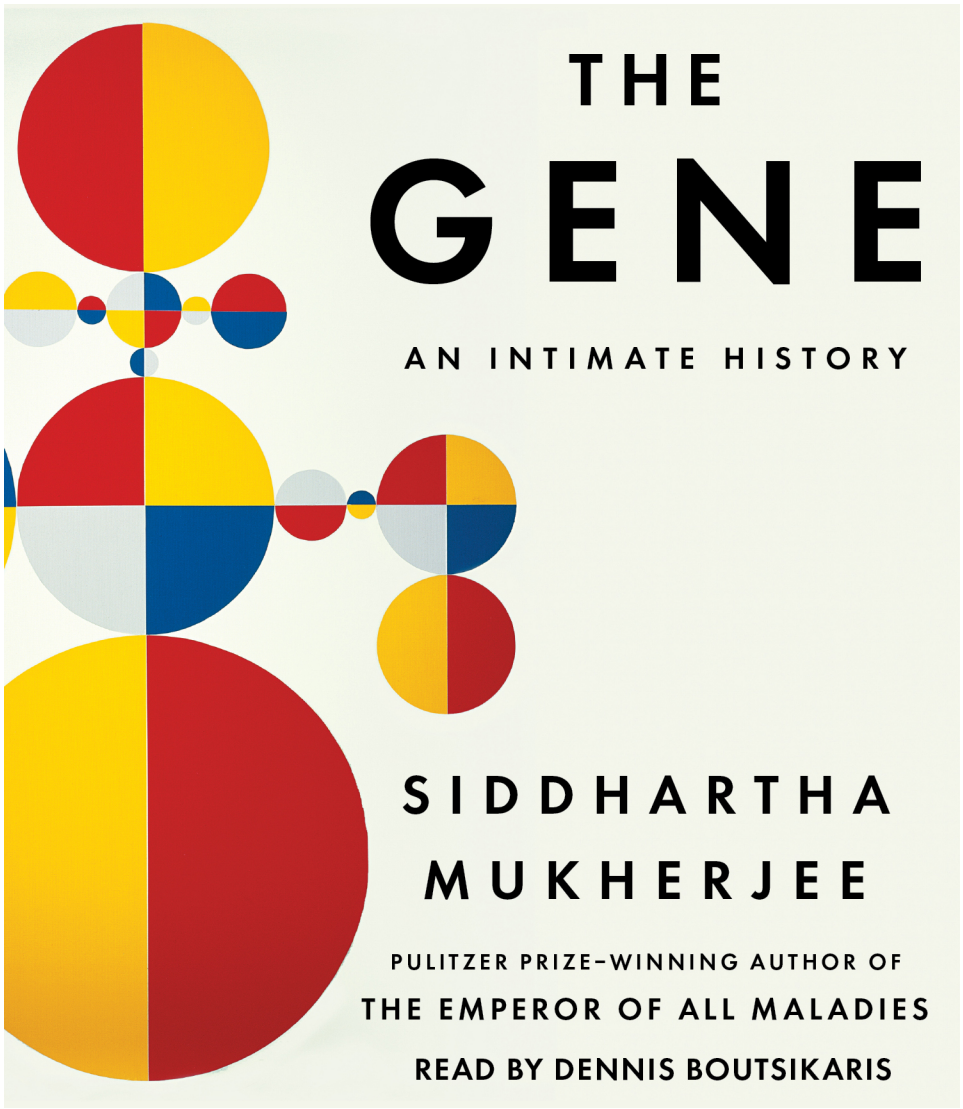
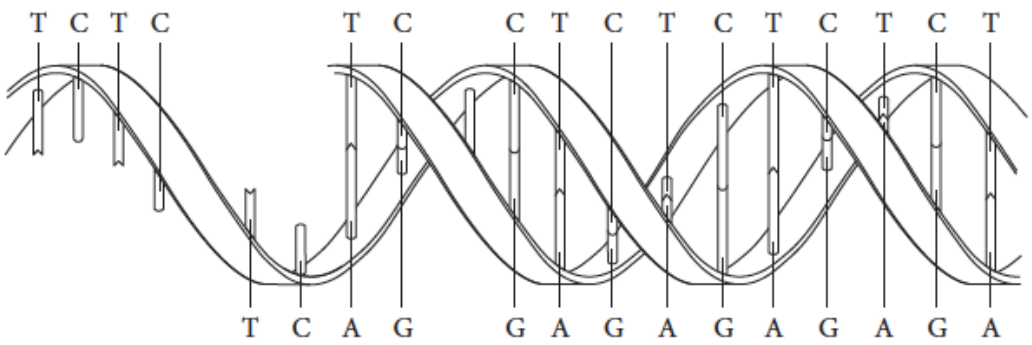
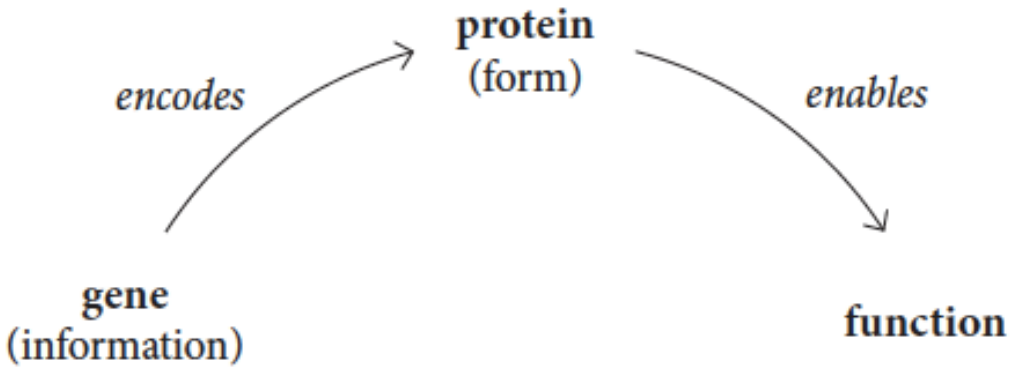

