GENES AND THE BIOLOGY OF CANCER

Harold Varmus
Robert A. Weinberg

SCIENTIFIC AMERICAN LIBRARY

A division of HPHLP
New York

Library of Congress Cataloging-in-Publication Data

Varmus, Harold
Genes and the biology of cancer / Harold Varmus, Robert A. Weinberg.
p. cm.
Includes bibliographical references and index.
ISBN 0-7167-5037-6
1. Carcinogenesis. 2. Cancer cells. 3. Oncogenes. I. Weinberg, Robert A. (Robert Allan), 1942– . II. Title.
[DNLM: 1. Cell Transformation, Neoplastic. 2. Neoplasms—etiology. 3. Oncogenes. QZ 202 V316c]
RC268.5.V37 1993
DNLM/DLC
for Library of Congress 92-20554
CIP

ISSN 1040-3213

Printed in the United States of America.

Scientific American Library
A division of HPHLP
New York

Distributed by W. H. Freeman and Company.
41 Madison Avenue, New York, New York 10010 and
20 Beaumont Street, Oxford OXI 2NQ, England

2 3 4 5 6 7 8 9 0 KP 9 9 8 7 6 5 4 3

This book is number 42 of a series.

CONTENTS

Preface *vi*

1 Genes, Cells, and Complex Organisms *1*

2 The Nature of Cancer *25*

3 Clues to the Origin of Cancer *45*

4 The Genetic Elements Governing Cancer: Proto-Oncogenes *67*

5 The Genetic Elements Governing Cancer: Tumor Suppressor Genes *101*

6 Distortion of Growth Control in Cancer Cells *121*

7 Multistep Carcinogenesis *155*

8 Cancer Genes in the Clinic *185*

Sources of Illustrations *208*

Index *210*

PREFACE

Recent discoveries have revolutionized the understanding of cancer. Our purpose in this book has been to show the motivated but non-expert reader how this understanding is structured and how these discoveries have emerged.

In large part, progress against cancer can be attributed to fundamental changes in the way organisms and diseases are now studied. Traditionally descriptive, biology has been transformed in the twentieth century into a science capable of explaining complex phenomena like cancer. This new power flows largely from a comprehension of genetics in molecular terms that descends both from Gregor Mendel's conception of the gene as the unit of heredity and from James D. Watson and Francis H.C. Crick's identification of genetic material as a double helix of DNA. But modern biology also depends upon biochemical methods for purifying molecules and characterizing enzymes; upon procedures for growing cells and viruses under controlled conditions; and upon physical techniques for describing molecules at the atomic level. All these approaches have had major roles in developing the radically new picture of cancer that we present in the pages that follow.

Even twenty years ago, ideas about the origins of cancer were obscure and conflicting. We now have a view of the genetic basis of cancer that, while still incomplete, is not only satisfyingly coherent but also remarkably precise in molecular detail. The essence of the story is disarmingly simple. Under normal conditions, the growth of each animal cell is exquisitely controlled—stimulated or retarded, according to need—by inherited (genetic) mechanisms. Cancer occurs when gene altera-

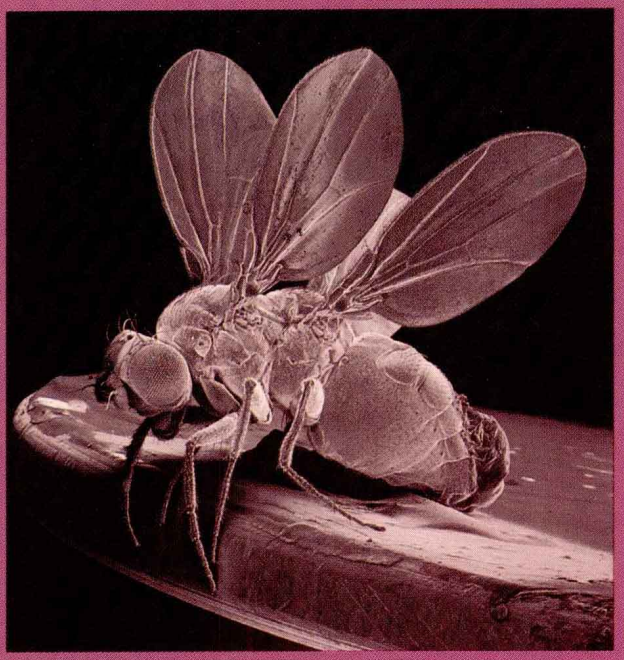
tions (mutations) distort the normal controls in an individual cell, prompting it to inappropriate, ultimately invasive, growth. Such mutations can act directly, provoking cell growth, or indirectly, by impairing the means to restrict cell division.

