

## Primer on Medical Genomics

### Part I: History of Genetics and Sequencing of the Human Genome

CINDY PHAM LORENTZ, MS; ERIC D. WIEBEN, PhD; AYALEW TEFFERI, MD; DAVID A. H. WHITEMAN, MD;  
AND GORDON W. DEWALD, PhD

In comparison with most other disciplines of science, the field of genetics is still in its youth. The majority of scientific work in genetics has been done in the past 150 years. The successful preliminary sequencing of the human genome was announced in 2001. Nonetheless, interest in heredity and in other concepts within the field of genetics has existed since the beginning of humanity. This article provides an account of the history of genetics, spanning from humankind's initial attempts to understand and influence heredity, to the early scientific work in the field of genetics,

and subsequently to the advancements in modern genetics. Additionally, the Human Genome Project is summarized, from inception to publication of the "first draft" of the human genome sequence.

*Mayo Clin Proc.* 2002;77:773-782

DOE = Department of Energy; HGP = Human Genome Project; NCHGR = National Centre for Human Genome Research; NIH = National Institutes of Health; SNPs = single nucleotide polymorphisms

Even though not recorded, there is reason to believe that the concept of genetics was contemplated by the first human beings. When specific traits were noted to be shared by parent and child, this likely raised questions about the phenomenon of inheritance (although it was not termed as such then).<sup>1</sup> The same phenomenon of shared traits between parent and offspring was then applied in the areas of plant cultivation and animal husbandry.<sup>2</sup> Decision making based on the concept that traits could be passed on from generation to generation laid the foundation for the modern genetics that we practice today.

Like their predecessors, geneticists of today are influenced by trends and developments in science. Currently, substantial energy and resources are being directed into efforts to sequence the human genome.<sup>3</sup> A complete sequence of the human genome holds the potential to revolutionize science and medicine. The first part of this overview gives an account of the history of genetics that spans from humankind's first attempts at understanding and influencing heredity,<sup>4</sup> to the early scientific work in the field of genetics,<sup>5</sup> and then to the advancements in modern genetics.<sup>6</sup> The second part summarizes the Human Genome

Project (HGP) from inception to the publishing of the "first draft" of the human genome sequence.<sup>3,7,8</sup>

---

[For editorial comment, see page 745.](#)

---

#### PRE-MENDELIAN AND MENDELIAN GENETICS

Early work in the cultivation of plants and the domestication of animals provides evidence that humans were interested in the inherited aspects of phenotypes and the manipulation of parent-parent pairings to improve future generations.<sup>9</sup> After the dioecious nature of the date palm was noted in 5000 BC, the ancient Babylonians and Assyrians practiced artificial pollination from the time of King Hammurabi in 2000 BC. The presence of cultivated plants and domesticated animals today may be the result of these first attempts at genetic experimentation. From these activities, we can infer that humans have long recognized that certain traits were preserved in subsequent generations.

The Greek philosophers recorded their speculations on the concepts of inheritance as they applied to humans.<sup>1,4</sup> Early writings date back to the pre-Socrates era from the sixth and fifth centuries BC.<sup>9</sup> They debated the relationship between heredity and environment on the origin of birth defects and on factors that affect sex determination. Most notably, from 460 to 322 BC, Greek philosophers Hippocrates, Aristotle, and Plato wrote about the inheritance of human traits.<sup>2,4</sup> They observed that certain traits were frequently (ie, in a dominant fashion) passed from parent to child. These early Greek philosophers believed that semen was in some way responsible for passing on

---

From the Division of Laboratory Genetics (C.P.L., G.W.D.), Department of Biochemistry and Molecular Biology (E.D.W.), Division of Hematology and Internal Medicine (A.T.), and Department of Medical Genetics (D.A.H.W.), Mayo Clinic, Rochester, Minn. Drs Wieben, Tefferi, Whiteman, and Dewald are members of the Mayo Clinic Genomics Education Steering Committee.

Individual reprints of this article are not available. The entire Primer on Medical Genomics will be available for purchase from the *Proceedings* Editorial Office at a later date.

traits, although they did not understand the exact contribution by the male or female parent to the offspring.<sup>2</sup> Plato is noted to have understood that humans were most likely to mate with other humans with similar characteristics and that on occasion this led to offspring who were less able to deal with “the challenges of life.”<sup>10</sup>

There were 2 main schools of thought on the conception and in utero development of human offspring. Philosophers such as Empedocles and Democritus favored the theory of preformation, which was in opposition to that of Aristotle, who favored a form of epigenesis.<sup>11</sup> In 1694, Hartsoeker’s *Essay de Dioptrique* was published supporting preformation.<sup>12</sup> This essay included his rendition of the human spermatozoon with an enclosed homunculus (a preformed male or female animalcule).<sup>13</sup> Later, even Hartsoeker dismissed this concept of infinite encasements of preformed animals within each species.<sup>13</sup> The supporters of epigenesis believed that the organism was formed from components contributed by 1 or both parents.

In the 18th century, the debate between preformation and epigenesis was still unresolved, yet new disagreements arose within the camp of preformationists.<sup>14</sup> Malpighi<sup>15</sup> believed that the preformed new organism was already present and preformed in the egg. Some of his contemporaries, such as van Leeuwenhoek,<sup>16</sup> supported preformationism, but they believed that the preformed organism was within the male sperm. This led to factions within the preformationist camp of ovids vs spermists.

Maupertuis was a mathematician with an interest in biology.<sup>17,18</sup> He bred various animals and observed that first-generation hybrids had characteristics of both parents. This was a strong argument against the popular preformationism theory of that time, which Maupertuis irreverently described as “mythical” and undeserving of consideration. He was a strong advocate for pangenesis, the belief that individuals were formed from elements from the seminal fluid of each parent. In 1752, Maupertuis described an autosomal dominant postaxial polydactyly in 4 generations of 1 family.<sup>18,19</sup> This may have been the first report of an inherited genetic disorder.

In 1814, Joseph Adams published *A Treatise on the Supposed Hereditary Properties of Diseases*, but this work received little recognition from his peers.<sup>20</sup> Adams recognized the difference between autosomal recessive and autosomal dominant conditions (although he did not use this terminology). He also recognized that it was not favorable to mate with one’s relatives, that hereditary diseases can express later in life, that some hereditary diseases require an environmental exposure to be expressed, that there was some intrafamilial correlation of diseases, and that the reproductive fitness of individuals with hereditary disease was diminished.