THE
GENE
AN INTIMATE HISTORY
SIDDHARTHA
MUKHERJEE
PULITZER PRIZE–WINNING AUTHOR OF
THE EMPEROR OF ALL MALADIES
READ BY DENNIS BOUTSIKARIS

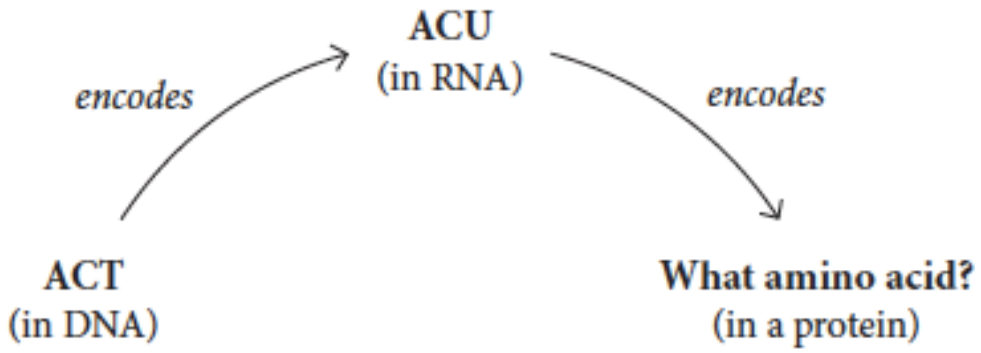


A schematic of the double-helical structure of DNA, showing a single helix (left) *and its paired double helix* (right). *Note the complementarity of bases: A is paired with T, and G with C. The winding "backbone" of DNA is made of a chain of sugars and phosphates.*

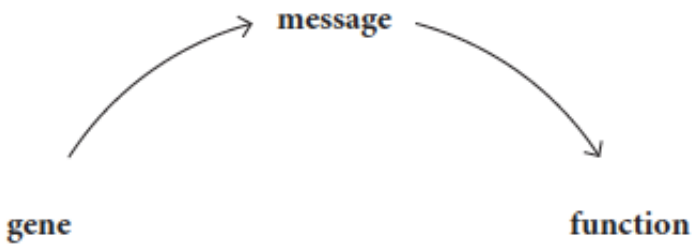
"A gene," Beadle wrote in 1945, "can be visualized as directing the final configuration of a protein molecule." This was the "action of the gene" that a generation of biologists had been trying to comprehend: a gene "acts" by encoding information to build a protein, and the protein actualizes the form or function of the organism.



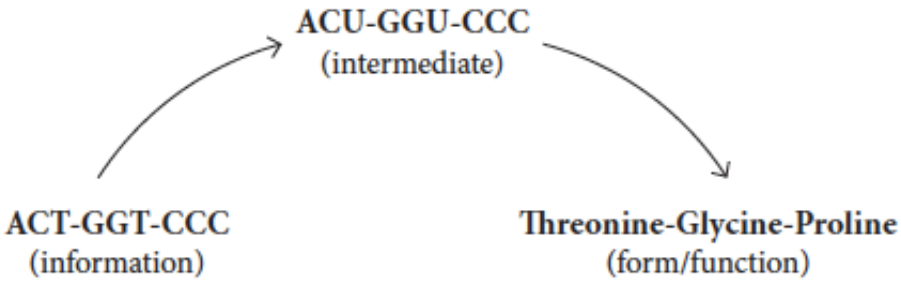
An analogy to natural language is illustrative. The letters A, C, and T convey very little meaning by themselves, but can be combined in ways to produce substantially different messages. It is, once again, the sequence that carries the message: the words act, tac, and cat, for instance, arise from the same letters, yet signal vastly different meanings. The key to solving the actual genetic code was to map the elements of a sequence in an RNA chain to the sequence of a protein chain. It was like deciphering the Rosetta Stone of genetics: Which combination of letters (in RNA) specified which combination of letters (in a protein)?
Or, conceptually:



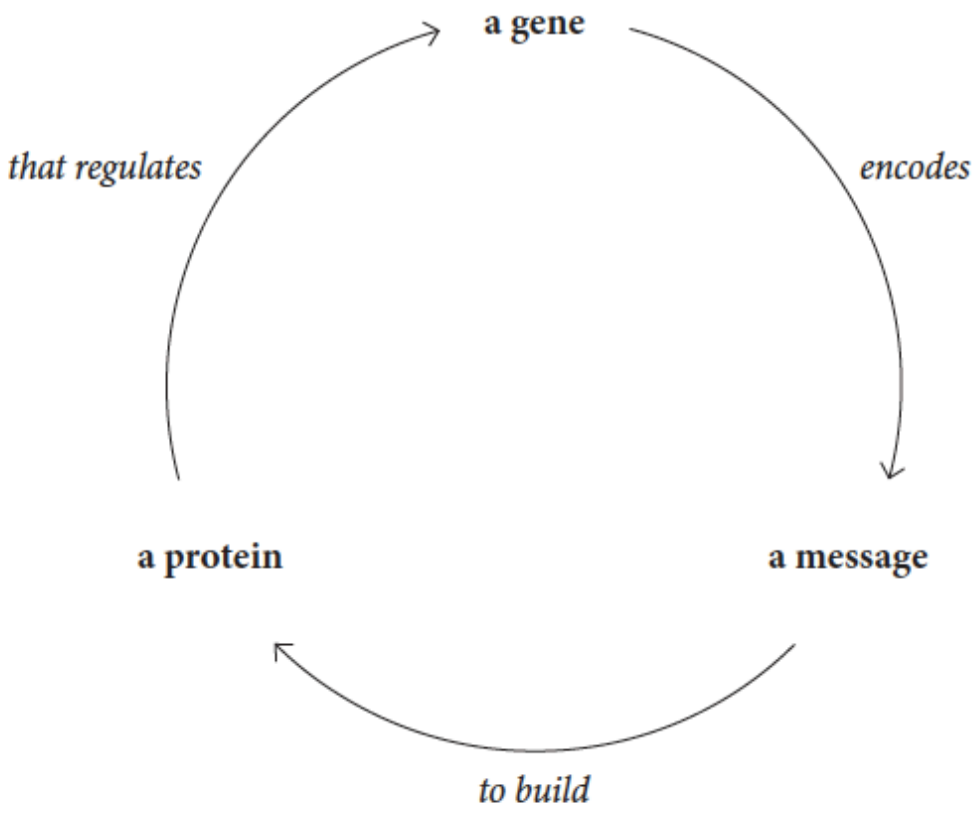
Francis Crick called this flow of information "the central dogma" of biological information. The word dogma was an odd choice (Crick later admitted that he never understood the linguistic implications of dogma, which implies a fixed, immutable belief)—but the central was an accurate description. Crick was referring to the striking universality of the flow of genetic information throughout biology. From bacteria to elephants—from red-eyed flies to blue-blooded princes—biological information flowed through living systems in a systematic, archetypal manner: DNA provided instructions to build RNA. RNA provided instructions to build proteins. Proteins ultimately enabled structure and function—bringing genes to life:



Or:



Proteins act as regulatory sensors, or master switches, in this process—turning on and turning off genes, or even combinations of genes, in a coordinate manner. Like the master score of a bewitchingly complex symphonic work, the genome contains the instructions for the development and maintenance of organisms. But the genomic "score" is inert without proteins. Proteins actualize this information. They conduct the genome, thereby playing out its music—activating the viola at the fourteenth minute, a crash of cymbals during the arpeggio, a roll of drums at the crescendo.
Or conceptually:



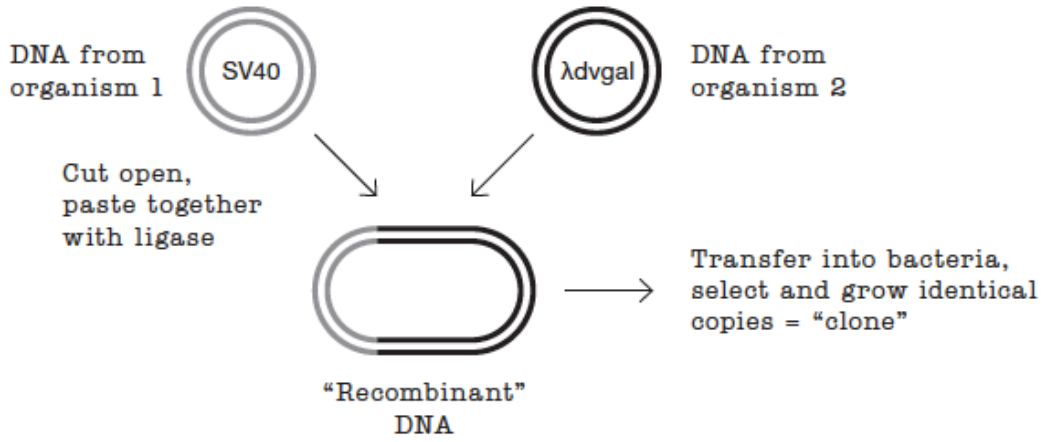
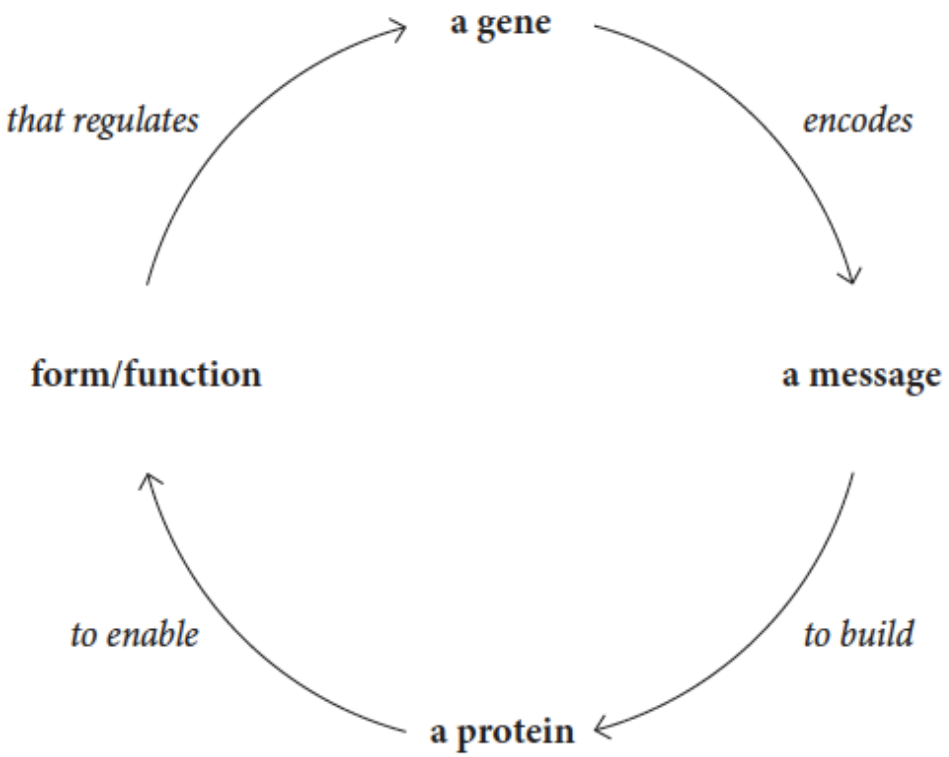