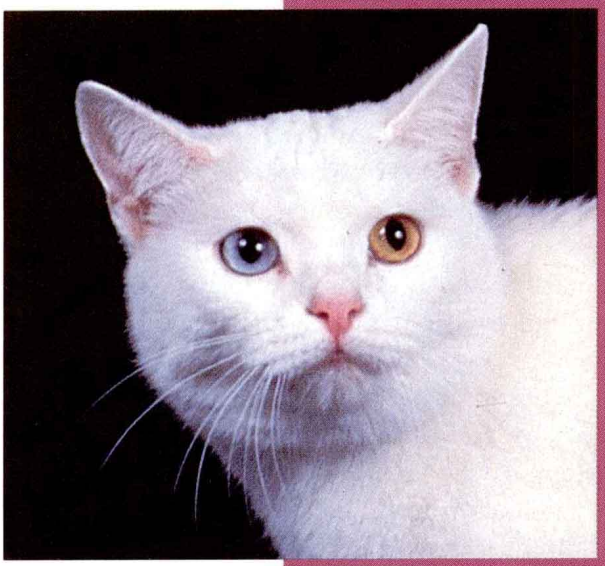
The most profound revelations, however, are to be found in the details of the story. We now know most of the major components—the genes, proteins, and other molecules—that cells use to control their growth; the pathways through which these components interact; the exact form of the mutations that disrupt growth control; and the causes of some of these mutations. This knowledge leads inevitably and provocatively to questions about what can be done to avoid, forestall, or rectify the changes that contribute to cancer.

The story we have tried to tell is intricate, and our understanding has become remarkably sophisticated. If our ambition to convey the many complex facts about cancer has not unduly challenged our readers, we must thank the staff at W. H. Freeman and Company—especially Amy Johnson and Diana Siemens, who improved our prose and clarified our drawings, and Travis Amos, who made the pages beautiful by securing superb photographs. We also owe a great debt to the many colleagues who supplied illustrations, advice, and encouragement, and to those who made the discoveries that inspired us to write this book.

Harold Varmus
Robert A. Weinberg

July 1992





Mutations change individual organisms: counterclockwise from top, a fruit fly's wings (doubled), a cat's eyes, zinnia petals, and corn kernels. Mutation, which drives evolutionary change, also drives cancer.

