

HHMI's 2002 Holiday Lectures on Science

Scanning Life's Matrix: Genes, Proteins, and Small Molecules

Lecture One

Reading Genes and Genomes

Eric S. Lander, Ph.D.

It started with a simple pea: The seeds of today's genetic revolution were sown more than 100 years ago by Gregor Mendel, an Austrian botanist and monk who teased the laws of heredity from the pea plants he was cultivating in his monastery garden. By the mid-20th century, scientists had unlocked the molecular secrets of heredity, showing that genetic information is encoded in long, winding stretches of DNA that must be faithfully copied by each cell before it divides to produce offspring. The genetic century closed with the goal of identifying every known gene in humans and other organisms. In this lecture, Dr. Lander takes us on a whirlwind tour of the remarkable advances in 20th-century genetics, from the rediscovery of Mendel's work in the winter of 1900 to the recent enumeration of the complete human genome sequence, the three billion letters of genetic instructions that make a human being.

Of course, determining the human genome sequence is not an end but a beginning. Thus, researchers are now working to decipher the massive amount of information contained in the letters of our genome, attempting to translate seemingly endless strings of As, Gs, Ts, and Cs into the genes that make the proteins in our bones, the enzymes that digest our food, the hormones that control our reproduction, and the brain chemicals that govern our moods, thoughts, and personalities. Dr. Lander reviews the powerful new technologies that are transforming biology and discusses how computers are being used to help researchers mine this cornucopia of genetic information. These DNA sequences, now collected in a public database called GenBank, stand to tell us not only where we came from but also how we function day to day.

The information contained in the genome, in essence, will provide us with the biological equivalent of the periodic table. Just as all matter can be described in terms of about 100 elements, the building blocks from which molecules are made, the behavior of all cells and organisms can be explained by the actions and interactions of the components of biology's periodic table—the tens of thousands of genes that make a human, a mouse, a fly, a tree, or any living thing on Earth. Having the genome sequence not just of humans but of many other organisms is like being given a key to greatest library on the planet—a place where evolution has been keeping notes. Thanks to the new science of genomics—the study of whole genomes—we can now pull up a chair and start to read.

Key Concepts

- The 20th century saw the birth of the genetic revolution. Over the course of the century, scientists unlocked the secrets of heredity, showing that genetic information is encoded in the long stretches of DNA that are copied carefully and passed on from generation to generation.
- In 2001, an international consortium of researchers determined the complete sequence of the human genome, the three billion letters of genetic instructions that make a human being. This string of instructions—a long list of As, Ts, Gs, and Cs—encodes the genes that make the proteins in our bones, the enzymes that digest our food, and the chemicals that give rise to our memories and our dreams.
- A vast array of new technologies, coupled with powerful computer programs, are being used to help researchers mine the cornucopia of genetic information. This data is collected in a public database called GenBank. It includes DNA sequences from a variety of organisms, such as humans, mice, worms, flies, yeast, plants, and bacteria.

- Together these genome sequences can be used to derive biology's periodic table—a set of genes whose actions and interactions are responsible for the existence and survival of all living things.

Lecture Two

Probing Genes and Genomes

Stuart L. Schreiber, Ph.D.

A genome sequence is both a parts list and a set of instructions for assembling those parts into a working organism. An organism's genome contains all its genes, and these genes direct the production of proteins, which form most of the structure and do most of the work in a cell. Proteins transport nutrients into the cell, copy its DNA, shuttle cargo around the cell, and carry messages to other cells. Before we can fully understand how muscles contract, how nerves conduct electrical impulses, how embryos develop, or how our bodies work, we must understand how proteins function inside the cell. To study what a protein does in a cell, scientists usually try to boost its activity or shut it down entirely. Although the latter approach may sound odd, seeing what happens to an organism when a particular protein is not working can offer clues about the role the protein normally performs.

