Annual Review of Genomics and Human Genetics Volume 1, 2000

ANNUAL REVIEW OF GENOMICS AND HUMAN GENETICS

VOLUME 1, 2000

ERIC LANDER, *Editor*
Whitehead Institute for Biomedical Research,
Massachusetts Institute of Technology

DAVID PAGE, *Associate Editor*
Howard Hughes Medical Institute,
Massachusetts Institute of Technology

RICHARD LIFTON, *Associate Editor*
Yale University

www.AnnualReviews.org science@annurev.org 650-493-4400

ANNUAL REVIEWS
4139 El Camino Way • P.O. BOX 10139 • Palo Alto, California 94303-0139

ANNUAL REVIEWS
Palo Alto, California, USA

International Standard Serial Number: 1527-8204
International Standard Book Number: 0-8243-3701-8

Annual Review and publication titles are registered trademarks of Annual Reviews.
⊗ The paper used in this publication meets the minimum requirements of American National Standards for Information Sciences—Permanence of Paper for Printed Library Materials. ANSI Z39.48-1992.

TYPESET BY TECHBOOKS, FAIRFAX, VA
PRINTED AND BOUND IN THE UNITED STATES OF AMERICA

Annual Review of Genomics and Human Genetics

Production Editors: NANCY P. DONHAM, TERI HYDE
Quality Control Coordinator: MARY A. GLASS
Subject Indexer: MARY GLASS

PREFACE

Biology's greatest achievement during the 20th Century was the elucidation of the mechanism of heredity–a tale of extraordinary and explosive scientific progress. The century opened with a paper published in January 1900, announcing the "rediscovery" of Mendel's laws governing inheritance. By the quarter-century, biologists had discovered the cellular basis of heredity: the chromosomes. By the half-century, we had revealed the molecular basis of heredity: DNA. By three-quarters of the way through the century, we had uncovered the information nature of the heredity—understanding the genetic code and developing recombinant DNA technology to read and manipulate DNA sequences. By the end of the century, we had read the sequence of entire genomes. Today we stand on the verge of having the complete sequence of the human genome.

What began as fundamental curiosity about the resemblance between parent and child ended up sparking a scientific program of relentless energy that is now providing us with Biology's equivalent of the Chemistry's Periodic Table—a comprehensive description of the elements of biology, in terms of which other cellular phenomena must be explained. The consequences for biological and biomedical research will be far-reaching.

The developments have had two dramatic effects on biology. First, they have spawned the new field of Genomics. Genomics seeks to understand biology by taking comprehensive, global views—studying the complete gene content of an organism, the entire symphony of gene expression during a cellular response, the full collection of protein-protein interactions, the repertoire of common human genetic variants. Genomics seeks nothing less than to understand the basis of cellular, developmental, and physiological circuitry by taking a comprehensive and integrative view of the parts. Genomics is quintessentially interdisciplinary–drawing on biology, biochemistry, engineering, mathematics, and computer science. Genomics is intimately connected with evolutionary studies in that it seeks to read and interpret the laboratory notebook of evolution's (successful) experiments over the past 3.5 billion years. In short, the field has no modest ambitions.

Second, it has given new life to human genetics. Human genetics has always posed a much more difficult challenge than genetic studies of experimental organisms that can be bred in the laboratory. Genomics has begun to remove many of these limitations. It has already become possible to study human diseases by extracting information from existing human family pedigrees, and it is rapidly becoming possible to extend this approach to studying the entire human population as if it were a single large pedigree, to directly study human tissue samples using global views of the complete DNA, RNA, and protein inventory. Nowhere will

the impact of genomics be more strongly felt than in human genetics. The result will be a deep understanding of the genetic contributions to human disease, with important implications for medicine and for society.

It is fitting to start the 21st Century by inaugurating a new Annual Reviews series aimed at distilling the insights from these two intertwined and important fields. We aim to cover these fields in the broadest sense—including within our mandate the biological, chemical, technological, computational, and social issues related to the fields. We are confident that the story of the century ahead will prove even more remarkable than that of the century past.