1

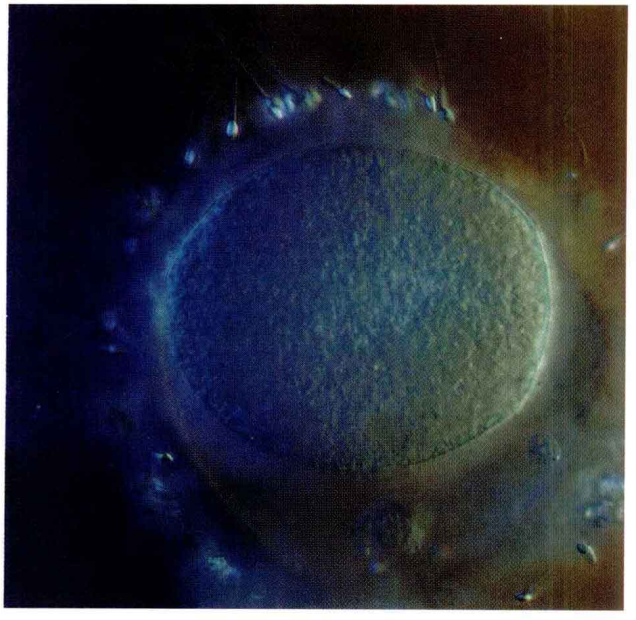
GENES, CELLS, AND COMPLEX ORGANISMS

Normal life processes are often directly illuminated by study of the abnormal. This book is another testament to the accuracy of that research adage; for the understanding it will record begins not with the beauties of living form, but with cancer, one of nature's aberrations. The fingers of a newborn child or the pattern of a butterfly's wing represent what we normally admire in biological systems: form; control; a unity of design and function that favors the survival of the organism. In cancer, all of these virtues are lost. Cancer cells divide without restraint, cross boundaries they were meant to respect, and fail to display the characteristics of the cell lineage from which they were derived. Yet it is from such

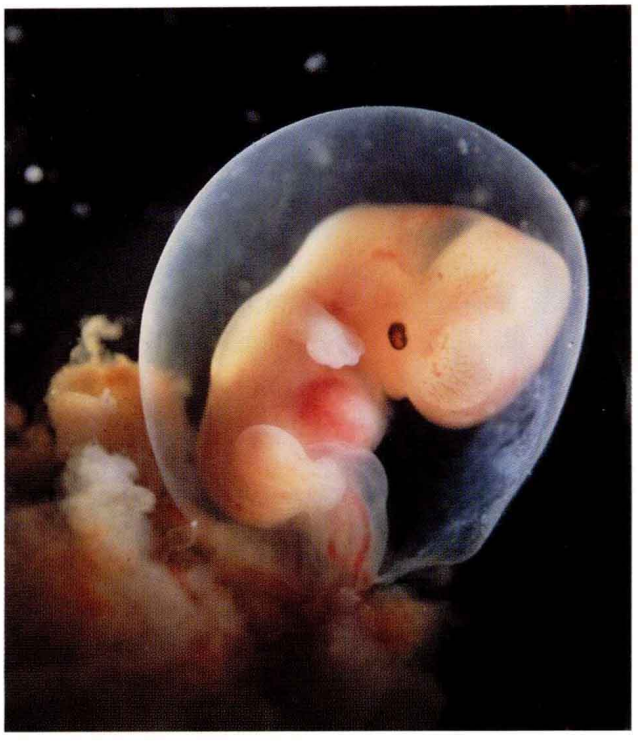
cells, with their feared consequences for the organism, that biologists have learned many key mechanisms in the drama of normal growth and development. To show why and how such mechanisms may fail—and may thereby initiate cancer—requires an initial review of our understanding of normal biological function, above all, of the general strategies that organisms use to govern their growth and differentiation.

All life on our planet depends upon one fundamental set of chemical components and processes. Living things are organized into functional units called cells; generate the same chemical, adenosine triphosphate (ATP), as their major source of energy; follow directions in a universal genetic code inherited in the same chemical form (deoxyribonucleic acid, or DNA); and, using nearly identical mechanisms, employ that code to make proteins and other components necessary for growth and, ultimately, the generation of new cells. (Viruses, whose credentials as living things remain controversial, conform to some but not all of these unifying features.) What is learned about one organism, then, may have significance for many others.

Although the universality of biological principles will be important at many stages of our story, a book about the molecular biology of cancer must emphasize at the outset a basic distinction that biologists make between simple, single-celled organisms, such as bacteria and yeast, and complex, multicelled organisms, such as plants, insects, and mammals. These complex organisms, composed of multiple tissues with different properties, must be endowed with genetic instructions for the growth



Sperm and egg unite (left); the single cell that results develops by successive divisions into a complex metazoan organism.