One powerful approach to understanding proteins has been to alter their activity by mutating the genes that encode them. In this lecture, Dr. Schreiber presents an analogous chemical approach—using small molecules to modulate the function of specific proteins directly, rather than via their genes. This exciting approach, called chemical genetics, has been made possible in part by recent advances in computational and chemical-synthesis technologies. Scientists can now generate vast numbers of diverse small molecules and then use different screening methods to identify the most illuminating ones. The specially shaped chemicals can, for example, serve as molecular monkey wrenches, jamming themselves into the critical crevices of proteins in a way that prevents them from doing their jobs. Using these small molecules, researchers can disable a cellular protein instantly and then reverse the effect by removing the small molecule, something not possible with traditional genetic methods. As Dr. Schreiber will discuss, scientists are now using small molecules to dissect a variety of dynamic

cellular processes, from the movement of chromosomes during cell division to the molecular events that drive the development of complex organisms.

Chemical genetics is the biological equivalent of hitting the pause button on a VCR. Adding a compound will freeze a cell in mid-action; washing it away allows the action to resume. Such fine control allows researchers to monitor the activity of proteins involved in complex cellular processes that require the coordinated movement and interaction of several proteins over a short period of time. Rapidly reproducing cells, for example, may divide in about an hour, and some of the key proteins that direct this intricately choreographed molecular maneuver act fleetingly, performing for 10 seconds before they retire to await the next round of division. With these small molecules as precision tools, scientists are better equipped to explore the fundamental mechanisms by which cells grow, communicate, reproduce, and even die.

Key Concepts

- An organism's genome encodes proteins. These proteins perform most of the activities we associate with life, directing cells to grow, communicate, reproduce, and even die.
- To determine what a particular protein does inside a cell, researchers can either boost its activity or turn it off entirely. Both methods provide clues about the role a protein performs in the life of the organism.
- In an approach called chemical genetics, scientists use small molecules to modulate the activity of proteins. These specially shaped small molecules can jam into the protein's critical crevices, preventing proper function.
- Large numbers of structurally diverse small molecules can be synthesized using a matrix technique, and the small molecules that illuminate a specific aspect of biology can be identified using the technique of screening.

- Chemical genetics is fast and reversible. The small molecules inhibit their target proteins almost instantly; washing the compound away restores the protein's activity.
- Small molecules provide an ideal tool for dissecting dynamic cellular processes, particularly those that involve the coordinated movement and interaction of multiple protein components over a short period of time. In particular, chemical genetics has already yielded insights into the molecular events that underlie cell division.

Lecture Three

Human Genomics: A New Guide for Medicine

Eric S. Lander, Ph.D.

The recent progress in genetics and genomics has sparked one of the most remarkable revolutions in history. Even before researchers have learned how to decipher all the instructions hidden in the human genome, they are finding that the sequence can help them unravel the secrets of human disease, human heredity, and the evolutionary history of the human species. Having access to the full suite of human genes is already transforming biomedical science.

Researchers can now study diseases such as cancer from a global point of view, monitoring the activity of thousands of genes at once. By analyzing these genetic profiles, geneticists think they will be able to classify different types of cancer and predict which type of therapy can best be used to combat the disease.

Why is one form of cancer more deadly than another? And why is one person more prone to a particular disease than another? With the exception of identical twins, no two human genomes are exactly alike. Each contains small spelling differences that make an individual unique. As Dr. Lander discusses, these spelling differences, called genetic variations, underlie the range of diversity in the human species—our differences in height, weight, hair color, and skin color and our susceptibility to diseases such as cancer, heart disease, asthma, and diabetes. Fortunately, the same variations also allow researchers to track down the genes responsible for these different traits or diseases.

Researchers have also used the spelling differences in our genomes to track human evolution and migration. Such studies of our DNA reveal that we are indeed all members of the same big family group, which traces its roots to Africa and a small founding population of perhaps 50,000 individuals. In fact, because we are such a young species—only 150,000 years and 7,000 generations separate us from our African ancestors—the human genome contains relatively

little variation. Any two humans on earth are 99.9 percent identical in their DNA sequences. In contrast, two chimps chosen at random will show at least double that variation; orangutans, eight times as much. Of course, that 0.1 percent difference, in a genome of 3 billion letters, amounts to some 3 million differences between any two people. So, each of us is an individual. But we are more than our genes. As Dr. Lander discusses, genes may generate a diverse range of human traits, but they do not limit human potential.