It is difficult to separate the discoveries in genetics from other discoveries about the cell because their functions are so closely intertwined. Cells were first described by Hooke<sup>21</sup> in 1665 using a primitive light microscope to study animal and plant tissues. Wolff<sup>22</sup> may have been one of the originators of the cell theory. He recognized in the 1760s that cellular structure was an essential mark of all living creatures and that cells had similar structures in plants and animals. In 1838, Schleiden and Schwann<sup>23</sup> suggested that cells with nuclei were the fundamental units of life. In 1855, Virchow<sup>24</sup> hypothesized that new cells can be formed only by the division of existing cells.

In 1859, Darwin<sup>25</sup> published *On the Origin of Species*, proposing evolution by natural selection, but the principles of genetics to defend his theory were not widely known at that time. In 1865, Mendel<sup>5</sup> published “Experiments in Plant Hybridization,” proposing the principles of heredity and introducing the concept of dominant and recessive genes to explain how a characteristic can be repressed in 1 generation but appear in the next generation. Today, he is widely considered the founding father of modern genetics because of his extensive experiments validating the basic tenets of genetics. Moreover, Mendel’s work eventually helped to partially explain Darwin’s concept of evolution.

In 1869, Miescher studied extracts of cellular nucleic acid and coined the term *nuclein*.<sup>26,27</sup> We now know that he had discovered the material basis of heredity, but another 80 years passed before nuclein was shown to be DNA.<sup>28</sup> Between 1879 and 1882, Flemming<sup>29</sup> used “new staining techniques” to see “tiny threads” within the nucleus of cells in salamander larvae that appeared to be dividing. In so doing, he discovered chromosomes that he named *chromatin* because of this material’s affinity for taking up stain. Shortly thereafter, in 1889, Weismann<sup>30</sup> published the first of a series of papers in which he theorized that the material basis of heredity was located on the chromosomes.

## MODERN GENETICS

Expanding on these early achievements, the 20th century brought particularly exciting discoveries in genetics, beginning with the discovery of the gene and ending with the first draft of the human genome.<sup>7,8,31</sup> On May 8, 1900, British zoologist William Bateson<sup>32</sup> boarded a train bound for London to give a lecture on problems of heredity. During the course of that trip, he purportedly read the published work of Mendel. Bateson was a highly vocal supporter of Darwin’s theory of natural selection and needed a “genetic” explanation for the theory of evolution. Bateson believed that Mendel’s discoveries met this need, and through his writings and lectures, Bateson contributed substantially to the legend of Mendel in genetics. Of interest, Mendel’s principles of genetics were independently

rediscovered in 1900 by 3 different geneticists: De Vries,<sup>33</sup> Correns,<sup>34</sup> and Tschermak.<sup>35</sup> Among Bateson's contributions was the coining of the term *genetics*.

Once Mendel's work became known to scientists, examples of traits that segregated in a mendelian fashion were soon identified in both plants and animals. One of the first traits described led to the birth of biochemical genetics. Inherited biochemical disorders were first described in 1902 by Garrod<sup>36</sup> in his landmark article, "The Incidence of Alkaptonuria: A Study in Chemical Individuality." Garrod learned of Mendel's work in 1909 from Bateson, and this led him to describe alkaptonuria and several additional disorders in his book *Inborn Errors of Metabolism*.<sup>37</sup> Because of this work, Garrod is now considered the founder of biochemical genetics. In 1929, Richard Schönheimer studied a patient with hepatomegaly due to massive glycogen storage and suggested that this disorder may be due to an enzyme deficiency. It was not until 1952 that Cori and Cori<sup>38</sup> found glucose-6-phosphatase to be deficient in patients with von Gierke disease (glycogen storage disease type I). This observation was the first time that an inborn error of metabolism was attributed to a specific enzyme deficiency.

In the early 1900s, geneticists from the United States began to make important contributions to genetic research. Working at Columbia University in 1910, Morgan<sup>39</sup> conducted experiments in fruit flies, mainly *Drosophila melanogaster*, and established that some genetically determined traits were sex linked. In 1913, Bridges,<sup>40</sup> one of Morgan's students, established that genes were located on chromosomes. In that same year, Sturtevant,<sup>41</sup> another student, determined that genes were arranged on chromosomes in a linear fashion, like beads on a necklace. Moreover, Sturtevant showed that the gene for any specific trait was in a fixed location or locus. Yet another Morgan student, Muller,<sup>42</sup> in 1926 discovered methods for artificially producing mutants in fruit flies by ionizing radiation and other mutagens. In so doing, he discovered the origin of new genes by mutations, a theory first proposed by De Vries<sup>43</sup> in the early 1900s.

In 1928, Griffith<sup>44</sup> studied *Streptococcus pneumoniae* and learned that a "transforming principle" can be transferred from dead virulent bacteria to living nonvirulent bacteria. Many investigators at that time believed that the transforming principle was protein. In 1941, Beadle and Tatum<sup>45</sup> performed experiments that suggested "one gene codes for one enzyme." In 1944, the work of Griffith was continued by Avery et al<sup>28</sup> who showed that the transforming principle was DNA, and thus DNA was the hereditary material in most living cells.

Linus Pauling, who originally trained as a chemical engineer, had a strong belief in chemistry as a means to

explain biological phenomena. Through his work in chemistry, he contributed substantially to the field of genetics. In 1949, Pauling et al<sup>46</sup> published their work on sickle cell anemia. This article revealed that the specific physicochemical changes in the sickled cells were a result of the mutated gene. Their publication is considered the first description of a disease on a molecular basis. One year later, Pauling et al<sup>47</sup> were the first to describe the 3-dimensional ( $\alpha$ -helix) structure of proteins.

In 1950, McClintock<sup>48</sup> discovered "jumping genes" or transposable elements in corn. However, the importance of her work was not appreciated until the 1970s when molecular biologists confirmed the existence of a gene for "transposase."<sup>49,50</sup> This enzyme enabled certain genes to self-duplicate and migrate to other areas of the genome.

In 1950, Chargaff<sup>51,52</sup> discovered that the ratio of the nucleic acid bases, purines to pyrimidines (adenine to thymine, guanine to cytosine) was always 1:1. This observation provided strong evidence that the nucleic acid bases form complementary pairs within the DNA molecule. In 1951, Franklin, Wilkins, and associates produced x-ray diffraction patterns of the DNA molecule.<sup>53,54</sup> This work revealed the helical structure and location of the phosphate sugar on DNA. Hershey and Chase<sup>55</sup> in 1952 produced conclusive evidence that bacteriophage DNA contained all the information necessary to create a new virus particle, including its DNA and protein coat. In 1953, Watson and Crick<sup>6</sup> elucidated the structure of the DNA molecule to be a double helix. Subsequently, Crick<sup>56</sup> introduced the central dogma of genetics, ie, DNA makes RNA makes protein.