Figure adapted from Paul Berg's paper on "Recombinant" DNA.
By combining genes from any organisms, scientists could engineer genes at will, foreshadowing human gene therapy and human genome engineering.

Introns are not the exception in human genes; they are the rule. Human introns are often enormous—spanning several hundreds of thousands of bases of DNA. And genes themselves are separated from each other by long stretches of intervening DNA, called intergenic DNA. Intergenic DNA and introns—spacers between genes and stuffers within genes—are thought to have sequences that allow genes to be regulated in context. To return to our analogy, these regions might be described as long ellipses scattered with occasional punctuation marks. The human genome can thus be visualized as:

This . . . . . is . . . . . the . . . . . (. . .) . . . s . . . truc . . . ture . . . . . of . . . . . your . . . . . gen . . . om . . . e;

"How does an organism transmit the information needed to create form and function to its embryo?" Aristotle had asked. The answer to that question, viewed through model organisms such as peas, fruit flies, and bread molds, had launched the discipline of modern genetics. It had resulted, ultimately, in that monumentally influential diagram that forms the basis of our understanding of information flow in living systems:

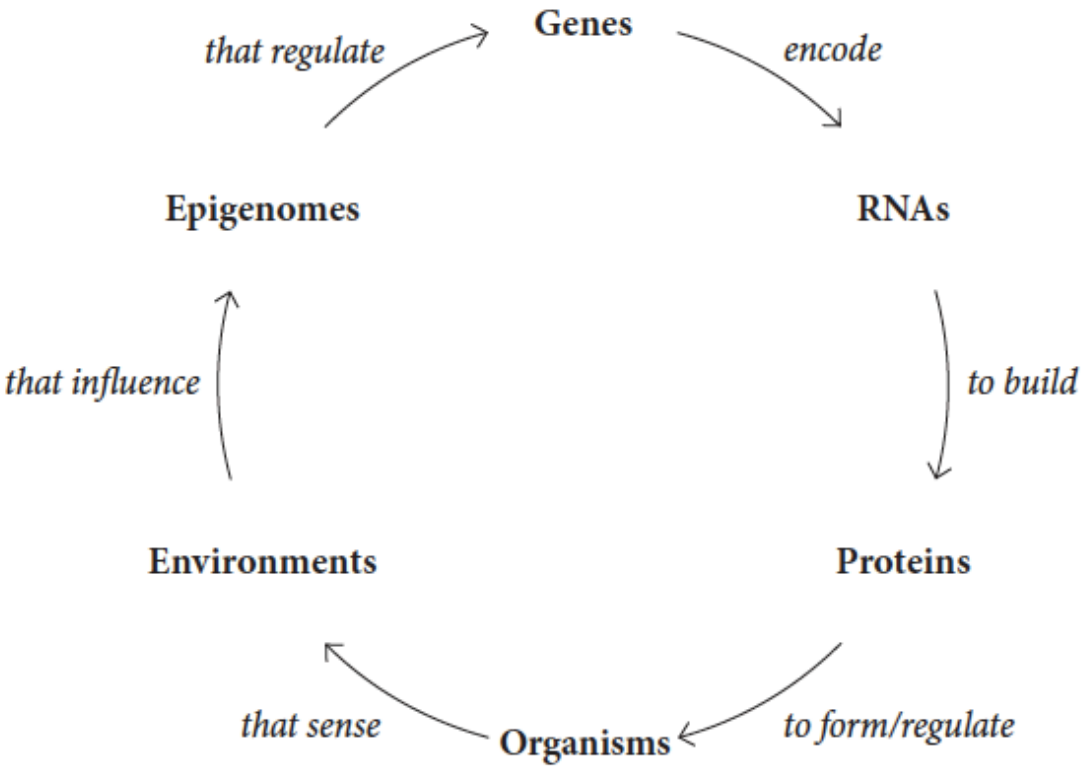


Returning to the sentence analogy, marginalia recorded in a book—pencil lines, underlined words, scratch marks, crossed-out letters, subscripts, and endnotes—modify the context of the genome without changing the actual words. Every cell in an organism inherits the same book, but by scratching out particular sentences and appending others, by "silencing" and "activating" particular words, by emphasizing certain phrases, each cell can write a unique novel from the same basic script. We might visualize genes in the human genome, with their appended chemical marks, thus:

*. . . This . . . . is . . . **the** . . . . . , , , ... struc . . . ture , ...*
of . . . Your . . . . . Gen . . . ome . . .

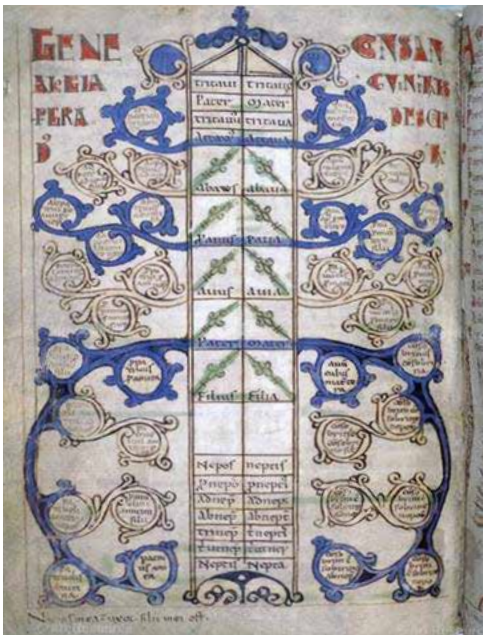
As before, the words in the sentence correspond to the genes. The ellipses and punctuation marks denote the introns, the intergenic regions, and regulatory sequences. The boldface and capitalized letters and the underlined words are epigenetic marks appended to the genome to impose a final layer of meaning.

The circular flow of biological information—

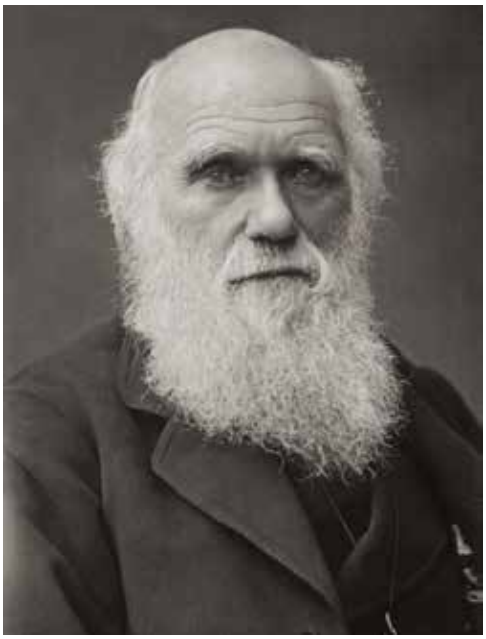


—is, perhaps, one of the few organizing rules in biology. Certainly the directionality of this flow of information has exceptions (retroviruses can pedal "backward" from RNA to DNA). And there are yet-undiscovered mechanisms in the biological world that might change the order or the components of information flow in living systems (RNA, for instance, is now known to be able to influence the regulation of genes). But the circular flow of biological information has been chalked out conceptually. This flow of information is the closest thing that we might have to a biological law. When the technology to manipulate this law is mastered, we will move through one of the most profound transitions in our history. We will learn to read and write our selves, ourselves.

This homunculus, wrapped inside human sperm, was drawn by Nicolaas Hartsoeker in 1695. Like many other biologists in his time, Hartsoeker believed in "spermism," the theory that the information to create a fetus was transmitted by the miniature human form lodged inside sperm.



