PROGRESS IN MEDICAL GENETICS

New Series, Vol. 7

MOLECULAR GENETICS IN MEDICINE

Editors

Barton Childs • Neil A. Holtzman

Haig H. Kazazian, Jr. • David L. Valle

PROGRESS IN MEDICAL GENETICS
New Series, Vol. 7

MOLECULAR GENETICS IN MEDICINE

Editors

Barton Childs, MD
Professor Emeritus

Neil A. Holtzman, MD
Professor

Haig H. Kazazian, Jr., MD
Professor

David L. Valle, MD
Professor

Department of Pediatrics
Johns Hopkins University School of Medicine
Baltimore

Elsevier
New York • Amsterdam • London

Elsevier Science Publishing Co., Inc.
52 Vanderbilt Avenue, New York, New York 10017

Distributors outside the United States and Canada:

Elsevier Science Publishers B.V.
P.O. Box 211, 1000 AE Amsterdam, the Netherlands

Library of Congress Cataloging-in-Publication Data

Molecular genetics in medicine.

(Progress in medical genetics; v. 7)
Includes index
1. Medical genetics. 2. Molecular genetics.
3. Recombinant DNA. 4. Genetic disorders – Diagnosis.
I. Childs, Barton, II. Series. [DNLM: 1. DNA – analysis.
2. Genetics, Medical. 3. Hereditary Diseases – diagnosis.
W1 RP6709 v.7 / QZ 50 M7185]
RB155.P7 vol. 7 616'.042] s[616'.042] 87-22215

ISBN 0-444-01254-0

Current printing (last digit):
10 9 8 7 6 5 4 3 2 1

Manufactured in the United States of America

Preface

That the conventional medical aims of diagnosis, treatment, and prevention have been forwarded by technological advances is a commonplace. Our insights into the causes and pathogenesis of disease are also based on that same technology. In pursuing their aims, investigators have employed a reductionist strategy, moving from physiologic explanations to biochemical descriptions of pathways, enzymes, and other proteins, then to molecular descriptions of structure and function of cellular mechanisms, and finally to detailed analyses of the genes themselves.

The questions asked by the participants in these discoveries are mainly categorical: how is the genetic message transcribed and translated? how are the peptides aggregated to form enzymes or cell surface receptors? how do these proteins work? or, apropos of disease, what are the molecular mechanisms of the pathogenesis of, say, diabetes or rheumatoid arthritis? Individual variation in all these processes has not been a salient concern.

But the study of the molecular basis of individuality and variation, although commanding less attention, has capitalized on the same technological progress to provide some understanding of differences (including different diseases) between individuals. Now, recombinant DNA and allied techniques make it possible to examine the individuality of disease by detecting the mutants involved, characterizing them, locating them in the chromosomes, and studying their role in pathogenesis. The tempo of this research suggests an impending avalanche of such information about all sorts of diseases, not only those we call genetic—information physicians cannot afford to ignore. This book provides the reader with a review of the impact of recombinant DNA analysis on molecular genetics and shows how information derived thereby is likely to become an important basis for understand-

ing and discussion of all aspects of disease. Specifically, the authors examine the methods themselves and how they are being used to study the structures of the genes, the nature of mutations, and gene linkage and mapping. Information of this kind is useful in diagnosis as well as in prevention, at present in antenatal diagnosis, and potentially in screening for disease predisposition. New treatment strategies are discussed, as well as the participation of the new biotechnology companies in basic and applied research. The theme of the book is that categorical descriptions of disease processes are not enough; diseases afflict individuals individually. Recombinant DNA analysis is revealing how extensive this individuality is and how it makes a difference in the pursuit of medical aims.

