Interview with Francis Collins
February 2003

Dr. Francis Collins
Photo courtesy of NHGRI

Francis S. Collins, M.D., Ph.D.
Director, National Human Genome Research Institute
National Institutes of Health
Bethesda, Maryland

Francis S. Collins, M.D., Ph.D., is the Director of the National Human Genome Research Institute at the National Institutes of Health. He oversees the Human Genome Project, a complex, multidisciplinary scientific enterprise directed at mapping and sequencing all of the human DNA, and determining aspects of its function. A working draft of the human genome sequence was completed in April 2003. From the outset, the project has run ahead of schedule and under budget, and all data has been made immediately available to the scientific community, without restriction on access or use.

Dr. Collins received a B.S. from the University of Virginia, a Ph.D. in Physical Chemistry from Yale, and an M.D. from the University of North Carolina. Following a fellowship in Human Genetics at Yale, he joined the faculty at the University of Michigan, were he remained until moving to NIH in 1993. His research led to the identification of genes responsible for cystic fibrosis, neurofibromatosis, and Huntington's disease. He is a member of the Institute of Medicine and the National Academy of Sciences.

Joseph McInerney, Executive Director of BSCS from 1985 to 1999 and now the Director of National Coalition for Health Professional Education in Genetics (NCHPEG), first met Francis Collins in the early 1990s, when McInerney was a member of the HGP's Working Group on Ethical, Legal, and Social Implications of Human Genome Research(ELSI program). While at BSCS, McInerney worked on genome modules developed under grants from Department of Energy's ELSI program. He was also a member of ERPEG, the ELSI Research, Planning, and Evaluation Group, which wrote the five-year plan for the ELSI program. Dr. Collins serves as the chairman of NCHPEG's board of directors. McInerney posed the following questions to Dr. Collins in February 2003.

This year, the BSCS newsmagazine, *The Natural Selection,* is celebrating the 50th anniversary of the classic paper by Watson and Crick in 1953, and the completion of the Human Genome Project in April 2003. As you think about the intervening 50 years, what do you regard as the most important scientific insights and technological breakthroughs that allowed us to go from the double helix to the sequence of the human genome?

There are many important milestones between Watson's and Crick's discovery of the double helix and the completion of the sequence of the human genome, and I fear that a complete enumeration would occupy many pages. Particular advances that had a profound impact would include the discovery of messenger RNA, the determination of the genetic code by which RNA can be translated into protein, the development of recombinant DNA technology, the development of efficient methods for sequencing DNA, first manually and then using automated machines, and the discovery of the polymerase chain reaction (PCR).

What were the most significant scientific and technological barriers the genome community had to overcome to meet the goals of the HGP?

The main challenge was simply the scale. Prior to the initiation of the HGP, even an excellent molecular biology laboratory would be hard pressed to generate more than a thousand base pairs of high-quality DNA sequence in a week. To sequence the 3 billion letters of the human genome, it was necessary to scale up to the point of being able to sequence a thousand base pairs every second, seven days a week, 24 hours a day. But there were many other technological challenges that had to be met, such as the ability to purify very large fragments of DNA in a stable form, the development of efficient methods for generating genetic and physical maps, and the evolution of an entirely

continued on page 6

new field of computational biology to record and analyze the prodigious amounts of data being generated by the HGP.

Was there ever a time when you had serious doubts about the feasibility or ultimate success of the HGP? If so, what were the sources of those concerns and how did you deal with them?
Yes, I had doubts all along the way. As originally proposed in 1990, the goals of the Human Genome Project were exciting, but the pathway to reach them was completely unclear. When I arrived at NIH in 1993, my reputation and that of hundreds of other scientists working on this project were on the line to make good on these bold promises. Yet, the hoped-for major breakthroughs in technology were hard to come by. Halfway through the 15-year period, we had only sequenced a tiny percentage of the human genome. Other major concerns swirled around the project, including whether all the data might end up in private databases (a very unappealing prospect, and one the HGP fought against from the very beginning), and whether the funding would be available to support the project. It took the combined creativity, energy, and optimism of hundreds of scientists working together in 20 laboratories around the world to make the dream come true.

It took the combined creativity, energy, and optimism of hundreds of scientists working together in 20 laboratories around the world to make the dream come true.

The HGP is remarkable in many ways, not least for its involvement of agencies, laboratories, and scientists from all over the world. Please outline some of the major players and their roles.
The HGP evolved through several phases. Of course, none of this could have ever gotten off the ground without the leadership of Jim Watson in the first two years of the genome project, using his considerable skills to successfully frame the effort in the early days. In the first five years, the emphasis was largely on mapping, with profound contributions from Maynard Olson, Jim Weber, and our French colleague Jean Weissenbach, who made the first really detailed genetic map of human DNA. As the emphasis shifted toward sequencing, John Sulston and Bob Waterston, colleagues from across the Atlantic, encouraged us all by their work on the roundworm genome. When the full-scaled sequencing of the human genome got under way, the major five laboratories (affectionately referred to as the G-5) were led by John Sulston, Bob Waterston, Eric Lander, Richard Gibbs, and Elbert Branscomb.