Eric S. Lander

RELATED ARTICLES

From the ***Annual Review of Biophysics and Biomolecular Structure***, Volume 29 (2000)

Comparative Protein Structure Modeling of Genes and Genomes, Marc A. Martí-Renom, Ashley C. Stuart, András Fiser, Roberto Sánchez, Francisco Melo, and Andrej Šali

From the ***Annual Review of Genetics***, Volume 33 (1999)

Molecular Genetics of Human Retinal Disease, Amir Rattner, Hui Sun, and Jeremy Nathans

Mitochondrial Genome Evolution and the Origin of Eukaryotes, B. Franz Lang, Michael W. Gray, and Gertraud Burger

Family Values in the Age of Genomics: Comparative Analyses of Temperate Bacteriophage HK022, Robert A. Weisberg, Max E. Gottesman, Roger W. Hendrix, and John W. Little

From the ***Annual Review of Immunology***, Volume 18 (2000)

Genomic Views of the Immune System, Louis M. Staudt and Patrick O. Brown

From the ***Annual Review of Medicine***, Volume 51 (2000)

Gene Therapy for Adenosine Deaminase Deficiency, Robertson Parkman, Kenneth Weinberg, Gay Crooks, Jan Nolta, Neena Kapoor, and Donald Kohn

Genetic Disorders Affecting Proteins of Iron Metabolism: Clinical Implications, Sujit Sheth and Gary M. Brittenham

Genetics of Psychiatric Disease, Wade H. Berrettini

The Genetics of the Amyloidoses, Joel N. Buxbaum and Clement E. Tagoe

From the ***Annual Review of Microbiology***, Volume 54 (2000)

Interim Report on Genomics of Escherichia coli, Monica Riley and Margarethe Serres

From the ***Annual Review of Neuroscience***, Volume 23 (2000)

Pain Genes? Natural Variation and Transgenic Mutants, Jeffrey S. Mogil, Lei Yu, and Allan I. Basbaum

ANNUAL REVIEWS is a nonprofit scientific publisher established to promote the advancement of the sciences. Beginning in 1932 with the *Annual Review of Biochemistry*, the Company has pursued as its principal function the publication of high-quality, reasonably priced *Annual Review* volumes. The volumes are organized by Editors and Editorial Committees who invite qualified authors to contribute critical articles reviewing significant developments within each major discipline. The Editor-in-Chief invites those interested in serving as future Editorial Committee members to communicate directly with him. Annual Reviews is administered by a Board of Directors, whose members serve without compensation.

Annual Reviews of

Anthropology
Astronomy and Astrophysics
Biochemistry
Biomedical Engineering
Biophysics and Biomolecular Structure
Cell and Developmental Biology
Earth and Planetary Sciences
Ecology and Systematics
Energy and the Environment
Entomology
Fluid Mechanics
Genetics
Genomics and Human Genetics
Immunology
Materials Science
Medicine
Microbiology
Neuroscience
Nuclear and Particle Science
Nutrition
Pharmacology and Toxicology
Physical Chemistry
Physiology
Phytopathology
Plant Physiology and Plant Molecular Biology
Political Science
Psychology
Public Health
Sociology

SPECIAL PUBLICATIONS

Excitement and Fascination of Science, Vols. 1, 2, 3, and 4

CONTENTS

Annu. Rev. Genomics Hum. Genet. 2000. 01:1–19

GENETICS, BIOLOGY AND DISEASE

Barton Childs and David Valle
Department of Pediatrics, The Johns Hopkins University School of Medicine, Baltimore, Maryland 21205; dvalle@jhmi.edu

INTRODUCTION

Medicine is always faced with decisions in its perennial struggle to define and renew itself. A question of today is how best to adapt to the deluge of biological information accumulating at exponential rates. One way is to continue the reduction of diseases to molecules, extending the number of specialties and the penetration of the vision of each. This way risks reducing the doctor-patient relationship to, on one side, an artisan who practices by algorithm, and on the other, a patient reduced to molecules. Another way differs, not in failing to grasp the significance of reductionism, but in seeking the principles of Disease that give coherence to diseases, producing a medical biology with its roots deeply implanted in the parent science. Actually there is no conflict between these seeming alternatives; the second easily accommodates the first and is more conducive to a mutually gratifying patient care and more compatible with medicine's traditional acceptance as a university subject. Further, since the principles of Disease must be biological and compatible with current biological thought, their elaboration should promote the already existing convergence of interests of biological and medical investigators with particular benefit to medical education. In encouraging this convergence the *Annual Review of Genomics and Human Genetics* will play a useful part.