The Editors

Contributors

Stylianos E. Antonarakis, MD
Associate Professor, Pediatrics, The Johns Hopkins University School of Medicine, Baltimore

Norman Arnheim, PhD
Professor of Molecular Biology, Department of Biological Sciences, University of Southern California, Los Angeles

Corinne D. Boehm, BS, MS
Assistant Professor, Pediatrics, The Johns Hopkins University School of Medicine, Baltimore

David Botstein, PhD
Professor of Genetics, Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts

Barton Childs, MD
Professor of Pediatrics, The Johns Hopkins University School of Medicine, Baltimore

Helen Donis-Keller, PhD
Senior Research Director, Department of Human Genetics, Collaborative Research, Inc., Bedford, Massachusetts

Henry A. Erlich, PhD
Senior Scientist and Director, Human Genetics Department, Cetus Corp., Emeryville, California

Neil A. Holtzman, MD, MPH
Senior Analyst, Office of Technology Assessment, Washington, D.C.

Haig H. Kazazian, Jr., MD
Professor of Pediatrics, The Johns Hopkins University School of Medicine, Baltimore

Grant A. Mitchell, MD
Research Fellow, Pediatrics, The Johns Hopkins University School of Medicine, Baltimore

Arno G. Motulsky, MD
Professor of Medicine, University of Washington, Seattle

Stuart H. Orkin, MD
Associate Professor, Pediatrics, Harvard Medical School, Division of Hematology, Childrens Hospital, Boston

John A. Phillips, III, MD
Professor of Pediatrics and Director, Division of Genetics, Vanderbilt University School of Medicine, Nashville, Tennessee

David L. Valle, MD
Professor of Pediatrics, Medicine and Molecular Biology and Genetics, The Johns Hopkins University School of Medicine, Baltimore

David A. Williams, MD
Assistant Professor of Pediatrics, Harvard Medical School, Division of Hematology/ Oncology, Childrens Hospital, Boston

Contents

CHAPTER 1

Introduction

Barton Childs, MD

> . . . If an equally close linkage were found between the genes determining blood group membership and that determining Huntington's Chorea we should be able, in many cases, to predict which children of an affected parent would develop the disease and to advise on the desirability, or otherwise, of their marriage. (Bell J Haldane JBS *Proc Roy Soc* 1936;123:119–150.)

It has become almost a cliché to speak of the "revolution" in molecular biology and genetics brought about by the use of recombianant DNA and allied techniques. Although the definition of the word revolution has proved more elusive than it once seemed, there is no doubt that these methods represent entirely new ways of investigating biologic problems. The scope and versatility of these techniques make it possible to test ideas previously open only to speculation, and there is something exponential in the rate at which ideas are tested and in the accumulation of information about each. Along with other aspects of biology, every medical field participates: immunology and infectious disease, endocrinology, oncology, and genetics, to name only some. And as for the conventional missions of medicine—the discovery of cause and pathogenesis, diagnosis, management, prognosis, and prevention of disease—each is served.

So there is general agreement that something quite new has been added to the technology of medical research, something with exceptional powers of inference and analysis. But the aims toward which the methods are used are not new; revolutionary means are being used to evolutionary ends. For example, among the genetic questions for which molecular strategies are providing answers are many that have been asked since the rediscovery of Mendel 87 years ago. The reference above is merely one example made timely by the current focus on Huntington's disease in consequence of the

discovery and mapping to chromosome 4 of restriction fragment length polymorphisms (RFLP) that segregate with the disease phenotype.[1] The point was made tellingly by Robson, who wrote an editorial on the occasion of the 50th year of publication of the Annals of Human Genetics entitled "Fifty years of human genetics: Plus ça change, plus c'est la même chose."[2] In the first volume of the Annals, just under one-third of the articles were concerned with new ways of doing linkage analysis, statistical maneuvers made necessary by the special qualities of human family data. In the 50th anniversary volume, a little more than half of the articles dealt with chromosome mapping based on linkage analysis, a strategy to which recombinant DNA methods have brought a new dimension of resolving power.[3] (See Chapter 2.)

