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.

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 under-detection, over-diagnosis, 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.
The Golden Age of Human Population Genetics
Molly Przeworski

Science 331 (6017), 547.
DOI: 10.1126/science.1202571