found high heritability for apoB, apoA-I and apoA-II levels (Table 1). Recently, we have examined members of another 156 monozygotic (MZ) twin pairs with respect to the same parameters. Heritability of quantitative lipoprotein parameters in this new series, calculated as the intra-class correlation coefficient ( $r = h^2$ ) in MZ pairs, is included in Table 1. The high estimates for apoB, apoA-I and apoA-II are so similar in the two independent studies that they reject a hypothesis of no genetic influence on the parameters. Apoprotein levels may be more meaningful expressions of amounts of lipoprotein particles than are lipid levels.

**TABLE 1** Estimates of heritability ( $h^2$ ) of serum lipid and lipoprotein levels from two studies of twins<sup>a</sup>

Parameter	Heritability ( $h^2$ ) estimate	
	Study 1	Study 2
Cholesterol	0.34	0.68
Triglycerides (fasting)	0.40	0.46
ApoB level	0.66	0.64
ApoA-I level	0.53	0.55
ApoA-II level	0.69	0.68

<sup>a</sup> Study 1, a series of Norwegian twins: 98 monozygotic and 100 same-sexed dizygotic pairs; Study 2, a new series of 156 monozygotic twin pairs.

### The Lp(a) lipoprotein

The Lp(a) lipoprotein (Berg 1963) forms a distinct subpopulation of lipoprotein particles that, with respect to density, is intermediate between low density lipoprotein (LDL) and high density lipoprotein (HDL). Lp(a)

lipoprotein contains apoB, but less lipid and more protein than LDL. Its unique feature is the presence of Lp(a) antigen, detectable with suitable antisera.

Lp(a) lipoprotein level is under strict genetic control (for review, see Berg 1983). The most extensive genetic study that has been conducted in recent years was reported by Morton et al (1985) who examined 227 families with 557 children. Quantitative Lp(a) lipoprotein determination was conducted blindly and independently in two different laboratories. There was excellent agreement between the two sets of results. A bimodal distribution of the composite was found, and a major locus was strongly indicated. There was no evidence against mendelian transmission of the trait.

The first study that pointed to a strong association between high levels of Lp(a) lipoprotein and CHD was conducted in Scandinavia (Berg et al 1974). Significant correlations were found both between Lp(a) lipoprotein and the clinical manifestation of CHD and between this lipoprotein and the degree of atherosclerosis demonstrable by coronary angiography. Several studies have since confirmed the relationship between Lp(a) lipoprotein and early-onset CHD.

The most extensive study of Lp(a) lipoprotein in relation to CHD has been conducted in Hawaii (Rhoads et al 1986). In this study, Lp(a) lipoprotein was found to be independent of other lipid or lipoprotein parameters related to atherosclerosis, and the Lp(a) lipoprotein level was significantly higher in persons who had had myocardial infarction than in controls. This difference was particularly pronounced in people below the age of 60, but was significant also in persons aged 60–69 years. The population-attributable risk was about one in four myocardial infarctions among men in the highest quartile of Lp(a) lipoprotein levels, below age 60, and one in eight for men aged 60–69 years. Thus, Lp(a) lipoprotein emerges as a very important genetic risk factor for early-onset CHD. In view of the low level of Lp(a) lipoprotein compared to LDL even in people in the upper quartile, the inferred atherogenicity of Lp(a) lipoprotein is striking. The reason for this atherogenicity is not known. Apparently intact Lp(a) lipoprotein has, however, been demonstrated in atherosclerotic lesions, and this lipoprotein has a very strong capacity to form aggregates *in vitro*. For the time being it seems reasonable to assume that its involvement in atherogenesis is caused by its becoming easily trapped in the arterial wall.

The finding that Lp(a) lipoprotein is associated with early-onset CHD also in Japanese males in Hawaii (Rhoads et al 1986) indicates that this association is a ubiquitous phenomenon.

### **Associations between genetic markers and lipid levels or atherosclerosis**

Several genetic markers have exhibited an association with plasma lipid

**TABLE 2** Blood group and serum type systems that have exhibited an association with plasma lipid levels or atherosclerotic disease

System	Association with:	
	Lipid	Atherosclerotic disease
Blood groups		
AB0	Yes	
Secretor	Yes	
Serum types		
Gm (IgG <sup>a</sup> heavy $\alpha$ -chain)	Yes	
Hp (haptoglobin $\alpha$ -chain)	Yes	
Lipoproteins		
Ag (x)	Yes	
Lp (a)	Very weak	Yes
apoE	Yes	Yes

<sup>a</sup> IgG, immunoglobulin G.

levels, and two out of three lipoprotein polymorphisms are directly associated with atherosclerotic disease (Table 2).

Previous findings of an association with lipid levels may not necessarily identify a given gene as a 'genetic risk factor'. Thus, we were unable to confirm a significant effect on total serum cholesterol of the haptoglobin (Hp) polymorphism as reported by Sing & Orr (1976). However, we found a significantly increased frequency of people with haptoglobin genotype 2-2 among the persons in the highest HDL cholesterol quartile (Børresen et al 1986, Berg 1986b). Thus, the effect uncovered by Sing & Orr may have been caused by an association between Hp 2 and HDL cholesterol. The haptoglobin locus is on chromosome 16 and it is closely linked to the enzyme lecithin-cholesterol acyltransferase (LCAT; EC 2.3.1.43, also called phosphatidylcholine-sterol acyltransferase). If normal genetic polymorphism exists at the LCAT locus, different LCAT genotypes could possibly have an effect on the serum level of HDL cholesterol. Linkage disequilibrium between the LCAT and Hp loci could theoretically explain an apparent effect of Hp on the HDL cholesterol level (Berg 1986b). The effect of the Ag(x) phenotype on lipid levels is also relatively small and difficult to detect in samples of young people. It has, however, been found in several populations (Berg et al 1976). People with phenotype Ag(x-) have higher levels of cholesterol, as well as triglycerides, than Ag(x+) people. It is possible that this reflects an increased amount of LDL particles in the lower part of the density spectrum. No direct association between Ag(x) type and atherosclerotic disease has been demonstrated.

It is well established that the apolipoprotein E (apoE) isoforms apoE-2, apoE-3 and apoE-4 are determined by three alleles at one single locus on