Mendel's predecessors could not imagine how traits could be transmitted independently, and so they were unable to conceive of any relationship between qualities that were plainly inherited and any genetically transmitted particle. That was Mendel's contribution. The drosophilists, in turn, showed how segregation and independent assortment could be accounted for by the behavior of the chromosomes in meiosis. Linkage was first suggested by Bateson in 1902, and the details were later worked out by the drosophilists. By 1910 the genes of drosophila were localized to specific chromosomes, and by 1913 there was a rudimentary map showing the linear arrangement of several genes in the X chromosome. Later, in the 1930s, when the physical basis of heredity was well established, attention was turned to physiological genetics, or what the genes do, and a strategy was established that we still use in resolving, layer by layer, the affinities between concretely defined phenotypes and abstract genes (Fig. 1.1). Even after the identification of deoxyribonucleic acid (DNA) as the genetic material and the discovery of the genetic code, the gene remained an abstraction, defined in genetic experiments by its inferred properties as units of function, each divisible into smaller units of recombination and mutation. Estimates of the size of each of these units were provided by Benzer, who showed, as a result of matings of strains of the bacteriophage T_4, that the various units could be measured in angstroms, allowing so many base pairs for each; for example, the unit of mutation could be as small as a single base.[4] But it remained for recombinant DNA analysis to give molecular reality to the genes, to make them all but palpable, to make it possible to locate them and to define the functional units in sequences of base pairs in strands composed of exons, introns, regulatory elements, and flanking sequences. The extraordinary versatility and power of these molecular strategies resulted from the discovery of bacterial enzymes (restriction enzymes) that cleave DNA at sites specific to each enzyme.[5] These remarkable properties enable the investigator, after isolating DNA from test subject, to choose how to cut the strands and then to manipulate the products

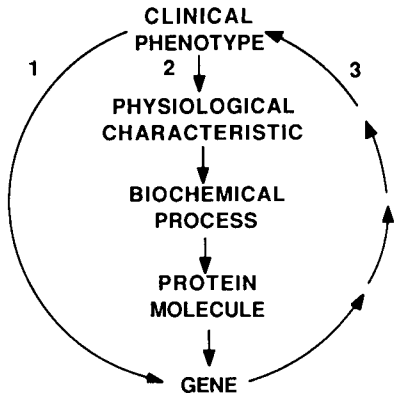

FIGURE 1.1 Strategies for relating phenotypes and genes. (1) Mendelian segregation reveals a relationship between a phenotype and a gene even when nothing is known of how the gene action is mediated. (2) The gene–phenotype relationship may be described at several levels, each a consequence of a previous one. (3) "Reverse genetics" is a mirror image of process 2. The structure of a gene may suggest a protein that may be found to have a biochemical function that plays a role in some homeostatic system accounting for the phenotype.

appropriately to test hypotheses that cannot otherwise be tested. For example, human genes can be transferred to other organisms where their functions can be studied under conditions not permissible in intact human beings. And in such experiments the person who is the source of the DNA, and whose genetic identity is to be defined with such precision, is merely a bystander, an observer who is thereby spared the intrusion of human clinical research.

WHAT DO THE METHODS DO?

General Applications

The methods have uses in all biologic investigations in which gene action and genetic variation are issues for study. Although only prokaryotic organisms produce the restriction enzymes upon which the techniques depend, the specific sequences they cleave are widely distributed. This means that the scope for the use of recombinant DNA techniques is limitless; obviously, only some aspects of the work are applicable to medicine. But medicine partakes of two kinds of information, some of general biological interest, some germane only to the pursuit of the practical aims of diagnosis and treatment. For example, mutation, polymorphism, and gene mapping have implications for all biology, but information generated in such investigations can be turned to practical account in tests for genetic heterogeneity, the

elucidation of pathogenesis, population screening, antenatal diagnosis, treatment, and prevention.

Mutation

Germ-line Mutation: Prior to the advent of recombinant DNA methods, our concepts of mutation were limited to what could be observed at the level of amino acid sequences; point mutations including missense, nonsense, and frameshifts were all known, as well as deletions and the consequences of unequal crossing over. "Regulatory" mutations were presumed to exist, of course, but their exact nature could not be imagined. But the ability to examine the DNA directly has added greatly to this list (see Chapter 3). Mutations that interfere with protein synthesis rather than amino acid sequence have been described; some involve the promoter or other regulatory elements, some the processes of RNA cleavage and splicing. Others, appearing in introns and flanking sequences, and so never translated, are both frequent and harmless. Furthermore, the fine structure of chromosomal mutations has been described. Deletions, translocations, and other aberrations too small to be seen with the light microscope but easily observed by restriction enzyme analysis have turned out to account for mendelizing phenotypes previously presumed to be due to single gene differences, but actually involving several contiguous genes.[6] Clearly, the adjective *mendelian* is indefinite as to what is segregating, an ambiguity the recombinant DNA methods may be expected to resolve. Also likely to yield to this kind of analysis is the identity of genes in the limited segments of chromosome 21 which, when trisomic, apparently account for Down's syndrome.