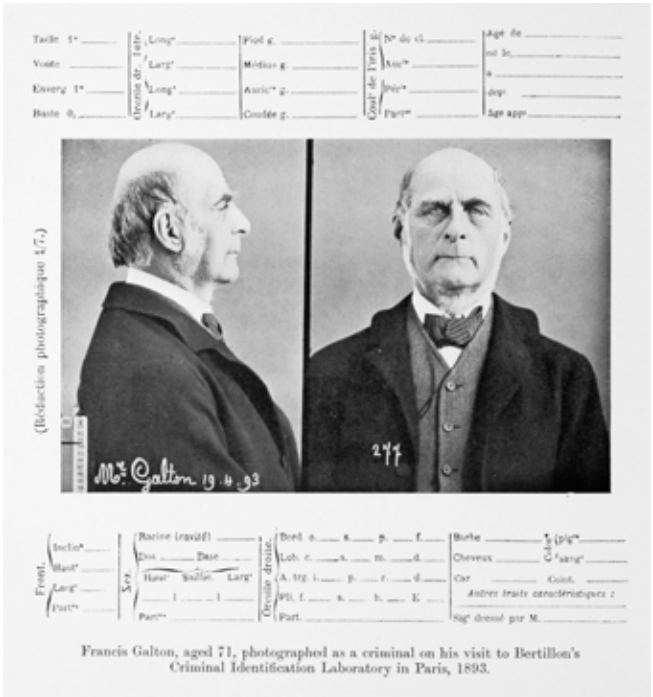
In medieval Europe, "trees of lineage" were often created to mark the ancestors and descendants of noble families. These trees were used to stake claims on peerage and property, or to seek marital arrangements between families (in part, to decrease the chances of consanguineous marriages between cousins). The word *gene*—at the top left corner—was used in the sense of genealogy or descent. The modern connotation of *gene*, as a unit of hereditary information, appeared centuries later in 1909.



Charles Darwin (here in his seventies) and his "tree of life" sketch, showing organisms radiating out from a common ancestral organism (note the doubt-ridden phrase "I think," scribbled above the diagram). Darwin's theory of evolution by variation and natural selection demanded a theory of heredity via genes. Close readers of Darwin's theory realized that evolution could work only if there were indivisible, but mutable, particles of heredity that transmit information between parents and offspring. Yet Darwin, having never read Gregor Mendel's paper, never found an adequate formulation of such a theory during his lifetime.

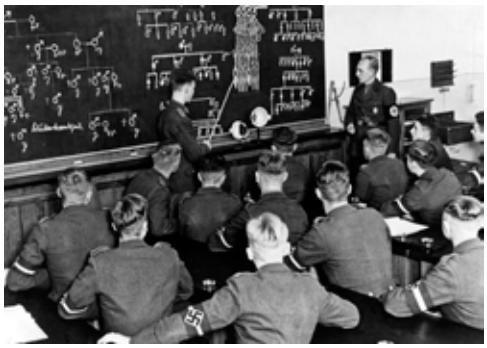
Gregor Mendel holds a flower, possibly from a pea plant, in his monastery garden in Brno (now in the Czech Republic). Mendel's seminal experiments in the 1850s and '60s identified indivisible particles of information as carriers of hereditary information. Mendel's paper (1865) was largely ignored for four decades, and then transformed the science of biology.

William Bateson's "rediscovery" of Mendel's work in 1900 converted him into a believer in genes. Bateson coined the term *genetics* in 1905 to describe the study of heredity. Wilhelm Johannsen (*left*) coined the term *gene* to describe a unit of heredity. Johannsen visited Bateson at his house in Cambridge, England; the two became close collaborators and vigorous defenders of the gene theory.



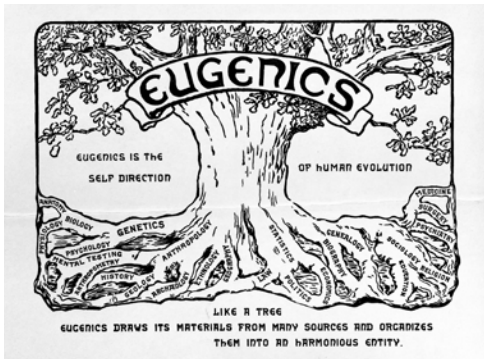
Francis Galton, aged 71, photographed as a criminal on his visit to Bertillon's Criminal Identification Laboratory in Paris, 1893.

Francis Galton—mathematician, biologist, and statistician—put himself on one of his own "anthropometry cards," in which he tabulated a person's height, weight, facial features, and other characteristics. Galton resisted Mendel's theory of genes. He also believed that the selective breeding of humans with the "best" features would lead to the creation of an improved human race. *Eugenics*, a term coined by Galton for the science of human emancipation through the manipulation of heredity, would soon morph into a macabre form of social and political control.



The Nazi doctrine of "racial hygiene" prompted a vast state-sponsored effort to cleanse the human race through sterilization, confinement, and murder. Twin studies were used to prove the power of hereditary influences, and men, women, and children were exterminated based on an assumption that they carried defective genes. The Nazis extended their eugenic efforts to exterminate Jews, Gypsies, dissidents, and homosexuals. Here, Nazi scientists measure the height of twins, and demonstrate family history charts to Nazi recruits.

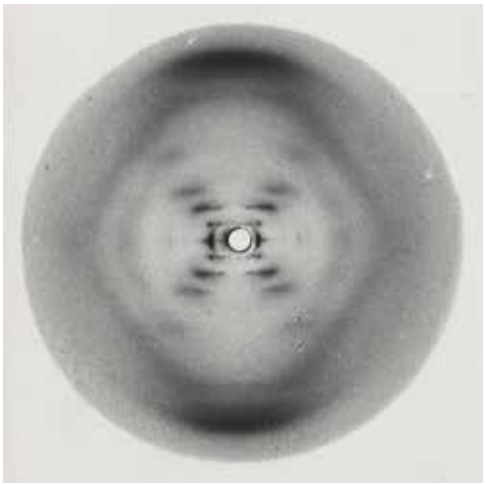
Better Babies contests were introduced in the United States in the 1920s. Doctors and nurses examined children (all white) for the best genetic features. Better Babies contests generated passive support for eugenics in America by showcasing the healthiest babies as products of genetic selection.