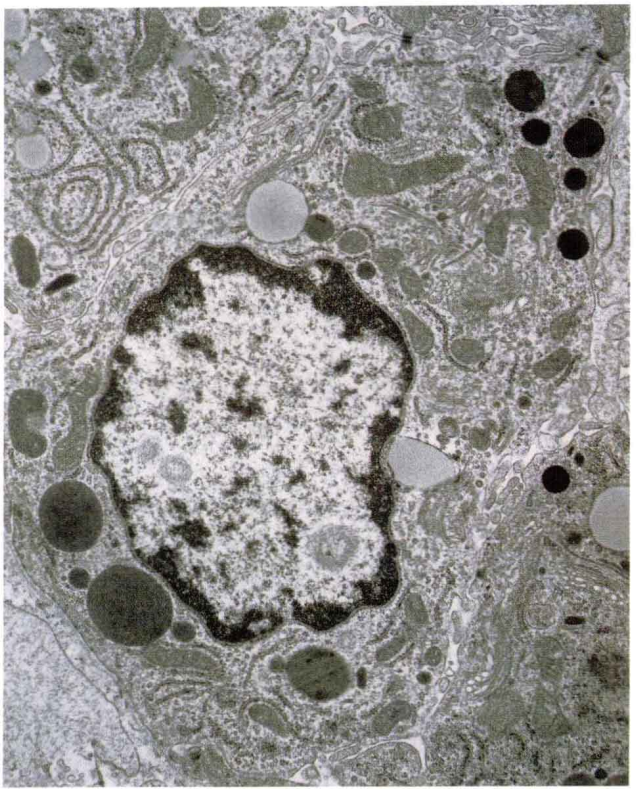
and maturation of many types of cells. When these instructions are distorted or fail, one possible result is the excessive growth of one cell type, often at the expense of others. In its most blatant form, the overgrowth invades locally and spreads to distant sites, finally killing the organism. When it affects human beings, we recognize this kind of derangement as cancer. The mechanisms underlying such events are the subject of this volume.

Disturbances of growth can and do occur in unicellular organisms, and they may be instructive for students of cancer. But true cancerous tumors (neoplasias) are found only in complex organisms and have been most fully characterized in animal species (metazoans). To understand neoplastic growth, therefore, it is necessary to establish some of the essential components of metazoans—the cellular units of which they are composed, the genetic instructions responsible for many of their properties, and the biochemical methods they use to put the instructions into action.

Essential Properties of Cells

The basic unit of biological growth and development is the cell. Each cell exhibits integrity of structure and a degree of autonomy, whether it exists as a single cell organism or as part of a complex multicellular organism composed of many cells differing in appearance and function.

At their peripheries, cells have physical boundaries that separate them from their envi-

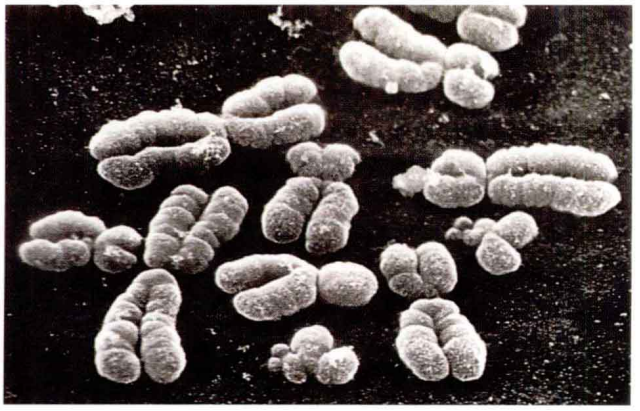


Transmission electron micrograph of an animal cell shows the large central nucleus and cytoplasmic molecular machines (organelles) that perform cell functions. Examples are the elongated mitochondria and the convoluted membranes of the Golgi apparatus and the endoplasmic reticulum.

ronments. Animal cells are packaged in simple fatty membranes, but plants, bacteria, and yeast manufacture, in addition, less penetrable cell walls that lie just beyond the membranes. Inside these barriers all cells consist of protoplasm: a dense mixture of large molecules—nucleic acids (DNA and ribonucleic acid, or RNA), proteins, carbohydrates, and fats—suspended in salt water.