Somatic Mutation: There are mutations in somatic cells as well as in the germ line; DNA replication occurs in mitosis, so point mutations and chromosomal aberrations are to be expected. Recently, somatic mutation has been demonstrated; for example, it explains variation in immunoglobulins beyond that accountable to gene shuffling; point mutations have been found in the variable elements of both light and heavy chains.[7]

For many years somatic mutation has been invoked to account for malignant change. Mutagenic agents are usually carcinogenic too, and although there are familial cancers, even mendelizing forms, the age-dependent incidence of most cancers is most easily accounted for by somatic change in response to something in the environment.[8,9] But, apart from the observation of aneuploidy in many tumors and the translocated Philadelphia chromosome in chronic myeloid leukemia, it was not clear just what the mutagens did. Now, as a result of the new molecular methods we know that we all possess proto-oncogenes, genes that produce important regulators of

cellular replication and that in their nucleotide sequences bear close resemblance to genes in retroviruses that are involved in the cancer of some animals and are capable of malignant transformation of mammalian cells.[10,11] (See Chapter 5.) And it has been demonstrated that somatic mutations activate the proto-oncogenes by subverting their normal processes of regulation. Clearly, this conversion of proto-oncogenes to oncogenes is not the only step in the generation of malignancy, nor is it the only mechanism, but it does seem evident that the actions of carcinogens in producing malignancies are mediated by genes at several loci.[10,11]

Polymorphism

Restriction enzyme analysis has a major contribution to make in estimates of the extent of human genetic variation. Such estimates are essential for population genetics and for our understanding of evolutionary processes.[12,13] And, as things are turning out, the discovery of the genetic variants by which such estimates are made has significance for medicine as well; they are useful as markers of susceptibility to disease and as genetically linked surrogates for genes as yet unidentified. (See Chapters 2, 4, 5, and 10.)

Until recently, such markers consisted entirely of protein variants—enzymes, serum proteins, blood group substances, and the like. To be useful as such, the marker gene must be "polymorphic," that is, it must be relatively common in the population. So a locus is defined as polymorphic when there are two or more alleles, the least common of which exists at a frequency of 1% or more. This somewhat arbitrary figure was chosen because it exceeds that which could be accounted for by mutation alone. That is, an appeal must be made to natural selection, random drift, founder effect, or some other cause to explain the presence of genes at such frequencies. Studies of electrophoretic migration of proteins of many organisms, including man, have revealed that about 30% of the loci qualify by this definition as polymorphic.[12,13] A determined search will turn up mutants in lower frequencies also, perhaps 0.5 percent or less, at most, if not all loci, and such mutants certainly contribute some share of the variation, perhaps more than we know. Each of us possesses an unknown number of them; but it is the 30% of polymorphic loci that furnish the bulk of this kind of genetic variation exhibited by individuals. The remaining 70% of the loci that, apart from rare mutants, are invariant, determine the qualities that distinguish the species—in our case our humanness. They supply the background upon which the polymorphic alleles stamp that variable effects.

The degree of polymorphism varies from one locus to another. Some have a principal allele with a frequency of nearly 90%, with one or two others at a few percent each; obviously, most people would be homozygous for the principal allele and, therefore, uninformative with regard to differences

within families. At the opposite extreme are the HLA loci, each with numerous alleles in frequencies such that nearly everyone is heterozygous.[14] The ABO and other blood groups, some complement loci and a few others lie in between these limits. Obviously, the more polymorphic the locus, the more useful it will be in providing marker alleles; that is, the more likely it will be to be useful in distinguishing individuals in both populations and families.

