GENOME-SEQUENCING ANNIVERSARY

A Celebration of the Genome, Part I

Ten years ago, the first peer-reviewed reports of the sequencing of the human genome were published. At that time we announced that "humanity has been given a great gift," which has proven to be the case in many ways but has also presented a great challenge. To commemorate the event, we have asked a cross section of insightful individuals—representing many viewpoints—to look at what it has meant to them or their communities to have access to human genome sequences. We will be publishing their comments throughout the month of February.

–Barbara R. Jasny and Laura M. Zahn

Faces of the Genome

Francis S. Collins
Director, National Institutes of Health, Bethesda, MD, USA.

When the draft sequence of the human genome was published in February 2001, *Nature* and *Science* featured human faces on their covers. As striking as these images were, they could be seen as more art than science, because systematic genome-wide sequencing had yet to be applied to individuals for medical purposes. What a difference a decade makes. Real faces are now appearing that demonstrate the medical value of comprehensive genome sequencing.

Researchers with NIH's Undiagnosed Diseases Program recently identified a genetic cause for a rare and debilitating vascular disorder that had baffled the medical field and evaded diagnosis. The discovery was spurred by the cases of Louise Benge and Paula Allen, two middle-aged sisters from Kentucky who had calcification of the large blood vessels and joints in their hands and feet, in the absence of any effect on coronary arteries. Thanks to genomic analyses (just published in the *New England Journal of Medicine*), they now know that their severe leg and joint pain stems from a a mutation in *NT5E*, which encodes a protein that converts AMP to adenosine. Better understanding of the disease mechanism will help to guide development of treatments for such patients, as well as illuminate metabolic pathways involved in calcification.

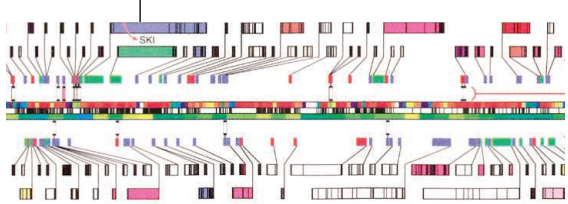
An even more impressive story of the revolution that genomic analysis is bringing to the clinic relates to Nic Volker (shown above), a 6-year-old Wisconsin boy who developed inflammatory bowel disease shortly before his second birthday. Multiple intestinal fistulas occurred, making it impossible for him to eat normally. Despite numerous tests and more than 100 surgeries, doctors remained at loss for a diagnosis and the little boy grew sicker. Then, researchers at the Medical College of Wisconsin carried out whole-exome sequencing, examining the protein-coding regions of every gene in Nic's genome. They identified a mutation in his *XIAP* gene. *XIAP* mutations were not previously associated with bowel symptoms, but had been linked to a severe blood disorder that is curable through bone marrow transplantation. The medical team raised the possibility of a transplant, which would not have been considered without a firm diagnosis. It was performed in July 2010, using stem cells from the cord blood of a matched, healthy donor. Seven months later, Nic appears to be on the road to recovery. While he is still on immunosuppressants, doctors report the new stem cells are stably engrafted, blood counts are good, and there's been no return of bowel disease (http://journals.lww.com/geneticsinmedicine/Documents/GIM200819_Revised.pdf). More important to Nic, he can finally eat solid foods!

These cases, and others that are now appearing in the literature, document that we stand at a significant juncture—the once-hypothetical medical benefits of individual genome sequencing are beginning to be realized in the clinic. This transition will require hard work and ingenuity; in fact, it's one of the many reasons that NIH plans to create a National Center for Advancing Translational Sciences. My hope is that when the day arrives to celebrate the 20th anniversary of the original human genome publications, we will be able to look at a world filled with the faces of people whose health has been improved by the sequencing of their genomes.

The Human Genome at 10: Successes and Challenges

J. Craig Venter
President, J. Craig Venter Institute, Rockville, MD, and San Diego, CA; CEO, Synthetic Genomics Inc., La Jolla, CA, USA.

Fifteen years ago, my team published the first complete genome sequence of a living species, that of *Haemophilus influenzae*, using our newly developed technique of whole-genome shotgun sequencing. This genome sequence was only 1.8 million base pairs in length and required 4 months of sequencing to produce. Five years later, as a result of dramatic changes in automation and massively parallel DNA sequencing, it was possible to sequence the human genome at 3 billion base pairs in only 9 months, a >1000-fold improvement. We published the first individual, diploid human genome sequence in 2007, and now, with single DNA sequencing instruments producing 100 million base pairs per day, individual genome sequences are becoming commonplace.



Most of the new generations of sequencing technologies, although faster and considerably cheaper, produce much shorter sequences (50 to

CREDITS (LEFT): MILWAUKEE (WISCONSIN) JOURNAL SENTINEL; (NATURE COVER) ERIC LANDER AND DARRYL LEJA; (BOTTOM) SCIENCE (SCIENCE COVER) ANN ELLIOTT CUTTING;

Published by AAAS


Xavier Cortada's Sequentia Exhibit, 2010.

200 bp) from smaller DNA fragments than did the strategies to produce the first human genome sequences. Long DNA sequence reads (800 bp) from the ends of long DNA clones (>100 kb) provide scaffolding and extensive DNA assembly by linking together subassemblies. The short sequences can only produce small clusters; these make sequence assemblies of substantial length improbable. Because of these technical issues, some investigators only layer their short sequences against a "reference" and do not try to assemble a sequence, which makes it problematic to define scientific standards for a "genome sequence."