In 1934, Fölling<sup>57</sup> was the first to ascribe a form of mental retardation to a metabolic disturbance. In this disorder, now known as phenylketonuria, the clinical characteristics are caused by an elevated excretion of phenylpyruvic acid in urine as a result of a deficiency in phenylalanine hydroxylase,<sup>58</sup> which in turn is a result of a mutation of the gene encoding for the enzyme.<sup>59,60</sup> The development of a treatment strategy for phenylketonuria in the early 1950s by the provision of a diet low in phenylalanine is another milestone in the history of biochemical genetics. Because of the potential to prevent mental retardation in affected infants, Guthrie and Susi<sup>61</sup> developed a cost-effective screening method for phenylketonuria, using small blood spots dried on filter paper that were collected from newborns. Population-wide newborn screening for phenylketonuria began in the 1960s. As a result of important technological advances, more than 1000 inborn errors of metabolism are now recognized. The application of tandem mass spectrometry to newborn screening has allowed the expansion of the number of metabolic disorders detectable in a single dried blood spot to more than 30.

Meanwhile, cytogeneticists were busy trying to establish the number of chromosomes in humans. In 1922, Painter<sup>62</sup> reported that humans had 48 chromosomes, based on studies of human testicular sections. Then in the early 1950s, several technical advances in chromosome methodology were discovered, including the discovery by Hsu<sup>63,64</sup> of hypotonic solutions to spread the chromosomes. Armed with these new methods, Tijo and Levan<sup>65</sup> working in Lund, Sweden, in 1956 established the correct chromosome number in humans to be 46. In 1959, 3 important chromosome syndromes were discovered. In France, Lejeune et al<sup>66</sup> described trisomy 21 in Down syndrome. Working in England, Jacobs and Ford identified 45,X in Turner syndrome and 47,XXY in Klinefelter syndrome.<sup>67-69</sup> Collectively, these observations marked the birth of clinical cytogenetics.

Based on his studies of sea urchins in 1914, Boveri<sup>70</sup> proposed the theory that chromosome abnormalities cause cancer. The opinion that chromosomes are responsible for the normal functions of a cell, plus his observation that aberrant chromosome combinations in dispermic eggs cause damage to such cells, led Boveri to this hypothesis. In 1960, Nowell and Hungerford<sup>71</sup> in Philadelphia discovered that an abnormality of chromosome 22 (the Philadelphia chromosome) was associated with chronic myeloid leukemia. This discovery led to the birth of cancer cytogenetics.

In the 1960s and 1970s, important discoveries led to modern techniques in the study of genetics. In 1964, Temin<sup>72</sup> worked with RNA viruses and discovered that Crick's central dogma (ie, DNA makes RNA makes protein) was not always true. Subsequently, Temin and Mizutani<sup>73</sup> and Baltimore<sup>74</sup> independently discovered an enzyme called *reverse transcriptase*. This enzyme used RNA as a template for the synthesis of a complementary DNA strand. In the mid-1960s, Holley,<sup>75</sup> Khorana,<sup>76</sup> and Nirenberg and Leder<sup>77</sup> all contributed to the cracking of the genetic code by determining the DNA sequence for each of the 20 most common amino acids.

In 1970, restriction enzymes were discovered by Smith and Wilcox<sup>78</sup> at Johns Hopkins, while studying a bacterium, *Haemophilus influenzae*. In this organism, restriction enzymes were used to cut up foreign DNA from invading organisms such as viruses. In 1975, Southern<sup>79</sup> developed a method to isolate and analyze fragments of DNA. Today, this method is referred to as *Southern blot analysis* and is commonly used in genetic studies in research and clinical practice. In 1977, Sanger et al<sup>80</sup> and Maxam and Gilbert<sup>81</sup> independently developed techniques to determine the nucleic acid bases for long sections of DNA. One year later, Kan and Dozy<sup>82</sup> discovered restriction fragment length polymorphisms. In 1983, Mullis et al,<sup>83</sup> working with the Cetus Corporation in California, invented the

polymerase chain reaction. This method amplified small fragments of DNA to make sufficient quantities available for DNA sequence analysis.

With use of these new genetic techniques, several genes for important human disorders were discovered in the 1980s. In 1982, Gusella et al<sup>84</sup> began studying patients with Huntington chorea and determined that the gene for this clinical disorder was located on the short arm of chromosome 4. During that period, Stephens et al<sup>85</sup> mapped the gene for neurofibromatosis type I to the long arm of chromosome 17.

Tsui et al<sup>86</sup> mapped the gene for cystic fibrosis to the long arm of chromosome 7 in 1985. This work was furthered in 1989, when the gene for cystic fibrosis was discovered. This research determined that 3 missing nucleic acid bases occurred in the altered gene of 70% of patients with cystic fibrosis.<sup>87</sup> This mutation is  $\delta F508$ .

In 1986, Royer-Pokora et al<sup>88</sup> in Boston isolated the gene for chronic granulomatous disease by studying a patient with a chromosome abnormality in the short arm of the X chromosome. Koenig et al<sup>89</sup> identified mutations associated with Duchenne muscular dystrophy in 1987. This gene was located close to the gene for chronic granulomatous disease on the short arm of the X chromosome. In 1991, King<sup>90</sup> found the first evidence that a gene on chromosome 17 could potentially be associated with an inherited predisposition to breast cancer and ovarian cancer. Collectively, the work of these investigators between 1980 and 1991 lead to the birth of modern clinical molecular genetics.

During that period, numerous important discoveries were made in cytogenetics. In the early 1970s, a variety of cytogenetic methods were discovered that produced distinct bands on each chromosome. Caspersson et al<sup>91</sup> developed a banding method that used quinacrine mustard, which allowed accurate identification of all 22 autosomes and the X and Y chromosomes. Thus, recognition of extremely subtle structural abnormalities and specific identification of the chromosomes involved in aneuploid situations became possible. In 1971, several investigators, including Seabright,<sup>92</sup> developed chromosome banding techniques that used Giemsa stain and a light microscope to view G bands. In 1991, a system of chromosome nomenclature based on these bands led to a systematic procedure for identifying each of these chromosome bands.<sup>93</sup> For example, Xq13 indicated a band on the X chromosome, q arm, region 1, band 3. This created a language of communication for gene and chromosome locations that is currently in use throughout the world. Today, cytogenetic studies are applied in 3 broad areas of medicine: (1) congenital disorders, (2) prenatal diagnosis, and (3) neoplastic disorders, especially hematologic disorders.

