Clinical Chemistry 55:12
2088–2092 (2009)

Q&A

By Eleftherios P. Diamandis[1]

Next-Generation Sequencing: A New Revolution in Molecular Diagnostics?

In 1980, Fred Sanger and Walter Gilbert were awarded the Nobel Prize in Chemistry for discovering novel ways for sequencing nucleic acids. In 2003, the human genome sequence was published, an effort that involved more than 3000 scientists from 6 countries. The work took 13 years to complete, at a cost of nearly $3 billion. Only 6 years later, nucleic acid sequencing technologies have advanced to a stage in which a human genome can be sequenced within weeks at a cost of $50 000 or less. These new sequencing technologies are about a million times more efficient than standard Sanger sequencing. Now, people are talking about the $1000 genome, and there is an X Prize worth $10 million for sequencing 100 human genomes within 10 days at a cost of <$10 000 per genome. International organizations are sequencing thousands of cancer genomes to find novel genetic changes, and individuals with money are paying for genomewide association studies in hopes of preventing diseases to which they are predisposed.

Although the technologies for high-throughput sequencing are here and although they are being perfected in terms of accuracy and reduced costs, many questions are being raised. Some of these questions are explored below with leading scientists from academia and industry.

Do you believe that we will ever reach the goal of the $1000 genome, given the complexity of the test, the expensive instrumentation, and the required massive bioinformatic analysis? What is a realistic expectation?

Karl V. Voelkerding[2]**:** Calculating the cost for sequencing a human genome needs to incorporate the level of sequencing "completeness" or "coverage" that will be required to accurately characterize both sequence and structural variation. Reagent and wet bench labor costs for sequencing a human genome should approach $5000 or less within three to five years, depending on the technology. It is difficult to price the costs for bioinformatic analysis, currently a lengthy and extensive process that varies depending on the questions being asked. New computational algorithms will definitely streamline this process. Beyond identification of sequence variants, functional interpretation requires cross-correlation with databases and the use of predictive software. Taken together, the cost of sequencing a human genome should be considered in light of all components that will be required to generate medically interpretable results.

Rade Drmanac[3]**:** Not only is the $1000 genome attainable, I believe that it will be a reality in the very near future. Complete Genomics' technology will, before long, be able to sequence multiple human genomes on a single microscope slide using DNA nanoarrays and provide the ability to read tens of genomes per instrument run using efficient imaging (over 100 frames per second and just a few pixels per spot). These advances will continue to reduce both material and instrument costs per genome. Data analysis is also becoming more inexpensive with recent advances in computing. Complete Genomics is taking advantage of the latest computing technologies and has deployed software architecture than can distribute the workload into a large cluster of computing and storage. This enables us to scale our computing capability as our business scales, so as to efficiently meet our customers' needs. With continued technology advancements, we hope to attain the goal of providing a $1000 genome in the not-too-distant future.

Received September 23, 2009; accepted September 28, 2009.

[1] Eleftherios P. Diamandis, Head of Clinical Biochemistry, Mount Sinai Hospital and University Health Network, Toronto, and Professor and Head, Division of Clinical Biochemistry, Department of Laboratory Medicine & Pathobiology, University of Toronto, Toronto, Canada.

[2] Karl V. Voelkerding, Associate Professor of Pathology, University of Utah, and Medical Director for Advanced Technology, ARUP Laboratories, Salt Lake City, UT.

[3] Rade Drmanac, Cofounder and Chief Science Officer, Complete Genomics, Mountain View, CA.

David Agus[4]: The simple answer is "yes, but with a caveat." It is important to put all of the variables out there. A $1000 genome is meaningless if the error rate is too high. The reason there is a race to perform high-throughput low-cost genome sequencing is to be able to use the information to help personalize each of our clinical disease prevention and treatment strategies. If there is even a small error rate, tests would need to be repeated, which would increase costs dramatically.

John McPherson[5]: The $1000 genome is a lofty goal and will be difficult to attain in the near future. First, one must define what a $1000 genome is with respect to the coverage needed for analysis. Assuming ultrahigh-throughput sequencing accuracy improves, there would still be a need for a minimal coverage to achieve completeness in the likely scenario that whole genome shotgun sequencing will be the method used. This sets the lower limit and will likely be 5- to 10-fold coverage, depending on the error rate. It may be possible with many new developments on the horizon to achieve this at the $1000 level in reagents, but there are other costs such as labor and overhead that must be added. Analysis and interpretation costs will also be required. A $5000 to $10 000 genome is likely more realistic in the near term.

What would you expect to find that would make a difference to, let's say, cancer patients if the genomes of 20 000 cancer patients are sequenced, as is planned by the International Cancer Genome Consortium (ICGC)[6]? Is such an initiative worth the price tag of hundreds of millions of dollars?