The most highly touted benefits of the HGP are those related to improved understanding of health and disease. How is genomics likely to improve human health, and how will it influence health care in this country?
The study of the genome will reveal over the next decade the hereditary factors that contribute to virtually all common diseases, including diabetes, cancer, heart disease, mental illness, hypertension, and many others. It will then become possible through those interested in this information to find out their future risks of illness and to practice more effective preventive medicine, focusing on their individual most important risks. Perhaps even more importantly, the discovery of the variants that confer risk to disease will also shine a bright light on the pathways involved in the causes of disease, which will suggest new, powerful, and more specific treatments for illness. Additional methods based on genomics will allow very early detection of disease, even before the appearance of any symptoms. This combination of approaches to prevention, diagnosis, and treatment is likely to result in a dramatic transformation of the practice of medicine by 2020.

There is some concern that the fruits of the HGP will be restricted only to countries or populations that can afford what is arguably very expensive technology. Does this concern you? What are some applications of genomics for developing countries, and how should the U.S. and the rest of the developed world help to bring them to fruition?
It is true that the technology associated with the HGP is currently complex. But the ability to read DNA sequence, or to identify specific variants in that sequence, is progressing rapidly with dramatic decreases in the cost, and associated increases in exportability. Thus, it is entirely possible that genomic approaches to disease will become significantly more widespread in the future. Already, for example, one of the most ambitious genome centers in the world has sprung up in Beijing, China, and other such efforts are under way in places such as South Korea, Estonia, and Nigeria. Realistically, however, the massive health disparities that exist across the world, many of them based upon the lack of nutrition and public health in impoverished nations, must be addressed head on. One way that genomics can assist this is the long overdue organized attack on diseases like malaria, tuberculosis, and other parasitic diseases which are the cause of untold suffering and death in other parts of the world. With genomes of many of these pathogenic organisms now determined, a full-scale assault

continued on page 7

on the development of new therapies, based on vaccines or new antibiotics, can be contemplated. However, given the low likelihood of significant commercial markets, this kind of advance will only realistically come about with major support from the governments and private foundations of the developed world.

Since its inception, the HGP has devoted a portion of its budget to the investigation of ethical, legal, and social implications of human genome research, known as the ELSI program. What are the major accomplishments of that program and what are the most significant challenges that remain?
The inclusion of the ELSI program within the Human Genome Project is a bold experiment. Never before has a scientific revolution of this sort attempted to anticipate the social consequences, and to put in place safeguards to maximize the benefits and minimize the risks. In its first decade, the ELSI program has accomplished much. A large cohort of scholars has been developed who are devoting their careers to a study of the ELSI issues and recommending policy options to encourage a good outcome. As an example, the question of genetic discrimination in health insurance and the workplace has been heavily studied, and nearly all experts agree that this requires a federal legislative solution in the United States in order for the public to be fully protected. Many states have now passed such bills, and a congressional legislative initiative is under active consideration in the U.S. Senate. Other important areas where much progress has been made include the issue of genetic privacy, the need for oversight of genetic tests to maintain their clinical validity, the ethical standards that should be applied in carrying out genetics research on human subjects, and the need for public and professional education. Many challenges still remain, as the science has continued to evolve. A major question is how our study of human variation will play out in interpretations of race and ethnicity. As we study more about the human genome, it becomes clear that it is scientifically indefensible to draw sharp boundaries around any particular population group, and yet socially that is often what people have done. Science thus offers an exciting opportunity to underline our similarities and discourage human prejudice, but there is much work to be done to implement that outcome. Finally, the ELSI program is wrestling with the question of boundaries to genetic research, particularly when it comes to the use of genetic technologies to enhance human traits rather than cure diseases. While at the present time most of those scenarios are not realistic, we will need to decide as a society how far we wish to go in applications of technology that fall outside the medical arena.

The ELSI program has supported educational programs for precollege students and teachers. Why is it important for those audiences to know about the HGP and human genetics?
First of all, we hope that many students will learn about the excitement and promise of the study of the genome, and will decide to consider pursuing a career in this remarkable type of science. The opportunities to understand human biology and cure human diseases are unprecedented, and we need the best and brightest of the next generation to roll up their sleeves and help find those answers. But of course, not all students will become scientists. All students will, however, sooner or later be consumers of genetic information as part of their health care. Thus it is critically important for teachers and precollege students to become familiar with the principles and applications of genetics, so that when they are asked to make a decision about whether or not to have a genetic test, or whether or not to buy genetically modified foods, they will be fully informed about the scientific facts.