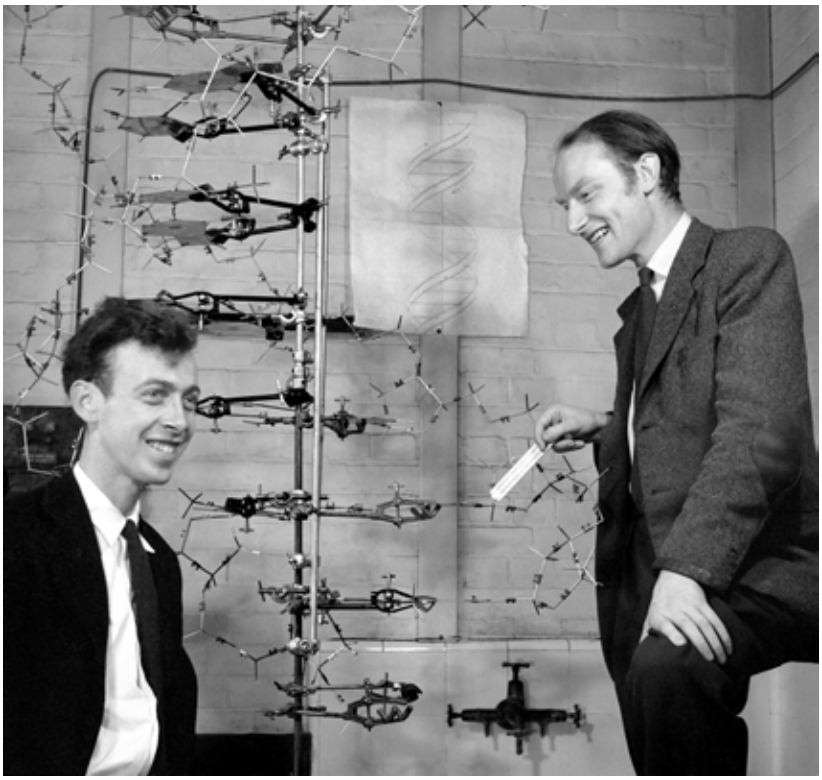
A "eugenics tree" cartoon from the United States argues for the "self-direction of human evolution." Medicine, surgery, anthropology, and genealogy are the "roots" of the tree. Eugenic science hoped to use these foundational principles to select fitter, healthier, and more accomplished humans.

In the 1920s, Carrie Buck and her mother, Emma Buck, were sent to the Virginia Colony for Epileptics and Feebleminded, where women classified as "imbeciles" were routinely sterilized. The photograph, obtained on the pretext of capturing a casual moment between mother and daughter, was staged to provide evidence of the resemblance between Carrie and Emma, and thus proof of their "hereditary imbecility."

At Columbia University, and subsequently at Caltech University in the 1920s and '30s, Thomas Morgan used fruit flies to demonstrate that genes were physically linked to each other, presciently predicting that a single, chainlike molecule carried genetic information. Linkage between genes would eventually be used to generate genetic maps in humans and lay the foundation for the Human Genome Project. This is Morgan in his Caltech Fly Room, surrounded by the milk bottles in which he bred his maggots and flies.



Rosalind Franklin looks down a microscope at King's College in London in the 1950s. Franklin used X-ray crystallography to photograph and study the structure of DNA. Photograph 51 is the clearest of Franklin's photographs of a DNA crystal. The photo suggested a double-helix structure, although the precise orientations of the bases A, C, T, and G were not clear from it.

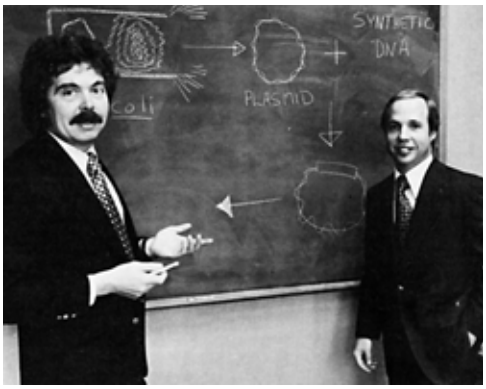


James Watson and Francis Crick demonstrate their model of DNA as a double helix in Cambridge in 1953. Watson and Crick solved the structure of DNA by realizing that the A in one strand was paired against the T in the other, and the G against the C.

At the Moore Clinic in Baltimore in the 1950s, Victor McKusick created a vast catalog of human mutations. He found that one phenotype—short stature, or "dwarfism"—could be caused by mutations in several disparate genes. Conversely, diverse phenotypes could be caused by mutations in a single gene.

Nancy Wexler's mother and uncles were diagnosed with Huntington's disease, a lethal neurodegenerative disease that spurs involuntary sinuous or jerking movements. The diagnosis launched her personal hunt for the gene that causes the illness. Wexler found a cluster of several patients with Huntington's disease in Venezuela, all likely descended from a founder with the disease. Huntington's disease was one of the first human diseases to be definitively identified with a single gene using modern gene-mapping methods.