Virtually all cells in a multicellular organism contain complete copies of the genetic instructions inherited from the organism's parents, to be passed in turn to its descendants. These instructions, in the form of long, two-stranded pieces of DNA, are present initially in a single cell, the fertilized egg—a precursor of the entire organism. This cell must undergo a large number of sequential divisions, each giving rise to two daughter cells, in order to produce the mature organism that, in humans, is composed of at least ten million million (10^{13}) cells. Following these inherited instructions, the cells not only multiply but also construct an imposing array of molecules that provide the cell architecture and perform the daily activities required both for survival and for specialized duties.

The complete set of genetic instructions in a cell is called its genome; the DNA molecules encoding the instructions are covered with proteins, together forming chromosomes. Cells are



A scanning electron microscope yields a view of human chromosomes; the genome comprises 22 pairs, plus the sex chromosomes.

Germ Cells and Somatic Cells

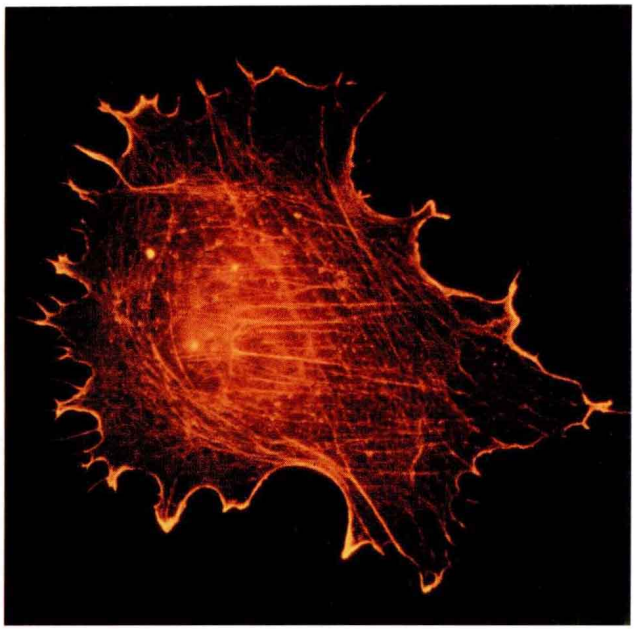
Unlike single-celled organisms, multicellular animals normally cannot be regenerated from any single cell, even though most cells contain the organism's complete genome and may grow in the absence of their usual neighbors. Instead, the business of making a new organism has been assigned to specialized cells that are found exclusively in the testes (male) and the ovaries (female). These are called the germ cells. Ordinary body (somatic) cells of animals are diploid, endowed with two copies of all chromosomes, except for the sex chromosomes. (The sex chromosomes are usually named X and Y, whereas all the others, called autosomal chromosomes, are numbered.) Unlike these somatic cells, germ cells are able to reduce their number of chromosomes by one-half through a special process called meiosis. The products of meiosis are haploid cells called gametes—sperm in males, eggs in females—with one copy of each autosomal chromosome and one sex chromosome. Gametes combine to form a single cell (a fertilized egg, or zygote) with a complete and unique diploid set of chromosomes, half from each parent. By successive divisions, the zygote ultimately gives rise to the large number of descendants that form a complete organism. The cell divisions of these somatic (nongerm) cells through the process of mitosis ensure the transfer of the entire complement of chromosomes, both maternally and paternally derived, to progeny cells.

Any change in the genetic information carried by a germ cell will affect the genome of the resulting zygote and in turn all the cells of the new organism. Since that change is present in both somatic and germ cells, it may be transmitted to succeeding generations. As we shall see in later chapters, certain inherited variations in genetic information can place many somatic cells in human beings at high risk for the development of cancer. Genetic change that arises in a somatic cell, however, affects only that cell and any cells derived from it by further rounds of cell division in the maturing individual; such a change, since it is not present in the germ cells, cannot be transmitted to the individual's offspring.

assigned to one of two categories, depending upon whether they contain a nucleus to isolate their chromosomes from other parts of the protoplasm. In eukaryotic organisms (including plants, yeast, and animals) the genome is fragmented into linear chromosomes and enclosed by a second, internal membrane that separates it from most of the rest of the cell (its cytoplasm). In prokaryotic organisms such as bacteria, the genome is usually one large, circular chromosome that floats within the cell, unconstrained by a nuclear membrane.