Marker genes are of two kinds: (1) those associated with the disease, whether in cause or pathogenesis, and (2) those that are genetically linked but functionally independent, that is, they are situated in the chromosome so near to the gene in question as seldom to be parted by recombination.[15] Indeed, a linked gene may act as a stand-in for a gene of unknown location or even of unknown function, and such a surrogate could be of value in diagnosis, perhaps especially in the antenatal diagnosis of conditions of unknown pathogenesis. But as long as we were dependent upon protein polymorphism, progress in the discovery of linkages and associations, to say nothing of their use in medicine, was disappointing; although there were some successes, particularly the association of HLA alleles with autoimmune diseases. But this limitation should be no surprise; phenotypic variation, even at the level of protein difference, is likely to be subject to selective constraint. First, there is that 70% of loci that are invariant, except for the rare mutants that are not useful as markers. Then the degree of variation for most of the polymorphic loci is just not enough to make many families informative. What is needed to make the use of linkage for diagnosis prosper is an exuberant variation, less constrained by selection. And that is just what has been exposed by restriction-enzyme analysis of the DNA; a restriction fragment length polymorphism (RFLP) is encountered on an average of once in every few hundred bases. Further, they appear mainly in the noncoding parts of the DNA, the introns and flanking sequences, and often in such population frequencies as to make them very likely to distinguish individuals in families. So they have increased enormously the potential use of such markers.

Mapping the Chromosomes

The details of how DNA polymorphisms are discovered and how they are used in linkage analysis and in mapping of human chromosomes are given in Chapter 2. It is enough here to point out the virtues of such maps for medical use.[16] First, the mapping is proceeding at a furious pace; the number of mapped genes and arbitrary DNA segments doubles about every 2 years, having reached a total of nearly 1,500 by 1985.[17] And there is no reason to suppose that the pace will slacken. Indeed, it may accelerate, because the greater the density of the map, the more easily is a new gene or fragment to be mapped. In the end, it is anticipated that there will be enough markers spaced

throughout the whole genome to facilitate quick location of any gene or fragment.[16]

Medical uses are obvious. A marker may be used to make a diagnosis in cases where the identity of the disease is in doubt or to strengthen a probable diagnosis, or merely to stand as one among other evidences for or against. More often they are likely to be used as evidence of susceptibility; a signal of some homeostatic vulnerability predisposing to a disease commonly precipitated by some experience. In any of these instances the presence of the marker is seldom *proof* of the presence of disease or even of susceptibility. Although the sib of a patient with Huntington's disease who shares the marker for that disorder may have a high probability of having it, for other diseases the pathogenesis may be too complex to be initiated by a single gene.

Reverse Genetics

No one is ready to stop with the detection of RFLP markers, however useful they may be. Such a discovery is merely a preliminary to efforts to find the relevant gene itself. For purely diagnostic purposes a very tightly linked RFLP may serve the purpose, but there are rewards for finding the gene itself.[6,18,19] That is, there may be ways to explain the phenotype in reverse order, inferring a protein from the base sequence of the gene and then drawing conclusions as to the role of the latter in homeostasis, perhaps by comparing the sequence of its amino acids with those of proteins of known function. The point of the exercise, after discovering the role of the gene in cause or pathogenesis, is to devise some strategy to reverse or to neutralize it. To attempt to direct treatment according to cause is conventional medical practice. What is unprecedented here is the idea of a genetic analysis that proceeds from the bottom up, rather than from the top down. (See Chapters 4 and 5.)

The process seems to close a circle, the first arc of which was described by Mendel. That is, reverse genetics is a mirror image of Mendelian analysis. In the latter, a phenotype is explained by physiological, biochemical, and molecular accounts of gene action (Fig. 1.1). But the properties of phenotypes often fail to give direction to such an effort, so there remain many phenotypes, for example, most of the disease entries in *Mendelian Inheritance in Man,* that are undistinguished by attributes suggestive of the mechanisms of the genes from which they originate. Recombinant DNA methods will expose the genes for many such phenotypes so that physiological explanations can be pursued in the opposite direction, that is, from the gene to a protein and on to the clinical phenotype (Fig. 1.1). But, in addition, genes will be discovered in the absence of phenotype, so we will have genes in search of phenotypes as well as phenotypes in search of genes.