In 1980, Bauman et al<sup>94</sup> attached fluorophores to specific DNA bases to perform in situ hybridization and visualized individual chromosome loci. Between 1986 and 1988, Pinkel et al<sup>95</sup> working in California used this method with chromosome-specific DNA sequences to detect trisomy 21 and translocations involving chromosome 4. Many other investigators soon developed other fluorescent-labeled DNA probes and other techniques to detect various chromosome abnormalities and gene mutations in patients with a variety of congenital disorders and neoplastic diseases. Collectively, this work gave birth to modern molecular cytogenetics.

### LAYING THE FOUNDATION FOR THE SEQUENCING OF THE HUMAN GENOME

In 1953, Watson and Crick<sup>6</sup> elucidated the double helical structure of DNA. Their proposed DNA structure was published in *Nature* on April 25, 1953, as a brief article that barely exceeded 1 page. This article heralded a new age of discovery in genetics and laid the foundation for the sequencing of the human genome.

Today the term *genome* is widely understood to be the genetic material of an organism. However, this common term existed well before modern genetics. The word *genome* was coined by Winkler in 1920 and is defined as the haploid set of chromosomes and all the genes they contain. Genome is derived from parts of 2 other important words: *gene* and *chromosome*. With advances in genetic research, we now further define a genome to be composed of a series of nitrogenous DNA bases (adenine [A], guanine [G], thymine [T], or cytosine [C]). In each organism, these bases are arranged in a specific order, and this order is the genetic code of the organism. In humans, the genome is made up of approximately 3 billion such bases. In 2001, a first draft sequence of the entire human genome was completed and made available to the public for study and research.<sup>7,8</sup>

The first draft of the human genome was greeted with great anticipation by geneticists and nongeneticists alike. Understandably, every person had some vested interest in this scientific endeavor because the material that the researchers were working on was the genetic code of the human species. The genetics community devoured detailed accounts of the project through record-length journal articles with numerous acronyms. The general public was informed of the sequencing of the human genome through extensive media coverage. Although the geneticists could understand the details and the future implications of this feat better than laypersons, the importance of this achievement was not lost on anyone, regardless of his or her professional training. It was understood that the sequence of the human genome held the answers to our uniqueness as

a species and that this information would likely be the basis for important future biomedical achievements.

The formal HGP was conceived fewer than 20 years ago, yet in that short time frame, it has become one of humankind's crowning achievements. Although it has occurred in such recent times, few individuals outside the genetics community know the history of the HGP.

By the mid-1980s, decades after Watson and Crick's historic discovery, technology had advanced far enough to make the idea of sequencing the human genome seem plausible.<sup>96</sup> These advances included the discovery of recombinant DNA in the mid 1970s, which allowed the cloning of DNA fragments hundreds to several thousands of base pairs in length.<sup>97</sup> Then the use of yeast artificial chromosomes expanded the size of the DNA fragments that could be cloned to hundreds of thousands of base pairs.<sup>98</sup> Rapid DNA sequencing and analysis became more of a reality in 1985 after Mullis et al<sup>83</sup> developed the polymerase chain reaction. By 1987, automated sequencers were developed, allowing work to be done more rapidly on large segments of DNA.

In the United States, the first formal meeting to discuss the feasibility of sequencing the human genome occurred in June 1985.<sup>96</sup> This meeting was convened by Robert Sinsheimer, then chancellor of the University of California at Santa Cruz. The concept of sequencing the human genome excited many of the prominent researchers in attendance and generated discussion of such a project within the scientific community. Separately, in late 1985, the director of health and environmental research at the Department of Energy (DOE), Charles DeLisi, was charged with assessing the utility of DNA sequencing in detecting induced mutations in survivors of Hiroshima and Nagasaki. In 1987, DeLisi started the Human Genome Initiative, the first government program on genome research. The US Congress began gathering data on human genome research in 1986. However, Congress did not decide to provide funding until 1988, after it concluded that the formation of administrative centers accountable to Congress could most efficiently manage issues such as databases, sharing of research materials, and cultivation of new technologies. Congress started by funding \$17.3 million to the National Institutes of Health (NIH) and \$11.8 million to DOE. These budgets increased progressively over the next few years. The NIH established the Office of Human Genome Research in September 1988. One year later, this office was renamed the National Centre for Human Genome Research (NCHGR). James Watson was the NCHGR's director until April 1992.<sup>3</sup> In this role, he pledged that 3% to 5% of the project's budget would be devoted to address ethical, legal, and social issues that arose from the study of the human genome. This program, the Ethical, Legal and Social Issues

program, was the largest bioethics program in the world. Watson left the NCHGR after some disagreements over issues related to gene patenting. Francis Collins was named director of the NCHGR in April 1993.<sup>99</sup>

The United States was not alone in this important endeavor. Soon after the initiation of the DOE project in 1987, the Italian National Research Council started a pilot project on genome research composed of 15 groups throughout Italy. The United Kingdom launched its project in February 1989. Then in 1990, the European Commission initiated a 2-year human genome project. France's government project followed months later in June 1990. Researchers from the former Union of Soviet Socialist Republics (USSR) gained funding in 1989. Although the program in the USSR was already one of the largest programs at that time, it expanded 1 year later in 1990.<sup>96</sup>

### FORMAL INITIATION OF THE HGP

Despite its relatively brief history and skepticism that it would fail, the HGP has seen several remarkable successes. When the HGP was first proposed in the 1980s, there was little support for pouring billions of dollars into another big science project with few obvious benefits. Because the human genome was estimated to consist of approximately 3 billion bases, completion of the project was likely to be expensive and labor intensive. At that time, the cost of a finished sequence was approximately \$10 per base. A well-equipped laboratory could produce about 500 bases of sequence a day, and solid data indicated that more than 90% of the 6 feet of DNA in every cell was "junk" (ie, repetitive sequences with either no known function or noncoding "spacer" DNA). Despite well-founded skepticism at that time, it is now difficult to find someone who admits to opposing the project in the early days. An engaging summary of the personalities and controversies surrounding the early days of the project was published in the genome issue of *Science* (*Science*. 2001;291:1145-1434).

The international effort to sequence the human genome was initiated in 1990 and was named the Human Genome Project.<sup>100</sup> The HGP originally developed a 15-year plan to map and sequence the human genome. The plan outlined several goals including (1) development of high-resolution genetic and physical maps of the human genome; (2) determination of the complete DNA sequence of humans and several other model organisms; (3) development of the capability for storing, analyzing, and interpreting these data (now called *bioinformatics*); and (4) development of the technology necessary to meet these goals, as well as assessment of the ethical, legal, and social implications of genomics. Objectives relating to research training and technology transfer were also elaborated at that time.