THE ANNUAL REVIEW OF GENOMICS AND HUMAN GENETICS

The evolution of *Annual Reviews* reflects the advances in the sciences they summarize, not only in the prevailing concepts and facts, but in general acceptance and timeliness. The first of these summaries was the *Annual Review of Biochemistry* (1932), and examination of volume 1 reveals a preoccupation with intermediary metabolism. The 1930s were a time when medical attention had shifted from the organ to the cell and the biochemical characterization of its physiology. So the new *Annual Review* was as timely for medicine as for biochemistry itself.

Such relationships to current conceptual developments were also observed in the start of the *Annual Review of Genetics* (*ARG*) beginning in 1967. That medicine

1527-8204/00/0728-0001$14.00

would be surveyed for its interest in genetics was clear from the start; the first article of volume 1 was a paper on human biochemical genetics, an account of new insights stemming from the Beadle-Tatum observation of the one gene–one enzyme principle. In the 32 volumes to date, there have been 85 papers on human genetics (15% of the total), and these are distributed evenly throughout. In contrast, the contents of the *Annual Review of Medicine* (*ARM*) (1949) has been more reflective of the growth of interest in the participation of heredity in disease, most notably the inborn errors. For example, in the first 11 volumes (up to 1960), there were no articles with genetic content. So while the perception of genetics in medicine was on the rise, that of medicine in genetics remained constant.

No surprise there. Who can doubt that the reviews of medical interest that appeared in *ARG* helped to promote the movement of genetics into medicine? A look at the titles of the genetic papers in *ARM* and the medical papers of *ARG* reveals in both a trend away from biochemical genetics to molecular interpretation. This shift attracted the attention of the editors of *ARG*, who in volume 20 (1989) made the following comment: "For some years, genetics and geneticists have faced occasional identity crises. From one viewpoint genetics ended with Watson and Crick, to be replaced by molecular biology; from another, molecular biology serves to define more precisely and answer more definitively the questions that geneticists were already asking" (20). In defending their position the editors might have cited Sewall Wright's claim for genetics as "the rootstock of biology" (55). But of course the editors have been resolute in adhering to their position, knowing that a genetic explanation always includes variation, while a molecular explanation need do no such thing.

The decision to publish the *Annual Review of Genomics and Human Genetics* will do much to resolve this misapprehension. The title is plain enough. Genomics has to do with the identification and characterization of genes and their arrangement in chromosomes, while human genetics is devoted to understanding the origin and expression of human individual uniqueness. This juxtaposition of genomics with human genetics has a special significance for medicine. Genomics is a strongly comparative study dedicated to advancing the unity of the biology of all organisms, while medicine is only just emerging from a state of autonomy in which it took from other sciences whatever it saw as useful for a parochial biology of medicine. The logic of the latter is embraced in the metaphor of the body as a machine that breaks from time to time and needs fixing. But that the machine derives from an evolutionary past that determines why it breaks is perceived, if perceived at all, to be irrelevant. Genomics and human genetics will help us to formulate and to answer such "why" questions as well as those that begin with "how" (42).

A further impetus to publish this new *Annual Review* might have been the recognition that medical genetics is going beyond the inborn errors to disorders of complex origin, diseases that include not only variation in more than one gene and their products, but also diversity in development, maturation, and aging, themselves shaped by variable experiences of the environment acting within an everchanging physiological matrix composed of variable gene products. In the elucidation

of diseases of complex disorders, molecular biology has been indispensable, but it is genomics that gives direction to molecular explanations (5, 36, 37). In fact, genomics is heir to the thinking of R.A. Fisher, who in 1918 concluded that continuous variation was no less due to the effects of single genes than monogenic phenotypes (21).