Metazoan organisms have existed as interdependent communities of cells for over a billion years. Yet despite this long evolution, the individual cells within a metazoan continue to display a considerable degree of independence. Each cell maintains a distinctive identity and the ability to perform most or all of the functions found in free-living, unicellular organisms. Almost all cells within a complex animal are endowed with the ability to synthesize organized units of large molecules (organelles) that efficiently perform many tasks essential to the cell or to the multicellular partnership. The jobs assigned to these molecular machines include protein synthesis (performed by ribosomes), ingestion and secretion of macromolecules (endosomes, endoplasmic reticulum, Golgi apparatus), degradation of certain wastes (lysosomes), and generation of chemical energy (mitochondria). In addition, the cytoplasm, far from being an amorphous sludge, is organized into a fibrous network (the cytoskeleton) that gives shape to cells lacking cell walls, moves cells through their environment, and transports internal units within cells. Many cells are also equipped with machinery for building a scaffolding of fibers in their surroundings; this extracellular matrix promotes cell adhesion, movement, and shape. Most cells, moreover, can duplicate all their constituents, including their chromosomal DNA, and divide, yielding two equally endowed daughter cells.

The remarkable materials packaged within the cell membrane—a full set of genes, many enzymes, and organelles—provide a metazoan cell with its potential for autonomous behavior. Removed from its natural setting and placed in another hospitable context, such as a petri dish supplied with nutrients and moisture, even a cell from a complex organism can retain the credentials of living things: the ability to grow and divide, to ingest and secrete, to build large molecules from small ones and degrade large ones to small. The perhaps unexpected autonomy of metazoan cells has both virtues and disadvantages. On the one hand, cells that can survive and multiply in a variety of contexts within the organism can travel throughout the body to provide specialized functions. On the other, cellular autonomy encompasses the potential for excessive growth, manifested ultimately as cancer.



A light micrograph displays the actin fibers that help to form the cytoskeleton, essential to the architecture of an animal cell. In this human epithelial cell from the lining of an organ, the actin matrix is dyed fluorescent orange with antibodies against actin.

Cellular Environments

Each cell in a complex organism, despite the remarkable autonomy we have emphasized so far, inhabits an environment—including extracellular fluids, dissolved chemicals, and insoluble fibers—that has been created by a wide variety of cells, both near and far. Even unicellular organisms possess devices (for example, sensors and propellers) for responding to changes in their surroundings, such as altered levels of nutrients or noxious substances. As might be expected from their need to intercommunicate, cells from metazoan organisms are especially well equipped in this respect. Such cells can both respond to and affect their environment in many ways, like individuals in a society. Cells in complex animals like humans are influenced by the nutrients, salts, and relative acidity of the fluids bathing them; by hormones secreted by distant organs; by factors produced by adjacent cells; and by the components of the extracellular matrix in which they reside.

Because cells can be removed from the normal environment in a tissue and grown in culture, a researcher can separate them from the normal array of signals that impinge on them and, in the confines of a petri dish, impose a new set of signals. Thus the timing of cell division, the direction of cell movement, and the execution of specialized cell functions can be manipulated by the experimenter, who can add or subtract components of the culture fluids, adjust the density and types of cells, or change the physical characteristics of the surface on which cells grow.

Alex Carrel, 1873–1944. Born in France, Carrel joined the Rockefeller Institute in 1906. He won a Nobel Prize in 1912 for research in vascular surgery and organ transplantation; in 1936 he and the aviator Charles A. Lindbergh invented a perfusion pump that they called an artificial heart. A pioneer in cell culture, Carrel kept cells from a chicken heart alive in his laboratory for 32 years.