In the late 1980s, as the scientific community debated whether to pursue the HGP, there was some concern that the project would deflect funds and attention from other areas of biology. Has that proven to be a valid concern?
A furious debate raged in the late 1980s about whether or not the human genome project was a good idea. Not only were many scientists concerned that the costs would be so great that they would distract from other types of scientific research, there was also concern that the project was simply not feasible technically. Fortunately, those gloomy scenarios did not come to pass. While the technical challenges were immense, a wonderfully creative group of biologists, engineers, computer scientists, and experts in automation managed to solve those problems quite successfully. With regard to costs, it is gratifying that the total cost of achieving the original goals of the human genome project turned out to be less than predicted, and the project was accomplished two and one-half years ahead of the originally projected schedule. It is now almost impossible to find a biologist who is not enthusiastic about the Human Genome Project. Most beginning graduate students can hardly imagine how science was done without this information.

This combination of approaches to prevention, diagnosis, and treatment is likely to result in a dramatic transformation of the practice of medicine by 2020.

continued on page 8

Its applications to health care aside, what are the most important things the HGP has taught us about biology?
The HGP has taught us a prodigious amount about human biology. We have learned that our genome seems to include only about 25,000 to 30,000 protein coding genes, less than a third the number expected. This must indicate that each gene is capable of packing a wallop, perhaps in part by the realization that on the average each gene can code for several proteins, using the mechanisms of alternative splicing. By studying the genomes of several model organisms, including the laboratory mouse, we have also learned about the remarkable similarity of living things at the DNA level. For instance, there is a 99 percent chance that any human gene that you decide to study will have a similar gene found in the mouse genome, and many human genes have identifiable homologues, as far away as worms, flies, or even yeast.

What are some of the most important technological contributions of the HGP to biology and to science in general?
The HGP has catalyzed a transition of the field of biology from one in which most work was done in relatively small laboratories, using primarily manual approaches, to a new era where many of the most exciting projects are done on a large scale, using much automation and computational assistance. The HGP has driven specific technical advances such as automated DNA sequencing and high throughput, low-cost genotyping, (measuring DNA variants). Perhaps one of the most important contributions has been the nurturing of the field of computational biology, which has mushroomed from a relatively modest and not well appreciated field to occupying a central place in the study of life processes.

NHGRI recently announced the next set of species whose genomes will be sequenced, including the honeybee, dog, and chimpanzee. How did the institute arrive at those decisions?
While sequencing a genome is becoming easier and cheaper all the time, it is still a major undertaking. We wanted to be sure that the resources available for large-scale sequencing were used in a way that achieved the maximum benefit to science and health. We set up a process whereby individual investigators, or in some instances communities of investigators, who were interested in seeing the genome of a particular organism obtained were asked to submit "a white paper," outlining the reasons why this would be of scientific value. That might be because the organism itself is a valuable experimental model, or it might be because it occupies the unique niche in the evolutionary tree that will teach us about biology; or, perhaps most importantly, because the genome will further help us understand how the human genome functions. The white papers are then reviewed by a distinguished panel of scientists and assigned high, medium, or low priority. When a sequencing center opens up capability in its pipeline, they negotiate with NHGRI staff to choose an organism from the high priority bin, and the project gets under way. We anticipate having draft sequences of as many as 20 vertebrates over the next five years, which will be an incredibly valuable source of information about biology.

BSCS is in a unique position to prepare and disseminate accurate and accessible educational materials for use in classrooms. It will become increasingly impossible for textbooks to be up-to-date, making other resources, such as the World Wide Web and the kinds of materials that BSCS produces, absolutely critical for teachers in the future.

BSCS will celebrate its own 50th anniversary in five years. What are some of the educational challenges and opportunities BSCS will face as a result of the HGP?
The pace of progress in genetics, genomics, and biology has been accelerating rapidly, and is likely to accelerate even further in the coming years. BSCS is in a unique position to prepare and disseminate accurate and accessible educational materials for use in classrooms. It will become increasingly impossible for textbooks to be up-to-date, making other resources, such as the World Wide Web and the kinds of materials that BSCS produces, absolutely critical for teachers in the future.

What pleases you most about the HGP and your involvement in it?
As a physician, I became interested in studying the genome because of the promise that it shows for diagnosis, prevention, and cure of disease. As we begin to glimpse the first dramatic examples of that vision coming true, I feel a great sense of excitement and hope. It has been an enormous privilege to serve at the helm of a project of such historic significance, and to be able to work with a truly remarkable group of top-notch scientists from many different disciplines. Finally, a defining aspect of the Human Genome Project, and one which all of us involved feel very good about, has been the decision to make all of the data immediately accessible over the Internet without any restriction to access, thereby empowering all the brains of the planet to work on understanding this remarkable instruction book and applying that understanding to the cure of disease. I am very proud to have been part of a group that undertook this effort with that kind of selfless and forward-looking vision.