Karl V. Voelkerding: The historical progression from morphologic to molecular analyses of cancers has led to improved diagnosis, prognosis, and therapy. Therapeutic interventions based on signature molecular lesions are now the standard of care for certain malignancies. What I would anticipate from the ICGC initiative is a deeper level of understanding of how somatic changes alter cell physiology and how this knowledge will translate to improvements in cancer care. In the context of the frequency of malignancies worldwide and their substantial burden at the personal and economic levels, the investment of hundreds of millions of dollars, in my estimation, is entirely justifiable.

Rade Drmanac: The amount of novel data generated from the sequencing of thousands of cancer genomes will allow researchers to explore the genetic basis of the particular cancers they are studying in a completely new way. For the first time, researchers will be able to investigate associations between the genetic changes observed in cancer genomes and compare them with normal genomes for an unprecedented number of samples. This completeness of data (i.e., developing a comprehensive list of changes in thousands of genomes) will allow for the molecular definition of cancer subtypes according to the type of change and affected pathways. In turn, this should enable the discovery of hundreds of highly informative diagnostic and prognostic markers, and lead to more predictive personalized disease prevention, earlier diagnosis, and improved treatments, while minimizing observed instances of drug resistance and harmful side effects. It is also quite likely that with this new knowledge some existing drugs can be repurposed as cancer treatments. Lastly, at the rate that sequencing technology is evolving, such studies won't be as costly in the foreseeable future—costing tens of millions of dollars instead of the current cost of hundreds of millions of dollars.

David Agus: We have learned a tremendous amount already from the National Cancer Institute's Cancer Genome Atlas Project, led by Drs. Francis Collins and Anna Barker. The cancer mutations seem to cluster around a select number of pathways which have the potential for being targeted. I think the proposed ICGC's project will continue to have impact and utility. The potential is there for the creation of a new lexicon. Instead of categorizing cancer by body part, we can actually describe a cancer by active pathways (it needs

[4] David Agus, Professor of Medicine and Director, USC Center for Applied Molecular Medicine, and Director, USC Westside Cancer Center, University of Southern California Keck School of Medicine, Los Angeles, CA.

[5] John McPherson, Director, Cancer Genomics and High-Throughput Screening, Ontario Institute for Cancer Research (OICR), and Professor, Department of Medical Biophysics, University of Toronto, Toronto, Canada.

[6] Nonstandard abbreviations: ICGC, International Cancer Genome Consortium; SNP, single-nucleotide polymorphism.

to be proven that this categorization scheme will affect patient outcome, but it certainly makes sense to me).

John McPherson: As a member of the ICGC, I am likely biased, but I do believe the price tag is worth it. The idea of the ICGC is to coordinate international efforts to minimize redundancy of research and to attain uniform standards to facilitate data integration. The 20 000 cancer genomes are spread across 50 tumor types and subtypes so that any single tumor type will be sequenced with sufficient statistical power to identify all significant underlying variants. This will be the foundation for a more complete understanding of the mechanisms of tumorigenesis and open the doors for individualized treatment of cancer patients that targets their specific tumor with improved theranostic potential.

If cost were not a factor, why do you think an individual would like to sequence his/her genome now, or in the near future? Please comment also on genomewide single-nucleotide polymorphism (SNP) analysis that is commercially available now. Why would a healthcare professional like to have such information available now, or in the near future?

Karl V. Voelkerding: There are individuals who are quite interested in understanding their genomic composition, for a variety of reasons. Whether at the level of genomewide SNP analysis, or, in the future, at the whole genome level, they are interested in understanding how genomic information can predict their risk for developing medical conditions and how to translate this information into preventive measures to achieve optimum health or to guide medical care when they develop an illness. In an analogous manner, this information is of interest to healthcare professionals who would employ the information in counseling their patients and in facilitating patient stratification and guiding therapeutic interventions. One interesting scenario is that an individual's whole genome sequence, once generated, will be revisited and reinterpreted throughout time as our understanding of genomic medicine increases.

Rade Drmanac: The information contained in an individual's complete genome sequence will be useful once sufficient information is available to accurately associate certain genetic profiles to disease or healthy states and to the likelihood of developing diseases, or to how an individual may respond or not respond to particular classes of therapeutics. In many cases, the genome sequence may indicate the need for early cancer testing (e.g., existence of disruptive mutations in one or more tumor suppressor genes) or other diagnostic tests. The accurate interpretation of all the genomic data generated will be the key to unlocking the utility of complete genome sequencing as a diagnostic or prognostic tool. Current genomewide methods of SNP analysis are extremely limited because they employ technologies that are a surrogate for complete genome sequencing. There is no reason to not use these tests or individual complete genome sequencing today if we are careful to avoid overinterpretation. "Nothing to report, behave as you would without the results obtained from this test" should be a frequent and beneficial result that people and service providers should be satisfied with.