Rapid progress (particularly in constructing genetic and physical maps) led to the extension and revision of the original goals, and a "5-year plan" was released in October 1993.<sup>101</sup> These revisions were mostly incremental, detailing revised and more ambitious goals, including the completion of 80 million bases of sequence (<3% of the total size of the human genome) for all targeted organisms by 1998.

Composed of more than 2000 scientists from 20 institutions in 6 countries, the collaborators in the sequencing of the human genome were collectively known as the International Human Genome Sequencing Consortium. The funding for the HGP came from various countries as government grants and public charities. In 1990 when the HGP was established, it was estimated that \$3 billion (US dollars) would be needed to complete the sequencing and to support the other activities related to the HGP. At the completion of the draft sequence, approximately \$300 million (US dollars) had been spent worldwide.<sup>100</sup>

The international partners in the genome project met in Bermuda in February 1996. At that meeting, they created the "Bermuda principles," a set of conditions that pertain to data access, including the release of sequence data into public databases within 24 hours. Because of this agreement, participating scientists deposited base sequences into 1 of 3 databases within 24 hours of sequencing completion. The data between the 3 databases were exchanged daily. Since these were public databases, access to the stored sequences was unrestricted.

The early 1990s also saw the beginning of a privately funded genomics effort, championed by J. Craig Venter. Venter was running a successful laboratory at the NIH when he decided to focus his sequencing efforts not on the genome itself, but on the gene products, messenger RNAs produced by each cell. In addition to the genes themselves, the genome is composed of a much larger amount of DNA whose function remains to be observed. This is often derisively referred to as *junk DNA*; however, important functions may eventually be ascribed to this portion of the genome. Venter began a large-scale project to sequence expressed sequence tags as a way of trying to identify new gene products and to estimate the degree of diversity in gene structure and function. (Expressed sequence tags are random fragments of complementary DNAs derived from the information in the RNAs, which contain all the information that is actually expressed in a given cell type.) Although this approach drew criticism for being too much of a departure from the gene-by-gene strategy that was standard at that time, it was the forerunner of the large-scale genomics experiments that have become standard.

Venter had some initial success and eventually left the NIH to establish a nonprofit organization, The Institute for Genomic Research, whose goal was to collect and interpret

vast numbers of expressed sequence tags. The informatics component of The Institute for Genomic Research was so robust that Venter and colleagues realized that vast quantities of short sequence fragments could be assembled into long contiguous strings of sequence after the fact, simply by having computers analyze the overlap between the small sequence fragments. Although NIH study sections refused to support this approach to large-scale sequencing, Venter and colleagues proved the utility of this method by determining the 1.8 million base complete genomic sequence of a free living organism, *H influenzae*.<sup>102</sup> This project, which took less than a year, was completed in 1995. By 1998, Venter had demonstrated enough successes with his approach that he announced the formation of a privately funded company, Celera Corporation, that would produce the entire human genome sequence within 3 years for a fraction of the cost of the publicly funded program. Additionally, Venter stated that the release of Celera's data would not be shared with the international databases, and therefore it would not act in accordance with the Bermuda principles. However, he added that the public will have "free access" to the Celera's sequence as long as certain provisions are met.

This announcement spurred a final "5-year plan" from the publicly funded project late in 1998.<sup>101</sup> The accelerated timetable not only moved the target date for completed sequencing of the genome up to 2003 (the 50th anniversary of the discovery of the double helix) but also projected that 90% of the genome would be completed in draft form by 2001. Of importance, several new goals were also introduced in this version, including the cataloging of sequence variation in the human genome and the complete sequencing of several other genomes (*Caenorhabditis elegans*, *D melanogaster*, and mouse).

The first human chromosome to be completely sequenced was chromosome 22, completed in a collaborative effort by scientists at the Sanger Institute in England, at the University of Oklahoma and at Washington University in St Louis in the United States, and at Keio University in Japan. Its sequence was published in December 1999.<sup>103</sup> Five months later, in May 2000, the sequence of chromosome 21 was published from a collaborative effort by German and Japanese groups.<sup>104</sup> Chromosome 20 was sequenced by the end of December 2001.<sup>105</sup> Like chromosome 22, chromosome 20 was also sequenced by collaborators at the Sanger Institute.

## ACHIEVEMENTS

By any measure, the genome project has been an outstanding success, surpassing the most optimistic projections of progress and costing far less than originally expected. The completion of the draft human genome sequence by both

the public project and the Celera project was announced in special issues of *Nature* and *Science* in February 2001.<sup>7,8</sup> By the end of 2001, nearly half of the genome had been deposited in "finished" form (<1 error every 10,000 base pairs), and all but 1.5% of the targeted sequence is present in GenBank (the public database) as either finished or draft form. Data from the public project are accessible through the National Center for Biotechnology Information Web site (<http://www.ncbi.nlm.nih.gov/>).

Other goals of the HGP also have been met or exceeded. Complete genomic sequences for *Escherichia coli*, *Saccharomyces cerevisiae* (bakers' yeast), *C elegans* (roundworm), *D melanogaster* (fruit fly), and many other microorganisms have been determined and published. As of this writing, the mouse genome is 96% complete in draft form,<sup>106</sup> and projects to complete the rat and zebra fish sequences have been initiated. Impressive progress has been made in detailing the nature and extent of sequence variation in humans. The initial goal for the cataloging of single nucleotide polymorphisms (SNPs) (variations in single bases of DNA detected in the population at large) was to create a map with 100,000 SNPs by 2003. As of October 2001, dbSNP (the repository for these data) had more than 4 million SNPs, which are actively being used by researchers in both academia and industry to identify genes that contribute to disease susceptibility and drug response. These developments promise to dramatically change the way in which medicine is practiced in the 21st century.

## FUTURE GOALS

The next important challenge is to make sense of all the data generated by the genome project. One of the major surprises from early analysis was that previous estimates of gene number appear to be inaccurate by a substantial margin. Most pregenome estimates agreed that humans had between 60,000 and 100,000 genes (although some estimates ranged as high as 120,000<sup>107</sup>). Early analysis of the complete genome sequence suggests that the true number of genes required to make a person may be only 30,000 to 40,000. The difficulty in finding actual gene sequences in 3 billion bases of complex data is underscored by a recent analysis that revealed the 31,780 genes predicted by the public project only partially overlap with the 39,114 genes identified by Celera.<sup>108</sup> Adding to the controversy is a recent computational reanalysis of the "completed" public sequence that concluded there might be as many as 75,000 genes in humans.<sup>109</sup> Thus, there is reason to believe that the human gene number may be greater than 30,000. Clearly, one of the immediate goals is to improve the informatics capability to allow more accurate and meaningful analysis of the sequence data.