Key Concepts

- Even before researchers have learned how to decipher all the instructions hidden in the human genome, they are finding that the sequence can help them unravel the secrets of human disease, human heredity, and the evolutionary history of the human species.
- With the exception of identical twins, no two human genomes are exactly alike. Each contains small spelling differences that make an individual unique. These genetic variations underlie the diversity of the human species—our differences in height, weight, hair color, and skin color and our susceptibility to diseases such as cancer, heart disease, asthma, and diabetes. Researchers can take advantage of these variations to track down the genes responsible for diseases.
- Researchers can also use the spelling differences in our genomes to track human evolution and migration. Such DNA studies reveal that humans are indeed part of one big family that traces its roots to a small founding population of only 50,000 individuals that lived in Africa about 150,000 years ago.
- Because we are such a young species, relatively speaking, humans show very little genetic variation. Any two humans are 99.9 percent identical in their genome sequences. Although each of us is unique, we are all very much the same.

Lecture Four

Chemical Genomics: New Tools for Medicine

Stuart L. Schreiber, Ph.D.

To probe thoroughly the intricate molecular circuitry that controls something as complex as a human cell, researchers would ideally like to generate a small molecule to modulate the activity of each of the more than 40,000 proteins encoded by the human genome. Designing this large set of highly specific molecules, one by one, would take an enormous amount of time and effort. Instead, investigators are taking a more systematic approach to synthesizing such a collection of chemicals—using automated techniques to perform a series of mix-and-match chemical reactions selected to generate a flood of diverse and complex compounds. Each of these small molecules could potentially interfere with a cellular protein to reveal its activity and provide a shortcut to treating diseases arising from protein malfunctions.

Working with millions of small molecules, researchers can begin to move from studying individual proteins to dissecting the intricate web of molecular interactions underlying everything that a cell or organism does. An appreciation of each protein's role in the cellular drama should lead scientists to a deeper understanding of complex biological processes, such as how an egg develops into an animal or how the activity of a neuron contributes to learning and memory. Dr. Schreiber discusses one example of how chemical genetics has provided insights into the hunger response, a process related to the medical disorder diabetes.

Drawing on the principles of genomics, which stress global, systemwide analyses, small-molecule-based explorations of biology can also take on a "chemical genomics" perspective. As Dr. Schreiber discusses, the information gained by treating cells and organisms with these molecular modulators will be used to generate a multidimensional matrix that connects each compound with

its target protein—eventually making connections to all proteins—and the effect the compound has on different types of cells or even different disease states. The matrix, called ChemBank, will serve as a repository for reams of experimental data collected from laboratories around the world. Armed with a set of custom-designed computer programs, researchers will be able to mine the database, searching through the multidimensional matrix for the patterns that will reveal how proteins and molecules work together to lay the foundations for life. Data gathered from ChemBank will also allow scientists to synthesize more "intelligent" molecules—compounds that are targeted toward an interesting protein or biological pathway. For example, by identifying the key properties shared by compounds that affect the activity of neurons in the brain, researchers interested in studying learning and memory can synthesize molecules that will be likely to influence these cognitive processes. As Dr. Schreiber explains, this merging of information science with advances in chemistry and biology is the basis for chemical genomics, a science that may hold the key to understanding life's matrix.

Key Concepts

- Designing enough molecules to modulate the activity of all 40,000 or so proteins in the human cell would take an enormous amount of time and effort. To streamline the process, scientists are taking advantage of automated techniques to generate a flood of complex yet diverse chemical compounds.
- Each of these small molecules can potentially interfere with the activity of a cellular protein. These compounds will not only reveal a protein's activity but may also provide a shortcut to developing drugs for treating human diseases that arise when proteins function incorrectly or not at all.
- The information gained by treating cells and organisms with millions of these molecular modulators will be fed into a multidimensional matrix called ChemBank. This dynamic database will connect each

compound with its known target protein and the observed effect exposure to the compound has on different types of cells.

- Researchers can navigate through the information contained in ChemBank, searching for the patterns that will reveal how proteins and molecules work together to build living systems.
- The merging of information science with advances in chemistry and biology is the basis for chemical genomics, a science that may hold the key to understanding life's matrix.