GENETICS IN MEDICINE

For the past 50 years genetics has been moving into medicine, at first timidly, now aggressively. Prior to 1950, genetics in medicine relied on pedigree analysis. It could not have been otherwise. Every step in the entry of genetics into medicine has been mediated by a further elaboration of the definition of the gene, and before the 1950s the gene, which had been defined operationally by the Drosophilists, was still an abstraction with no known function. It was not until Beadle and Tatum gave the gene a functional definition in their demonstration of the one-to-one relationship of gene and enzyme that biochemical genetics got its start (4). Then it was possible to complete Garrod's description of the inborn error, which, although inclusive of hereditary enzyme deficiency and accumulated intermediate, lacked reference to any gene. But, based on the Beadle-Tatum concept, the description of inborn errors took off on an exponential trajectory from which it has never departed. The discovery, also in the 1950s, of chromosomal anomalies, together with formal recognition of biochemical genetics as a legitimate pursuit, led to the establishment of a medical genetics enterprise, with careers to make in medical school divisions and departments. So by the late 1960s when volume 1 of *ARG* appeared, genetics was firmly embedded in medicine. But it had not changed medical thinking; mutant genes and chromosomal anomalies had simply joined microbes and toxic substances as proximate causes of disease.

In the meantime, as the double helix, its code, and all its works preoccupied geneticists, the gene was given a structural definition manifest in the colinearity between base pairs of DNA and amino acids in its protein product (56). This demonstration of molecular intimacy between DNA and protein initiated a subtle shift in emphasis away from the gene in the direction of its variant protein products, which soon became the center of attention of the effort to unscramble the pathogenesis of the inborn errors. In time, the inborn errors came to be perceived as abnormalities not only of metabolism but of all of the homeostatic devices that maintain the integrated open system.

GENETICS IN MEDICINE TODAY

Molecularization proceeded apace with a new, more complex molecular definition of the gene as a unit of transcribable DNA with some flanking nontranscribed controlling elements. A spin-off in medicine was the use in the 1970s of RFLPs to

track mutants in monogenic diseases; among the first was the antenatal diagnosis of abnormal hemoglobins (33). And with the advent of PCR, human biochemical genetics was eclipsed by molecular genetics, and inborn errors were now defined by specific molecular mutants. Then, beginning in the late 1980s, with genomics well established, attention could be given to multifactorial disorders, now called diseases of complex origin, with expectation of success in finding genes whose products would figure prominently in pathogenesis. Now, early in the twenty-first century, arrays of such genes, or markers of genes, are available for diabetes, hypertension, and other diseases (38, 50). In time, the gene products and the homeostatic systems to which they belong will be identified and the participation of their variants in pathogenesis will be characterized. Further, we shall learn how many of which alleles, derived from how many loci and in how many different combinations, are needed to produce the same disease in different patients, as well as just how the effects of the gene products interact in nonlinear ways to produce the variations in clinical expression. Increasingly, in the mind's eye, we will see the variant proteins in their functional contexts, rather than the genes, and our vision of the pathogenesis of the complex disorders will join that of the monogenic diseases in being understood in terms of variations in the proteins that constitute the feedback loops, circuits, cascades, and pathways already well known to biochemistry and physiology (8, 25). It is not that the genes will be in any way downgraded. It is simply that the elucidation of pathogenesis requires understanding of proteins, not genes.

CHANGES IN MEDICAL THINKING

There is no question that molecular genetics has changed medicine; not least, it has made genetics familiar to all. But it has changed our thinking too, and in at least two ways. The first is in introducing the idea that if we are to understand any disease, it will be through recognition of variations in the elements of the biochemical and physiological apparatus of the cell, and those elements are protein products of genes. This idea that most, perhaps all, disease is somehow genetic has been around for a long time, even in the nineteenth century when it was called "diathesis," but until genetics came into medicine, it was an idea without specificity (24, 45). But the concept of genetic variations in proteins transcends disease to include the uniqueness of all individuals, human and otherwise, so the second profound change in medical thinking is that human genetic uniqueness is expressed no less in the diseases we experience and in their clinical variation, than in, say, our appearance, wherein hereditary variation has been accepted for centuries. Medical thought tends to be typological, emphasizing the central tendency in which the characteristics of each case are measured against those of the classical case, but genetics exemplifies what Ernst Mayr calls "population thinking," in which all populations are perceived to be composed of unique individuals (41). Although we will continue to honor the time-encrusted concept of the body as a machine, we

are now beginning to recognize that each machine is one of a kind and that each breaks in its own specific way. But these changes in medical thinking, however profound, are still far from complete in both concept and dissemination. For example, the idea of genetic individuality is not yet a staple of medical education. Nor have we reconciled epidemiological concepts like risk factors and evidence-based medicine with the idea that medicine is first of all for individual patients and then for populations (39, 51). Even so, it would not be unreasonable to claim that the genetic viewpoint has begun to pervade medicine.