Activities of Disrupted Cells

Cell biology focuses on the action of whole cells, but ultimately the behavior of each cell must be traced to the action of the multitude of molecules within it, the chemical reactions in which they participate, and the structures that they form. We are now in the realm of modern biochemistry and molecular biology, much of which is based upon the principle that many important components of cells—whether enzymes, organelles, or chromosomes

—may retain their essential structural and functional properties even after being removed from their usual intracellular contexts. Of course, once the integrity of the cell is violated, the orchestrated display of simultaneous functions that we call life is no longer possible. Disruption of the outer boundary of the cell—the membrane—will render the internal components unable to grow and divide, to move, or to respond to signals. Yet under appropriate conditions, many (but not all) of the individual processes essential to life can be continued in the material prepared from a disrupted cell: many cell proteins can still function as enzymes to catalyze certain chemical reactions; molecules containing genetic information may remain intact and may be manipulated to direct the synthesis of proteins; and multicomponent cellular machines, such as ribosomes and cell membranes, will continue to perform their normal tasks. Experimental success in isolating individual components and maintaining their activities after rupture of cells has been essential to the development of techniques that permit tinkering with—and promote an understanding of—fundamental cell activities, including many that are central to the problem of cancer.

The Genetic Blueprint and Its Variations

Nowhere has this ability to dissect the essential ingredients of cells advanced our view of life processes more profoundly than with respect to the genetic blueprint. The blueprint, which contains the information necessary to make all the proteins involved in both normal cell growth and cancer, is laid out in the chromosomes that organisms inherit from their parents and individual cells acquire from their cellular precursors. The genetic instructions exist in the form of a double-stranded helix of DNA composed of four basic units, the nucleotides (or, more properly, deoxyribonucleotides), named after their bases: adenine, cytosine, guanine, and thymidine—abbreviated A, C, G, and T. The A's on one strand always pair with T's on the other; reciprocally, C's pair with G's. The information content of DNA is embodied in the sequential arrangement of the nucleotides. The same is true of the other major nucleic acid, RNA, whose bases are identical with those of DNA except that uracil (U) takes the place of thymidine. (Some viruses, unlike all other organisms, carry their genetic information in the form of RNA rather than DNA.)

If we could scan a DNA sequence from one end of a chromosome to the other, we would find instructions, or signals, that determine the precise chemical composition of many proteins. Closely associated with these protein coding sequences are regulatory sequences that determine when a protein will be made, and in what quantities. Together the protein coding and regulatory sequences comprise a gene. Most genes are about 10,000 to 100,000 nucleotides in length; about 5000 of them are arranged along the average human chromosome. Although the information for making any single protein is almost always encoded by a single gene, one gene may sometimes carry the information needed to make several related proteins.

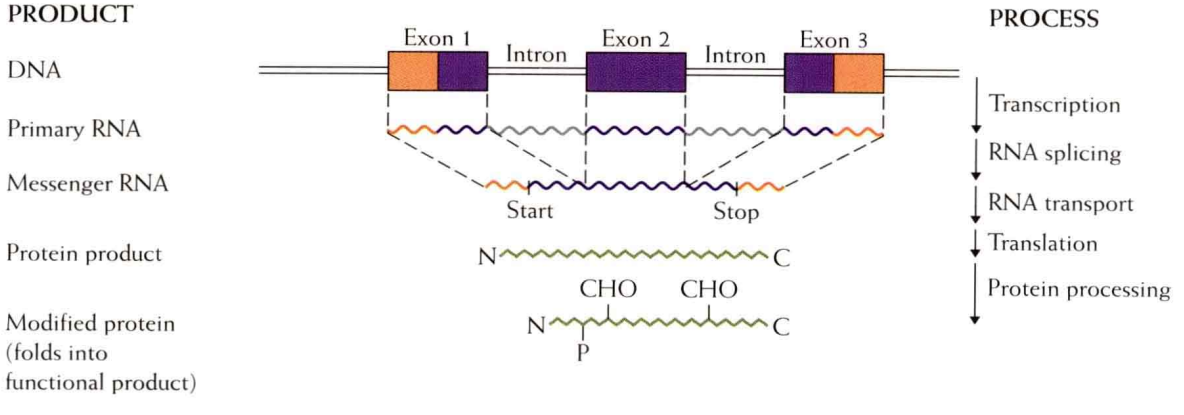
Actual production of proteins occurs in the cytoplasm of the cell. Since the instructions

for how to carry out this production are stored in the chromosomal DNA in the nucleus, an intermediary molecule—messenger RNA—must be synthesized to convey information from the nuclear storage repository to the cytoplasmic protein factory. The enzyme RNA polymerase transcribes the genetic information into an RNA copy from the nuclear DNA template. Because the protein-coding information is interrupted by irrelevant sequences (introns), the RNA must be edited (spliced) to remove the intron sequences and join the coding sequences (from exons). The messenger RNA that results from these transcription and processing events then moves to the cytoplasm, where it is used to generate protein.