David Agus: As a cancer physician, I realize we aren't very good presently at treating manifest diseases. We are most effective promoting health and wellness when appropriate preventive medicine is practiced. The power of genome association studies—and their resulting data—is that we can personalize prevention. The cardiovascular field is used to dealing with relative risk and prevention through cholesterol testing. Now, through genomewide association studies and disease biomarkers, data are becoming available to assign relative risk for many other diseases. While the genome association studies will not yield a binary answer as to who will get disease, more information is of paramount importance for making personal health choices and decisions. I am one of the cofounders of Navigenics, a consumer genomics and wellness company formed with a message of "know thyself"—using genetic data to personalize disease prevention strategies.

John McPherson: Given the opportunity, I would sequence my own genome. This desire is largely driven by curiosity. I could while away countless hours looking at the variants that would be revealed. Similar, but certainly less complete, snapshots of one's genome are available through genomewide SNP analysis. These too are largely recreational at the present time. Even with complete genome sequence, the significance of the observed changes is largely unknown. There are very few genetic variants that provide meaningful insight into potential health issues at present. Family history is largely a better predictor, and exercise and a healthful diet and lifestyle are better means of ensuring health. I do not see that a health professional will be prescribing complete genome sequencing anytime soon without a means of interpreting the findings. An exception in the near term will be for cancer diagnostics and treatment. Whether through whole genome sequencing or targeted sequencing, gaining insight into the operative pathways in a tumor will help guide treatment soon.

Do you envision that within the next few years all newborns would have their genome sequenced? And why? And who would pay for it? Would the benefit be enough for governments or insurance companies to pay for it?

Karl V. Voelkerding: Sequencing of all newborn genomes raises a provocative question for preventive medicine. I do not envision this occurring within the next few years. It is useful to reflect on the fact that current newborn screening programs for metabolic disorders have required lengthy implementation processes preceded by extensive dialogue within the medical field, with a primary focus on public health considerations including population cost–benefit analyses. These discussions have involved multiple private and public stakeholders at state-by-state, regional, and national levels. I anticipate a similar course would be followed with regard to the consideration of newborn genomic sequencing.

Rade Drmanac: In the next five years, advances in sequencing technology will provide the capacity to affordably sequence the genomes of the four million infants born each year in the US alone. Their complete genome sequences will serve as a universal and complete genetic test to be used throughout individuals' lives to improve their development and help them lead healthier lives. As such, a newborn's sequence should ideally be obtained as early as possible to reduce potential health and developmental risks. However, personal genomic information will be useful only to the extent that the associations between the genetic sequence and diagnosis or prognosis can be accurately made in large numbers of people. Most of these association studies have yet to be carried out, but one can foresee that improved diagnostic and prognostic methods would lead to superior health economics and patient outcomes, despite the likelihood of finding a "healthy" genome in the majority of newborns. Alternatively, to ignore the genetic indicators of potential disease risk would almost certainly result in much higher costs, not only for patients but also for governments or insurance companies, when compared to the cost of sequencing and analyzing a genome. We believe that with the positive healthcare economics rationale, governments or insurance companies will choose to pay for genomic sequencing as a health-screening test.

David Agus: I am not sure we will all have the need to have our genomes sequenced. We are all going to benefit from the heroes who today are giving their DNA and medical history to scientists for study. Whether physicians use targeted sequencing or whole genome sequencing depends on the data generated over the next few years.

John McPherson: No, not within the next few years but perhaps within the next decade it will be feasible. Although technically possible, it would bring up enormous social issues that still need to be resolved concerning various forms of discrimination. Again, the information would be largely unusable from a health perspective in the very near term. It is more likely that genome sequencing will be performed as needed for diagnostic purposes as sequencing costs will continue to fall, making population sequencing more tenable in the more-distant future.

Can you make a projection as to when whole human genome sequencing will become a "routine test" in primary healthcare institutions?

Karl V. Voelkerding: I anticipate that whole genome sequencing will first find clinical utility in cancer diagnosis, prognosis, and management. Whole cancer genome sequencing is expected to be increasingly incorporated into oncology clinical research studies over the next several years. Projecting that medically useful information will be derived from these studies, I expect translation from clinical research into oncology practice to follow. Although our discussion is focused on whole genome sequencing, more immediately, next-generation sequencing is being developed for diagnostic targeted resequencing of medically relevant sets of genes whose protein products participate in physiological pathways or multiprotein structural complexes. In these cases, one needs to sequence a panel of genes with a large mutational spectrum that leads to an overlapping clinical phenotype. Considerable academic and commercial efforts are ongoing to optimize targeted genomic enrichment methods coupled to next-generation sequencing, and these efforts are taking next-generation sequencing into the diagnostic realm.