How to advance this cause? One way is to bring up to date previous efforts to elaborate a conceptual basis for disease; a study of Disease, not diseases. We shall be greatly helped in discovering these principles by the product of the Human Genome Project. There should be ways to conjure with it to extract the generalizations that characterize all the diseases of our species.

THE HUMAN GENOME PROJECT AND MEDICINE

Today we are witnessing the Human Genome Project (HGP) in full spate. Now extended to dozens of organisms, it was its benefits to medicine that convinced Congress to invest in it, and we are on the threshold of discovering what those benefits are (9, 15–17).

That a principal aim of the project is the advancement of medical aims should cause those of us in medicine to reflect. How is it that knowing all human genes can give us insights into how our patients die of heart attacks, experience inflammatory bowel disease, or have chicken pox?

First, there is something of a paradox here. The Human Genome Project is a natural next step in the history of the examination of the human body for medical purposes. More than a hundred years ago physical diagnosis with attention to organs and organ systems was enhanced by morbid anatomy observed at autopsy. Then in the early twentieth century, attention shifted to the cell, first its structure, then its biochemistry, and then its molecular properties and their participation in pathogenesis. And gradually over the past 50 years, the genes have made themselves known, both as agents of disease and as determiners of the specificity of the molecules that are the engines of the cell and the central feature in pathogenesis. The next logical step in this descent is to create a complete list of the genes whose variant protein products instigate all disease. This step was taken not by physicians, but by biologists, who with sublime confidence envisioned the HGP and set it in motion, engaging the interest of ingenious innovators who designed the technological means to make it work (18). It was a biological project which, in the fashion of biology, would seek generalizations about how things work to attain congruence of open systems with their environments. But it was noted that in time, the project would call attention to states of incongruence that predispose to disease too. The benefits to medicine are of two kinds: (*a*) The HGP will advance conventional medical aims, and (*b*) it will be helpful in the study of Disease as opposed to diseases.

CONVENTIONAL MEDICAL AIMS

Medical aims are diagnosis, treatment, prognosis, and prevention. Benefits of the HGP compatible with these ends are easy to suggest. Accepting the validity of the metaphor of the body as a machine that breaks, the question for each disease is how and where did the machine break, and how can it be fixed? So diagnosis and treatment will be helped immeasurably by the discovery of genes relevant to each disease. Their variant protein products are the principal instruments of pathogenesis, and attempts will be made to design treatments to nullify the adverse effects of each (9, 15–17). So we may expect progress limited only by the number of investigators, their ingenuity and persistence, and their financial backing. Current attention is on the genetic origin of complex diseases. These are variably familial but nonsegregating, and their genes are exposed by genomics. Success in the application of these methods to the study of diabetes and hypertension is measured by the discovery of numerous genes whose variant products contribute to the phenotype (38, 50). Final answers will differ from monogenic disorders in that no single gene product is so prominent in pathogenesis as to allow others to be ignored. No doubt there will be difficulties in detecting how the gene products participate in pathogenesis, as well as in describing their salience and participation, each in relation to others, and their integration in biochemical and physiological systems that may increase, or compensate for, their deviance. Still, whatever the snags, medicine's conventional aims will prosper mightily. Perhaps the treatments will attain that "high technology" that Lewis Thomas yearned for in the 1970s: simple and effective medications designed for precisely defined targets and with few side effects (49).

CONCEPTS OF DISEASE

The second benefit of the HGP for medicine is that of enhancing the study of Disease as opposed to that of diseases. Each disease exhibits its own qualities that enable us to diagnose and classify it. But in attaining that distinctive state, even while traducing the rules of congruence, it was constrained by them. That is, the integrated whole responds, but always within the limits of the powers expressed in its own individual version of congruence. So since we are all human beings, there are generalizations of incongruence and disease, no less than of congruence and health. The question asked in this section is: How will the HGP contribute to concepts of Disease, to generalizations that provide a conceptual infrastructure for practical actions taken in medicine and that can help a medical student in integrating the preclinical teaching with that of the clinical years? We all have such conceptual bases for everything we do; legal, religious, ethical—even for riding a bicycle. Medicine has them too—the body as a machine is one—and they have been the object of study by philosophers among others (13, 48). But the elaboration of concepts of Disease was, in the past, much handicapped by a lack