This process, called translation, is accomplished by threading the messenger RNA through ribosomes, like a tape through the head of a tape player, to decode the information and assemble amino acids into chains. Each adjacent group of three nucleotides (a codon) specifies the next amino acid to be linked to the growing protein chain; translation, in effect, reads codons as if they were words, calling out one of the 20 possible amino acids for specific addition to the protein under synthesis. The properties of each protein are determined by the sequence of amino acids used to construct it, and this sequence in turn is determined directly by the nucleotide sequence of the messenger RNA template.

Regulation of Gene Expression

Although a mammalian cell possesses the genetic instructions to make on the order of 50,000 to 100,000 different proteins, only a subset of those proteins, perhaps 10 to 20 percent, are found in any single cell. This obser-



Gene transcription in an animal cell is initiated by a regulatory sequence of DNA nucleotides. After transcription into RNA, the genetic information is edited and spliced to omit the introns. Messenger RNA, transported out of the nucleus, is translated into protein on the cytoplasmic ribosomes. The protein then folds into its functional form and may be chemically modified (e.g., by addition of carbohydrate [CHO] or phosphate [P]).

vation introduces a new level of complexity into our understanding of genes, for it suggests that in each type of cell some genes are read out (by synthesis of messenger RNA molecules that, in turn, are used as templates to make diverse proteins), whereas other genes are silent (not read out). Moreover, different sets of genes are expressed in different types of cells, and among those genes that are read out (expressed), the outputs of gene products (proteins) can vary up to a millionfold. This means that in addition to the sequences that encode a protein, each gene must contain instructions that regulate production of the protein in correct amounts, at the correct times, for each cell type.

Gene expression can be regulated at several points. Most obvious and common are differences that result from the control of transcription: the amount of messenger RNA produced from a single gene can vary by many orders of magnitude. For example, the synthesis of messenger RNA for the protein component of hemoglobin occurs at extremely high rates in red-blood-cell precursors but at negligible levels in most other cells, even though they also carry copies of the gene for making hemoglobin. Later steps in the overall process of gene expression also afford points of control. The concentration of messenger RNA can be regulated by how fast it is synthesized in the nucleus; by its rate of transport to the ribosomes, where translation occurs; and by its relative ability to survive, once in the cytoplasm. If the messenger RNA is stable, it can be used repeatedly to make new protein chains; if quickly degraded, its capacity to direct protein synthesis will be correspondingly limited.

Still other constraints operate upon translation, influencing, for example, the likelihood that the ribosome factories will use a given messenger RNA molecule to make a protein. The long- or short-term stability of a newly synthesized protein can determine whether it will accumulate in a cell, and chemical modifications can influence its biological activity. Some proteins, for example, must have certain chemical groups—sugars, phosphates, or sulfates—attached to them in order to perform their normal functions; others must be broken apart (cleaved) at defined positions. The ability to make these modifications can create differences between cells as important as those achieved by different rates of RNA synthesis.

Genetic Variation, Mutation, and Viruses

This information for regulating protein synthesis, encoded using the four "letters" of the DNA "alphabet," makes biological variability possible at levels from cell type through species. Even individual variability within a species may be ultimately traced to subtle differences in the structure of proteins or the programs that dictate their production.

Biological variability implies that the underlying genetic blueprint is enormously plastic. The information content of DNA can be altered dramatically by minute changes in the order of its nucleotides or by gross rearrangements of its sequences. In other words, the meaning carried by a gene can be affected by even the subtlest alteration of its text.