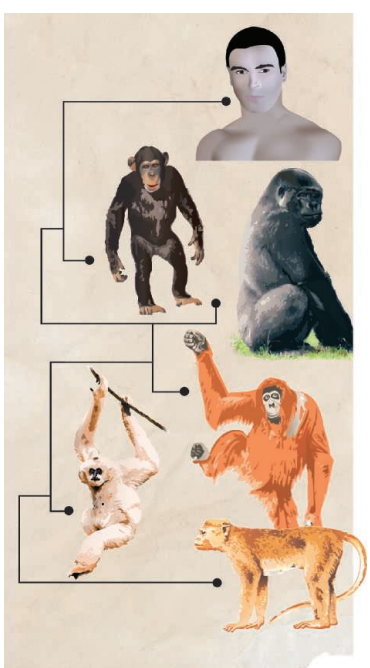
As important as sequence quality standards are, a much larger issue rests with the current state of our ability or inability to interpret human genome sequence. Among the many improvements that are needed in human genome research, the most important is the collection of human phenotypes (according to agreed-upon parameters and standards), in conjunction with tens of thousands of accurate human genome sequences. Such data sets will be the foundation for accurately predicting clinical outcomes from DNA sequence information. This is true not only for diagnosis but also in foreseeing and avoiding drug side effects, as well as monitoring stem cell genome mutations and/or variations before cell therapies.

Although many "genome" companies and researchers are promoting personal genomics for medicine and/or life choices, regulation of data quality and standards is lacking, which has made deceptive marketing a reality in some instances. We have sequence and genetic data quality that is suitable for some scientific analyses but no standards adequate for clinical practice or even for informing individuals of results that exist. We have come a long way in genomics; however, for genome sequencing to reach its full potential we still have a long way to go.

The Golden Age of Human Population Genetics

Molly Przeworski
Professor, University of Chicago, Chicago, IL, USA, Howard Hughes Medical Institute Early Career Scientist

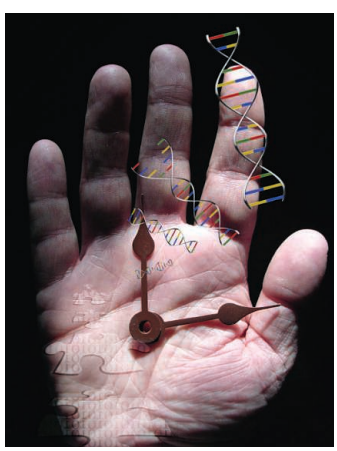
The first draft of the genome provided the road map for the past decade of research in human genetics, allowing for the design of platforms that have been used to query variation in populations worldwide and helping to drive down the cost of sequencing by several orders of magnitude. Within years, tens of thousands of complete genome sequences will be available from humans and from extinct hominids, as well as from thousands of other species. Given the human mutation rate, we will soon know of variation among individuals at almost all sites in the genome. For population genetics, this ushers in a previously unimaginable opportunity to reconstruct the entire genealogical and mutational history of humans and pushes us against the limits of what we will be able to infer about the evolutionary and genetic forces that affected every region of the genome. Why are disease mutations present in human populations? What is the genetic basis of our cognitive and physiological adaptations? What was the sequence of demographic events that led to the colonization of the globe by modern humans? Stay tuned, and before long, we should know as much as genetic data alone can tell us.



CREDIT (TOP): XAVIER CORTADA; (RIGHT) HEMERA/THINKSTOCK; (BOTTOM) N. KEVITIYAGALA/*SCIENCE*

Genomics and Clinical Relevance

Tom Hudson
President and Scientific Director, Ontario Institute for Cancer Research, Toronto, Ontario, Canada



When I interact with physicians, I realize that the clinical questions they think are most relevant are more sophisticated than those posed by the first wave of whole-genome studies. When I ask questions that are clinically relevant, they are not what I asked myself in 2001, when the landmark human genome–sequencing papers were published. At that time, my thinking process was DNA-driven: What is functional? What is normal genome variation? What amazes me in retrospect is that I did not appreciate the fact that genomic information and technologies would grow more than a millionfold in the following decade and, in a way, leap-frog other critical initiatives in health research. Now, clinicians are more and more concerned by overdetection, overdiagnosis, and overtreatment of diseases, as a result of sensitive tests (e.g., prostate-specific antigen for prostate cancer). Whether the disease involves cancer, metabolism, inflammation, or neurodegeneration, it becomes apparent that we have a limited knowledge of disease processes over time and, consequently, limited knowledge of when to intervene and to what degree. In some patients, this leads to unnecessary complications, whereas in others, the failure to act early is irreparable.

If I could move the clock back to 2001 and change course, I would invest significantly more in developing large clinical resources with detailed clinical histories, deep phenotyping, and longitudinal follow-up in order to better understand outcomes and treatment responses. If genomics was now being integrated with such resources, we would be closer to achieving a form of personalized medicine that clinicians would be eager to adopt.

Published by AAAS

The Human Genome at 10: Successes and Challenges

J. Craig Venter

Science **331** (6017), 546-547.
DOI: 10.1126/science.1202812