Whatever the final number of genes in the human genome, a more important goal is to understand the function of each gene. The function of approximately 15,000 such gene products is known. Thus, despite years of study, we understand the function of fewer than half of the human genes. Do some of the unknown genes code for products that will help us better understand cancer or heart disease? What new drug targets remain to be discovered? Now that the DNA sequence is largely complete, there is interest in trying to understand not only the function of genes (genomics) but also the function of proteins (proteomics) and how groups of proteins in common pathways combine to produce physiological responses (metabolomics). These efforts are already under way and will lead to insights about cell physiology.

Another area of intensive work is to understand how differences in human DNA sequences affect our daily lives. It is clear that common minor changes in DNA sequence can affect not only susceptibility to rare genetic disease but also many more common conditions. As a preview of what lies ahead, information from the genome project was recently used to identify susceptibility genes for non-insulin-dependent diabetes mellitus<sup>110</sup> and inflammatory bowel disease.<sup>111</sup>

Applications will not be limited to prognostic indicators and diagnosis. The NIH recently launched a program to correlate sequence variations with drug responses (pharmacogenomics). The long-term goal of the effort is to tailor drug therapies to the individual patient.

The wealth of genetic information has raised some concerns as well, including the confidentiality of genetic information and the possible stigmatization of people based on their DNA sequences. These issues deserve continued study and scrutiny.

## CONCLUSION

Humankind's interest and study of genetic-related concepts are recorded throughout history. However, the greatest advances in the field have occurred within the past 150 years. The future promises more progress in genetics and in related areas of medicine and science. An early statement by a well-known geneticist, Alfred Sturtevant, in regard to the future of genetics is as applicable now as it was when he wrote it in the last century: "We should like to conclude the account with an outline of the future of genetics; but one of the things that makes science interesting is that new developments are likely to come in unexpected directions. One thing can safely be predicted; genetics will continue to develop and to include new fields of work—it is not static."<sup>112</sup>

The completion of the sequencing of the human genome will surely lead to further progress in genetics, science,

medicine, and other related fields. The greatest benefits from the sequencing of the human genome are yet to be realized. Although continued advances in genetics research will be recognized most immediately by those working in genetics, the achievements and their social and ethical implications will affect all humanity. An understanding of the history of genetics and how we have arrived at where we are today will help prepare us to meet tomorrow's challenges.

## REFERENCES

1. Capelle W. *Die Vorsokratiker*. Stuttgart, Germany: Kroner; 1953.
2. Barthelmess A. *Vererbungswissenschaft*. Freiburg, Germany: Alber; 1952.
3. Watson JD. The human genome project: past, present, and future. *Science*. 1990;248:44-49.
4. Aristotle. *De Generatione Animalium*. Bk 2, 1; *Historia Animalium*. Bk 6, 37.
5. Mendel G. Experiments in plant hybridization. *Verh Naturforsch Vereines Brunn*. 1865;4:3-47.
6. Watson JD, Crick FHC. Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature*. 1953;171:737-738.
7. Lander ES, Linton LM, Birren B, et al. International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome [published corrections appear in *Nature*. 2001;412:565 and 2001;411:720]. *Nature*. 2001;409:860-921.
8. Venter JC, Adams MD, Myers EW, et al. The sequence of the human genome [published correction appears in *Science*. 2001; 292:1838]. *Science*. 2001;291:1304-1351.
9. Stubbe H. *History of Genetics, From Prehistoric Times to the Rediscovery of Mendel's Laws*. Waters TRW, trans. Cambridge, Mass: MIT Press; 1972.
10. Plato. *Politikos*. Sec 310.
11. Paracelsus. *Etliche Tractatus des Hohergarnen. I. Von natürlichen Dingen*. Strasbourg, France; 1570.
12. Hartsoeker N. *Essay de Dioptrique*. Paris, France: J Anisson; 1694.
13. Cole FJ. *Early Theories of Sexual Generation*. Oxford, England: Clarendon Press; 1930.
14. Needham J. *A History of Embryology*. 2nd ed. New York, NY: Abelard-Schuman; 1959.
15. Malpighi M. *De Formatione Pulli in Ovo*. Londoni: Apud Joannem Martyn; 1673.
16. van Leeuwenhoek A. Of the generation of animals. *Philos Trans*. 1724;32:438-440.
17. Maupertuis PLM. *Systeme de la Nature: Essai sur la Formation des Corps Organises*. Erlangen, Germany; 1751.
18. Bowler PJ. *The Mendelian Revolution: The Emergence of Hereditarian Concepts in Modern Science and Society*. Baltimore, Md: Johns Hopkins University Press; 1989.
19. Maupertuis PLM. *Lettres de M. de Maupertuis*. Dresden, Germany: Chez G C Walther; 1752.
20. Motulsky AG. Joseph Adams (1756-1818). *Arch Intern Med*. 1959;104:490-496.
21. Hooke R. *Micrographia*. London, England; 1665.
22. Wolff CF. *Theoria Generationis*. Berlin, Germany; 1759.
23. Schleiden M, Schwann T. *Beitrage zur Phytogenesis*. Hamburg, Germany; 1838.
24. Virchow R. *Die Cellularpathologie*. Berlin, Germany; 1858
25. Darwin C. *The Origin of Species by Means of Natural Selection*. London, England: John Murray; 1859.
26. Portugal FH, Cohen JS. *A Century of DNA: A History of the Discovery of the Structure and Function of the Genetic Substance*. Cambridge, Mass: MIT Press; 1977.