Students protested a genetics meeting in the 1970s. The novel technologies of gene sequencing, gene cloning, and recombinant DNA raised anxieties that new forms of eugenics would be used to create a "perfect race." The link to Nazi eugenics was not forgotten.

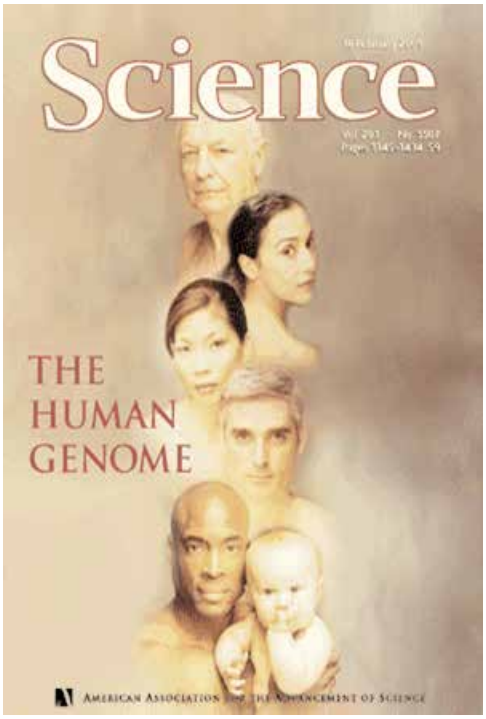


Herb Boyer (*left*) and Robert Swanson founded Genentech in 1976 to produce medicines out of genes. The drawing on the blackboard shows the scheme to produce insulin using recombinant DNA technology. The first such proteins were produced in enormous bacterial incubators under Swanson's watchful eye.

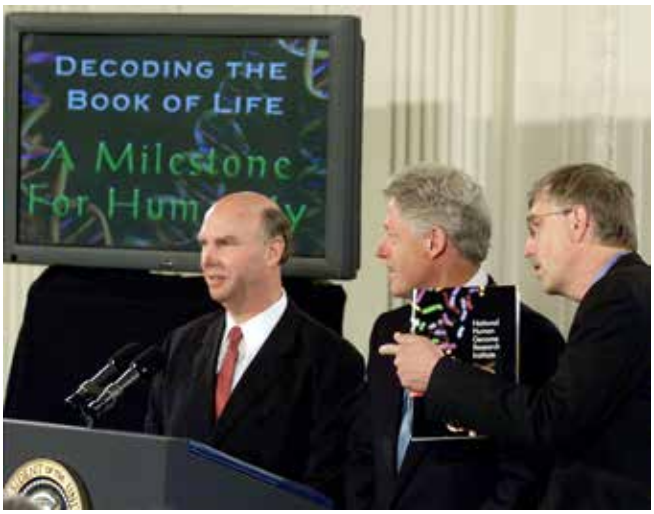
Paul Berg speaks to Maxine Singer at the Asilomar meeting in 1975, while Sydney Brenner takes notes in the background. Following the discovery of technologies to create genetic hybrids between genes (recombinant DNA) and produce millions of copies of these hybrids in bacterial cells (gene cloning), Berg and others proposed a "moratorium" on certain recombinant DNA work until the risks had been adequately assessed.

Frederick Sanger examines a DNA sequencing gel. Sanger's invention of a technique to sequence DNA (i.e., read the precise stretch of letters—A, C, T, and G—in a gene's sequence) revolutionized our understanding of genes, and set the stage for the Human Genome Project.

Jesse Gelsinger posed in Philadelphia a few months before his death in 1999. Gelsinger was one of the first patients to be treated with gene therapy. A virus was designed to deliver the correct form of a mutated gene into his liver, but Gelsinger had a brisk immunological response to the virus, resulting in organ failure and death. Gelsinger's "biotech death" would spur nationwide responses to ensure the safety of gene-therapy trials.



The February 2001 cover of *Science* magazine announced the draft sequence of the human genome.



Craig Venter (*left*), President Bill Clinton, and Francis Collins announced the draft sequence of the human genome on June 26, 2000, at the White House.

Even without subtle techniques to alter human genomes, the capacity to assess a child's genome in utero has led to vast dysgenic efforts around the world. In parts of China and India, the assessment of male versus female gender by amniocentesis, and the selective abortion of female fetuses, has skewed the sex ratio to 0.8 females to 1 male, and caused unprecedented alterations of population and family structures.

Faster and more-accurate gene-sequencing machines (housed inside gray boxlike containers) linked to supercomputers that analyze and annotate genetic information can now sequence individual human genomes in months. Variations of this technique can be used to sequence the genome of a multicelled embryo or a fetus, enabling preimplantation genetic diagnosis and in utero diagnosis of future illness.

Jennifer Doudna (*right*), a biologist and RNA researcher at Berkeley, is among those working on a system to deliver targeted, intentional mutations in genes. In principle, the system can be used to "edit" the human genome, although the technology still remains to be perfected and assessed for safety and fidelity. If intentional genetic changes were introduced into sperm, egg, or human embryonic stem cells, the technology would portend the genesis of humans with altered genes.