Rade Drmanac: In the first place, complete genome sequencing can be used as a universal genetic test and should be performed at the earliest possible point in a person's life, independent of any health indications, and the results can be used at all levels of healthcare management (including primary care). The second application for genome sequencing is as a somatic test (e.g., in case of cancer) to identify changes in an individual's genome to diagnose and treat genetically based disease. To use sequencing as a routine test would require the sequencing capacity of several million genomes per year in the US alone. We expect this capacity to be made available before 2015. We expect Complete Genomics' genome sequencing services to play a big role in making this happen. In the next five years, through sequencing of millions of diseased and control genomes, scientists will develop an efficient and com-

puterized genome interpretation process, which is critical for the adoption of complete genome sequencing for broad medical use. At the same time, regulatory agencies and the general public will have to understand the value of these technologies, which will lead to important advances in diagnosis, treatment, and overall improved health outcomes. We expect that the unsustainable increase of healthcare costs will motivate society at large to be open to the early adoption of new methods of preventive and predictive medicine based on complete genetic knowledge.

David Agus: I believe the technology and, most importantly, the utility for the use of the whole genome sequencing will have the potential to be commonplace in 5–7 years. Whether whole genome sequencing, or another technology to evaluate the genome, will be used clinically remains to be seen.

John McPherson: Within 10 years.

Do you foresee any other ethical issues related to whole human genome sequencing? Would this technology facilitate discrimination against those with a "weak" genome?

Karl V. Voelkerding: In 2008, the Genetic Information Nondiscrimination Act was signed into law with the goal of protecting individuals against discrimination based on their genetic information, especially with regard to access to health insurance and employment. Whole genome sequencing fundamentally adds to the complexity of genetic information that will be available, and this law will be increasingly important as we move further into the realm of genomic medicine.

Rade Drmanac: As mentioned previously, we see complete genome sequencing as a promising tool with great potential to improve healthcare and overall human health, and especially to prevent diseases in people with certain genetically based predispositions. We understand that there are important ethical considerations that have to be taken into account in genome research, e.g., the availability of research results to participants and the concern that consumers are likely to misinterpret the genetic information, as well as the obligations, if any, to participants' relatives and the future use of residual samples and data, among many others. Overall, however, we believe that the benefits of having access to one's individual genomic sequence for use in managing one's health are potentially of enough importance to make the work required to manage the ethical risks extremely worthwhile. We believe that continued education, proper regulations, and guaranteed privacy provisions—as with other medical data—will ensure responsible use of this technology to society's benefit and not to its harm.

David Agus: Thomas Jefferson once said, "I know no safe depository of the ultimate powers of the society but the people themselves, and if we think them not enlightened enough to exercise that control with a wholesome discretion, the remedy is not to take it from them, but to inform their discretion." There will be no "weak genome" found; rather, each of us has health-related strengths and weaknesses identified in our genome. Our genes and our behavior and our environment together will dictate our future health. With knowledge will come the ability to focus prevention strategies towards better health and wellness.

John McPherson: We are grappling now with these issues, and fledgling legislation is being enacted to attempt to prevent genetic discrimination. All information about an individual, not just genome sequence, can and will be used in a negative manner by some extremist elements in our society. It is the mandate to us all to ensure that our societal values are not lowered to a level where genetic discrimination is accepted in any form. Genome sequencing is revealing an underappreciated level of imperfection in all our genomes, and this variability should be embraced and accepted as the normal human condition.

Author Contributions: *All authors confirmed they have contributed to the intellectual content of this paper and have met the following 3 requirements: (a) significant contributions to the conception and design, acquisition of data, or analysis and interpretation of data; (b) drafting or revising the article for intellectual content; and (c) final approval of the published article.*

Authors' Disclosures of Potential Conflicts of Interest: *Upon manuscript submission, all authors completed the Disclosures of Potential Conflict of Interest form. Potential conflicts of interest:*

Employment or Leadership: D. Agus, Navigenics.
Consultant or Advisory Role: None declared.
Stock Ownership: D. Agus, Navigenics.
Honoraria: None declared.
Research Funding: None declared.
Expert Testimony: None declared.

Role of Sponsor: The funding organizations played a direct role in the preparation and final approval of the manuscript.

Previously published online at DOI: 10.1373/clinchem.2009.133389