27. Miescher F. Basel, Switzerland; 1869.
28. Avery O, MacLeod C, McCarty M. Studies on the chemical nature of the substance inducing transformation of pneumococcal type. *J Exp Med.* 1944;79:137-158.
29. Flemming W. *Zellsubstanz: Kern & Zelltheilung.* Leipzig, Germany: FCW Vogel; 1882.
30. Weismann A. *Essays Upon Heredity and Kindred Biological Problems.* Oxford, England: Clarendon Press; 1889.
31. Johannsen W. *Elemente der exakten Erblchkeitslehre.* Jena, Germany: Gustav Fischer; 1909:143-144.
32. Bateson W. Problems of heredity as a subject for horticultural investigation. *J R Hort Soc.* 1900;25:54-61.
33. De Vries H. Sur la loi de disjonction des hybrides. *Compt Rend Acad Sci (Paris).* 1900;130:845-847.
34. Correns C. G. Mendel's Regel uber das verhalten der Nachkommenschaft der rassen Bastarde. *Ber Dtsch Botan Ges.* 1900;18:158-168.
35. Tschermak E. Uber kunstliche Kreuzung bei pisum Sativum. *Ber Dtsch Botan Ges.* 1900;18:232-239.
36. Garrod AE. The incidence of alkaptonuria: a study in chemical individuality. *Lancet.* 1902;2:1616-1620.
37. Garrod AE. *Inborn Errors of Metabolism.* 2nd ed. London, England: Henry Frowde and Hodder & Stoughton; 1923.
38. Cori GT, Cori CF. Glucose-6-phosphatase of the liver in glycogen storage disease. *J Biol Chem.* 1952;199:661-667.
39. Morgan TH. Sex-limited inheritance in *Drosophila*. *Science.* 1910;32:120-122.
40. Bridges CB. Direct proof through non-disjunction that the sex-linked genes of *Drosophila* are borne on the X-chromosome. *Science.* 1914;40:107-109.
41. Sturtevant AH. The linear arrangement of six sex-linked factors in *Drosophila*, as shown by their mode of association. *J Exp Biol.* 1913;14:43-59.
42. Muller HJ. Artificial transmutation of the gene. *Science.* 1927;66:84-87.
43. de Vries H. *Intracellular Pangenesis.* Chicago, Ill: The Open Court Publishing Co; 1910.
44. Griffith F. The significance of pneumococcal types. *J Hyg.* 1928;27:113-159.
45. Beadle GW, Tatum E. Genetic control of biochemical reactions in neurospora. *Proc Natl Acad Sci U S A.* 1941;27:499-506.
46. Pauling L, Itano HA, Singer SJ, Wells IC. Sickle cell anemia: a molecular disease. *Science.* 1949;110:543-548.
47. Pauling L, Corey RB, Branson HR. The structure of proteins: two hydrogen-bonded helical configurations of the polypeptide chain. *Proc Natl Acad Sci U S A.* 1951;37:205-211.
48. McClintock B. The origin and behavior of mutable loci in maize. *Proc Natl Acad Sci U S A.* 1950;36:344-355.
49. Cohen SN, Shapiro JA. Transposable genetic elements. *Sci Am.* 1980;242:40-49.
50. Roeder GS, Farabaugh PJ, Chaleff DT, Fink GR. The origins of gene instability in yeast. *Science.* 1980;209:1375-1380.
51. Chargaff E. Chemical specificity of nucleic acids and mechanism of their enzymatic degradation. *Experientia.* 1950;6:201-209.
52. Chargaff E. Structure and function of nucleic acids as cell constituents. *Fed Proc.* 1951;10:654-659.
53. Franklin RE, Gosling RG. Molecular configuration in sodium thymonucleate. *Nature.* 1953;171:740-741.
54. Wilkins MHF, Stokes AR, Wilson HR. Molecular structure of deoxyribose nucleic acids. *Nature.* 1953;171:738-740.
55. Hershey AD, Chase M. Independent functions of viral protein and nucleic acid in growth of bacteriophage. *J Gen Physiol.* 1952;36:39-56.
56. Crick F. On protein synthesis. *Symp Soc Exp Biol.* 1966;31:3-9.
57. Fölling A. Ueber Ausscheidung von Phenylbrenztraubensaure in den Harn als Stoffwechselerkrankung in Verbindung mit Imbezillitaet. *A Physiol Chem.* 1934;277:169.
58. Jervis GA. Phenylpyruvic oligophrenia deficiency of phenylalanine-oxidizing system. *Proc Soc Exp Biol Med.* 1953;82:514-515.
59. Woo SL, Guttler F, Ledley FD, et al. The human phenylalanine hydroxylase gene. *Prog Clin Biol Res.* 1985;177:123-135.
60. Woo SL, DiLella AG, Marvit J, Ledley FD: Molecular basis of phenylketonuria and potential somatic gene therapy. *Cold Spring Harb Symp Quant Biol.* 1986;1986;51(pt 1):395-401.
61. Guthrie R, Susi A. A simple phenylalanine method for detecting phenylketonuria in large populations of newborn infants. *Pediatrics.* 1963;32:338-343.
62. Painter TS. Studies in mammalian spermatogenesis, II: the spermatogenesis of man. *J Exp Zool.* 1923;37:291-336.
63. Hsu TC. Mammalian chromosomes in vitro, I: the karyotype of man. *J Hered.* 1952;43:167-172.
64. Hsu TC. *Human and Mammalian Cytogenetics: An Historical Perspective.* New York, NY: Springer-Verlag; 1979.
65. Tjio JH, Levan A. The chromosome number of man. *Am J Obstet Gynecol.* 1956;130:723-724.
66. Lejeune J, Gautier M, Turpin R. Etude des chromosomes somatiques de neuf enfants mongoliens. *Compt Rend.* 1959;248:1721-1722.
67. Jacobs PA, Strong JA. A case of human intersexuality having a possible XXY sex-determining mechanism. *Nature.* 1959;183:302-303.
68. Jacobs PA, Baikie AG, MacGregor TN, Harnden DG. Evidence for the existence of the human "superfemale." *Lancet.* 1959;2:423-425.
69. Ford CE, Miller OJ, Polani PE, de Almeida JC, Briggs JH. A sex-chromosome anomaly in a case of gonadal dysgenesis (Turner's syndrome). *Lancet.* 1959;1:711-713.
70. Boveri T. *The Origin of Malignant Tumors.* Boveri M, trans. Baltimore, Md: Williams & Wilkins; 1929.
71. Nowell PC, Hungerford DA. A minute chromosome in human chronic granulocytic leukemia. *J Natl Cancer Inst.* 1960;25:85.
72. Temin HM. Homology between RNA from Rous sarcoma virus and DNA from Rous sarcoma virus-infected cells. *Proc Natl Acad Sci U S A.* 1964;52:323-329.
73. Temin HM, Mizutani S. RNA-dependent DNA polymerase in virions of Rous sarcoma virus. *Nature.* 1970;226:1211-1213.
74. Baltimore D. RNA-dependent DNA polymerase in virions of RNA tumour viruses. *Nature.* 1970;226:1209-1211.
75. Holley RW. Structure of a ribonucleic acid. *Science.* 1965;147:1462-1465.
76. Khorana HG. Nucleic acid synthesis in the study of the genetic code. In: *Nobel Lectures: Physiology or Medicine (1963-1970).* New York, NY: Elsevier; 1973.
77. Nirenberg MW, Leder P. RNA codewords and protein synthesis. *Science.* 1964;145:1399-1407.
78. Smith HO, Wilcox KW. A restriction enzyme from *Hemophilus influenzae*, I: purification and general properties. *J Mol Biol.* 1970;51:379-391.
79. Southern EM. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol.* 1975;98:503-517.
80. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A.* 1977;74:5463-5467.
81. Maxam AM, Gilbert W. A new method for sequencing DNA. *Proc Natl Acad Sci U S A.* 1977;74:560-564.
82. Kan YW, Dozy AM. Polymorphism of DNA sequence adjacent to human beta-globin structural gene: relationship to sickle mutation. *Proc Natl Acad Sci U S A.* 1978;75:5631-5635.
83. Mullis K, Faloona F, Scharf S, Saiki R, Horn G, Erlich H. Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harb Symp Quant Biol.* 1986;51(pt 1):263-273.
84. Gusella JF, Gilliam TC, Tanzi RE, et al. Molecular genetics of Huntington's disease. *Cold Spring Harb Symp Quant Biol.* 1986;51(pt 1):359-364.
85. Stephens K, Riccardi VM, Rising M, et al. Linkage studies with chromosome 17 DNA markers in 45 neurofibromatosis 1 families. *Genomics.* 1987;1:353-357.

86. Tsui LC, Buchwald M, Barker D, et al. Cystic fibrosis locus defined by a genetically linked polymorphic DNA marker. *Science*. 1985;230:1054-1057.
87. Lemna WK, Feldman GL, Kerem B, et al. Mutation analysis for heterozygote detection and the prenatal diagnosis of cystic fibrosis. *N Engl J Med*. 1990;322:291-296.
88. Royer-Pokora B, Kunkel LM, Monaco AP, et al. Cloning the gene for the inherited disorder chronic granulomatous disease on the basis of its chromosomal location. *Cold Spring Harb Symp Quant Biol*. 1986;51(pt 1):177-183.
89. Koenig M, Hoffman EP, Bertelson CJ, Monaco AP, Feener C, Kunkel LM. Complete cloning of the Duchenne muscular dystrophy (DMD) cDNA and preliminary genomic organization of the DMD gene in normal and affected individuals. *Cell*. 1987;50:509-517.
90. King MC. Localization of the early-onset breast cancer gene. *Hosp Pract (Off Ed)*. 1991;26:121-126.
91. Caspersson T, Zech L, Modest EJ. Fluorescent labeling of chromosomal DNA: superiority of quinacrine mustard to quinacrine. *Science*. 1970;170:762.
92. Seabright M. A rapid banding technique for human chromosomes. *Lancet*. 1971;2:971-972.
93. Mitelman F, ed. *ISCN 1995: An International System for Human Cytogenetic Nomenclature (1995): Recommendations of the International Standing Committee on Human Cytogenetic Nomenclature, Memphis, Tennessee, USA, October 9-13, 1994*. Basel, Switzerland: S. Karger; 1995.
94. Bauman JG, Wiegant J, Van Duijn P, et al. Rapid and high resolution detection of in situ hybridisation to polytene chromosomes using fluorochrome-labeled RNA. *Chromosoma*. 1981;84:1-18.
95. Pinkel D, Landegent J, Collins C, et al. Fluorescence in situ hybridization with human chromosome-specific libraries: detection of trisomy 21 and translocations of chromosome 4. *Proc Natl Acad Sci U S A*. 1988;85:9138-9142.
96. Watson JD, Cook-Deegan RM. Origins of the Human Genome Project. *FASEB J*. 1991;5:8-11.
97. Davies KE. The application of DNA recombinant technology to the analysis of the human genome and genetic disease. *Hum Genet*. 1981;58:351-357.
98. Chumakov IM, Rigault P, Le Gall I, et al. A YAC contig map of the human genome. *Nature*. 1995;377(6547, suppl):175-297.
99. Collins F, Galas D. A new five-year plan for the U.S. Human Genome Project. *Science*. 1993;262:43-46.
100. Dennis C, Gallagher R, eds. *The Human Genome*. New York, NY: Nature/Palgrave; 2002.
101. Collins FS, Patrinos A, Jordan E, Chakravarti A, Gesteland R, Walters L. New goals for the U.S. Human Genome Project: 1998-2003. *Science*. 1998;282:682-689.
102. Smith HO, Tomb JF, Dougherty BA, Fleischmann RD, Venter JC. Frequency and distribution of DNA uptake signal sequences in the *Haemophilus influenzae* Rd genome. *Science*. 1995;269:538-540.
103. Dunham I, Shimizu N, Roe BA, et al. The DNA sequence of human chromosome 22 [published correction appears in *Nature*. 2000;404:904]. *Nature*. 1999;402:489-495.
104. Hattori M, Fujiyama A, Taylor TD, et al. Chromosome 21 Mapping and Sequencing Consortium. The DNA sequence of human chromosome 21. *Nature*. 2000;405:311-319.
105. Deloukas P, Matthews LH, Ashurst J, et al. The DNA sequence and comparative analysis of human chromosome 20. *Nature*. 2001;414:865-871.
106. Check E. Draft mouse genome makes public debut. *Nature*. 2002;417:106.
107. Liang F, Holt I, Perlea G, Karamycheva S, Salzberg SL, Quackenbush J. Gene index analysis of the human genome estimates approximately 120,000 genes [published correction appears in *Nat Genet*. 2000;26:501]. *Nat Genet*. 2000;25:239-240.
108. Hogenesch JB, Ching KA, Batalov S, et al. A comparison of the Celera and Ensembl predicted gene sets reveals little overlap in novel genes. *Cell*. 2001;106:413-415.
109. Wright FA, Lemon WJ, Zhao WD, et al. A draft annotation and overview of the human genome. *Genome Biol*. 2001;2:research0025.1-research0025.18.
110. Horikawa Y, Oda N, Cox NJ, et al. Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus [published correction appears in *Nat Genet*. 2000;26:502]. *Nat Genet*. 2000;26:163-175.
111. Hugot JP, Chamaillard M, Zouali H, et al. Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature*. 2001;411:599-603.
112. Sturtevant AH. *A History of Genetics*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press; 2001. Available at: [www.esp.org/books/sturt/history/title3.html](http://www.esp.org/books/sturt/history/title3.html). Accessibility verified June 26, 2002.